Mentos Regimes: How Individual Uncertainty Affects the Explosive Strength of Resistance Movements—People are the Real Freshmakers.

Joseph Helbing
University of Chicago
helbing@uchicago.edu

Abstract

This article looks at the empirical data on protests in authoritarian countries as a function of regime type and information control, then constructs an agent based model to examine how the effect of uncertainty can help explain the differences in protest and resistance movements in these differing regime types. The agent based model instantiates two agent types, Citizen and Security, and shows how an inverse relationship between uncertainty with regards acceptable public opposition, the probability of suffering costs, and the ability to accurately perceive local regime support lead to differences in resistance movements. Analysis focuses on the speed of resistance spread between agents as a function of individual agent level uncertainty, and how this affects total resistance size, either full equilibrium flips, i.e. successful revolutions, or protracted unrest. Investigation of empirical data shows reduced frequency of protests in more authoritarian regimes and regimes with higher levels of information control. Modeling dynamics further confirms this behavior and shows a potential connection between lower information control and more frequent but slower spreading, smaller scale resistance events while higher information control is connected with with faster inter-agent resistance spread, and larger resistance levels at a reduced frequency.

GitHub Link: https://github.com/JoeHelbing/cascade

1 Literature Review and Research Background

Large protest or resistance movements within authoritarian countries are characterized by cascade transitions, or small acts of dissension sometimes quickly ballooning into societal wide movements. Obvious examples would be the fast-moving events of the Eastern European states in 1989 or the Arab Spring. I hypothesize that authoritarian control over the information sphere within a country reduces the frequency of resistance movements, which is investigated using regression methods, but that lower frequency of protest events is counterbalanced by an increased speed of protest spread and protest size when these movements occur, which will be investigated theoretically using agent based modeling. The word resistance is chosen specifically to imply a more general set of opposing actions than protests per say, as protests can often morph into resistance movements, and vice versa.

Resistance, either as acts of moderate dissension to expected behavior or even outright public protest should not be understood as an individual advocating for liberal democratic change or regime collapse. Resistance to the regime is particular to the information environment of that regime and resistance that creates regime change, successful revolutions, or equilibrium flips is agnostic as to the reasoning. The language of resistance and the nature of individual actions that feed into these movements can and often is couched in the political and ideological framework of the existing regime Hellbeck

(1996), where sup-par regime performance is seen as failing to uphold the ideological underpinnings the regime promotes. Cascades that overthrow a regime can as easily start to protect a regime's founding ideology as much as they can to overturn it.

The hypothesis this model was designed to test originates from work by Kuran (1991), Kuran (1989), Kuran (1995) and his theory on resistance cascades, *public* vs *private preferences*, the gap between those two concepts, and how those two attributes operate in information-controlled environments. In his model, the decision to oppose the government publicly can carry with it various punishments. Given this, an individual may profess public support for the government while privately holding a negative view. The individual has control over their *public preference* while the *private preference* is essentially fixed. Where the *public preference* and *private preference* diverge, the individual is engaging in *preference falsification*. If deciding to actively oppose a disliked regime, the likelihood of persecution is a function of the number of people publicly opposing the government S, where the more people are publicly opposing the less likely persecution is. The reward to any individual agent for making a negative *private preference* their *public preference* is removing the psychological cost associated with *preference falsification*. An individual i has some private preference x_i where the cost is the distance between *public* and *private preferences*. Therefore *public preference* depends on S and x_i such that as public opposition grows, while *private preferences* remain constant, the balance between the external costs and internal costs has a tipping point. Kuran refers to this point as an individual's *revolutionary threshold* T_i. Extending this to a small bottle society example of 10 people, each has their own revolutionary threshold and where total public opposition is represented by S.

$$S = \{\mathbf{0}, 2, 2, 3, 4, 5, 6, 7, 8, 10\} = 1$$

In this model, one person has a revolutionary threshold of 0, and so publicly opposes the government where the then reaches a public opposition of $S = 1$. In this scenario, revolution does not occur as the next person requires more public opposition for feel driven to publicly oppose the government. Our next person in this global information scenario has a negative *private preference* yet maintains a supportive *public preference* which is shown by their low revolutionary threshold of 2. At current the gap between these two, their *preference falsification* cost, is not high enough to push them to a lower revolutionary threshold. Supposing that one of the two individual i of $T_i = 2$ encounters some negative interaction with the government that reduces their *private preference* x_i by 1 point, this reduces their *revolutionary threshold* T_i by 1. Our theoretical 10 person bottle society would then look like this.

$$S = \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{7}, \mathbf{8}, 10 = 9$$

Here we see that the revolution then cascades from 0 to 9 quickly, as each person coming out against the government then prods another person to come out publicly against the government as well. Kuran refers to this as a *revolutionary bandwagon*. Importantly, even those who have a very high revolutionary threshold are capable of announcing public opposition in this model, even those whose *private preference* is pro-regime. Kuran's model looks at costs, and in a fast moving resistance cascade the cost of failing to announce opposition may exceed the *preference falsification* cost if one is pro-regime causing pro-regime agents to change their *public preference* to anti-regime to avoid the potential cost of being one of the last pro-regime holdouts.

Lohmann, 1994 expands on this idea with her work on information transfer within resistance movements, adapting theories from Hirschman, 1970 "Exit, Voice, and Loyalty". In Hirschman's 1970 work, he proposed a market framework for how individuals respond to declining firm performance. Hirschman presents three possible responses: Exit—When dissatisfied parties choose to leave, disengage or discontinue purchasing from the entity. In a commercial context this could be represented as a customer ceasing to purchase goods or for an employee to leave a company, while in a non-commercial context it could understood as a citizen leaving a political party. Exit serves as a market-based mechanism of expressing dissatisfaction where the actors exit puts pressure on the under-performing entity to improve or risk further loss of customers or members. Voice—Instead of leaving, people can choose to voice their frustrations with the intention of seeking improvements. Voice, in Hirschman's model, is a non-market mechanism to drive change. Loyalty—the moderating moderating factor between Exit and Voice. Loyal entities may be more willing to use Voice to express their dissatisfaction and drive change before they decide to Exit, but too much loyalty can lead to inaction, stagnation and further decline. Lohmann discusses the East German resistance movements in the context of Hirshman's Exit, Voice and Loyalty framework, specifically in the interplay of Exit and Voice as informational transfer vehicles. In information-controlled societies where explicitly discussing one's private preferences on the regime in power is verboten, there is limited ability for any individual to know true regime popularity. Thus, implicit and sometimes incidental methods of expressing regime preference become an important method of information transfer between parties. The chaotic emigration from East to West Germany in the summer of 1989 when Hungary dropped its border restrictions with Austria, an Exit choice in Lohmann's theory, was an implicit and incidental informational transfer of regime preference to those still in East Germany. Lohmann's primary contribution in this respect was the informational transfer aspect, but she posited that different types of agents would have different levels of informational transfers. Namely that "extremist" agents, people that are willing to activate even at incredible cost, transfer less information than ordinary people. Seeing a highly politically active ideologically motivated actor expressing public opposition is less convincing than another ordinary person. In some sense this may be true, but there are many cases where "extremist" agents acted as the seeding nucleus to resistance movements, ranging from Mohamed Bouazizi's self-immolation leading to full-scale revolutions, to the Sitong Bridge Protest and the following low level acts of copycat dissension. This theory also assumes an ability and willingness to discern the difference between these two types of actors.

Let us imagine two societies, one high information-controlled, and the other low information-controlled. Information control in this scenario can be roughly understood as authoritarian by layman definitions, but more specifically it is a society where access to information is strictly constrained. There are the obvious ways in which this manifests, state controlled media, internet restrictions, but a highly information-controlled environment also maintains that level of information and societal control by clearly communicating individual behavioral expectations, the consequences of deviating from those expectations and the likelihood of suffering consequences. It is the individual knowledge of these "red lines" and what happens if they are crossed that maintains the closed information environment through self censorship. Vaclav Havel, writing in late 1970s Czechoslovakia Havel and Keane (2016) explains the individual experience of high information control environments beautifully and it is worth including at length.

> The manager of a fruit-and-vegetable shop places in his window, among the onions and carrots, the slogan: "Workers of the world, unite!" Why does he do it? What is he trying to communicate to the world? Is he genuinely enthusiastic about the idea of unity among the workers of the world? Is his enthusiasm so great that he feels an irrepressible impulse to acquaint the public with his ideals?...it can safely be assumed that the overwhelming majority of shopkeepers never think about the slogans they put in their windows, nor do they use them to express their real opinions. That poster was delivered to our greengrocer from the enterprise headquarters along with the onions and carrots. He put them all into the window simply because it has been done that way for years, because everyone does it, and because that is the way it has to be. If he were to refuse, there could be trouble. He could be reproached for not having the proper decoration in his window; someone might even accuse him of disloyalty. He does it because these things must be done if one is to get along in life. It is one of the thousands of details that guarantee him a relatively tranquil life "in harmony with society,"
>
> Obviously the greengrocer is indifferent to the semantic content of the slogan on exhibit; he does not put the slogan in his window from any personal desire to acquaint the public with the ideal it expresses. This, of course, does not mean that his action has no motive or significance at all, or that the slogan communicates nothing to anyone. The slogan is really a sign, and as such it contains a subliminal but very definite message. Verbally, it might be expressed this way: "I, the greengrocer XY, live here and I know what I must do. I behave in the manner expected of me. I can be depended upon and am beyond reproach. I am obedient and therefore I have the right to be left in peace." This message, of course, has an addressee: it is directed above, to the greengrocer's superior, and at the same time it is a shield that protects the greengrocer from potential informers.

The state's ability to impose self censorship through clear expectations has the inverse effect on an individual's ability to correctly perceive the true opinions of their neighbors, friends and family. Whereas an individual is given strong signals about the boundaries of acceptable behavior, they have extreme uncertainty about the opinions of those around them. In low information-controlled societies, the effect is reduced inversely. Where behavioral boundaries are less clear cut, there more possibility for minor acts of dissension that transmit information on true opinion and thus less uncertainty by each individual on their perception on the opinions of others. Alternatively, in societies where expectations are strong and mutually reinforced, each Citizen is an agent of transmission and continuance, but also the application of punishment of this power structure simply by their acquiescence to behavioral expectations.

> Individuals need not believe all these mystifications, but they must behave as though they did, or they must at least tolerate them in silence, or get along well with those who work with them. For this reason, however, they must live within a lie. They need not accept the lie. It is enough for them to have accepted their life with it and in it. For by this very fact, individuals confirm the system, fulfill the system, make the system, are the system.

In this, Havel talks obliquely about the how *preference falsification* is the cost in both these societies, but in one of these, moderate escape from this cost is more possible. Between these two hypothetical societies, high vs low information control, we would expect several differences around the birth and evolution of a resistance movement. In the any authoritarian environment open opposition is verboten but not all systems are able to maintain or even desire to maintain the type of clear rules around what is acceptable public behavior nor the citizen level mutual reinforcement of punishment for breaches. In a low-control environment where open opposition is still verboten, the application of punishments may be sporadic or unpredictable. Citizens are sometimes unsure about is an is not acceptable public behavior, and transgressions of "red lines" may be more frequent. As the range of acceptable public behavior is more blurry, the more frequent acts of dissension transfer information about regime support between citizens, giving agents a more accurate understanding of the general opinions of their neighbors. We would then expect that breakouts of opposition in the low-control society would be more frequent, but the spread and size of these opposition movements to be slower than in high-control environments. In high-control environments, that lack of information on regime support means an act of dissension carries more information.

2 Empirical Observations of Protest Events in Authoritarian Polities

This section focuses on analyzing available empirical evidence of regime type differences in public displays of opposition. The dependent variable of interest comes from the Integrated Crisis Early Warning System (ICEWS) Boschee et al., 2018 which started as a DARPA program in 2008 and is maintained by Lockheed Martin. This data set was then merged with the V-Dem data set Coppedge et al., 2022 which contains the polity score metrics of interest for each country year. To complete the dependent variable construction, population information was used to create a count of protest events per country per year per million people. Finally a set of economic covariates pulled from the World Bank (2023) was appended to data set used in this analysis.

Analysis focuses on the V-Dem polyarchy scale, defined as the rating scale variant for electoral democracy. The polyarchy metric examines whether elections are free and fair, whether suffrage is extensive, and the extent to which political and civil society organizations can operate freely. V-Dem uses multiple country experts to create ratings which are then aggregated to create point estimates. The V-Dem High-level Democracy indices are composites of those aggregated expert coded indices. The rating system is grounded in the work of Dahl, 1971, who defined polyarchy as a political system characterized by the presence of political pluralism, competition, and participation. This particular metric, v2x_polyarchy, combines various indicators, such as electoral competitiveness, freedom of association, and inclusive suffrage, to provide a comprehensive assessment of a country's level of polyarchy. The specifics the creation of the compound Polyarchy metric can be found in the V-Dem Codebook.

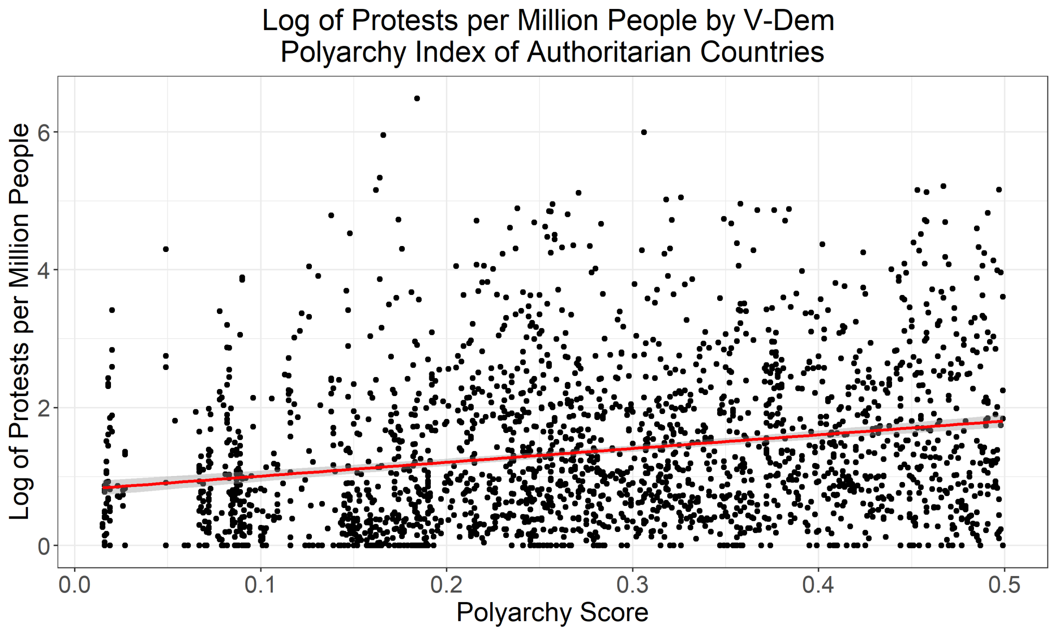

Figure 1: Model 1: Polyarchy Rating Using Original 0.0 - 0.5 V-Dem Scale vs Log Transformed Protests Per Million Per Country Year

The V-Dem polyarchy scales begins at 0.0 meaning least polyarchic, aka most authoritarian, to 1.0, most polyarchic aka most democratic. For purposes of this analysis, below 0.5 is defined as authoritarian and above or equal to 0.5 is defined as democratic. All country year data above the 0.5 democratic cutoff is dropped from the analysis leaving the analysis on a set of 2226 autocratic country years. The count of protests per million was log transformed to pull in extreme values and all protest counts were increased by 1 to avoid log zero errors. In Figure 1 we see a a simple bivariate regression line of the original V-Dem polyarchy score against the log transformed count of protests per million people where a clear positive linear relationship is observed. As countries become less authoritarian on the pohyarchy index, the frequency of protests per million people increases.

All tables were created using the Stargazer R package (Hlavac, 2022). To simplify interpretation of regression tables, the polyarchy scale is multiplied by 100 to create a scale from 0 to 100, so a one unit increase corresponds to a 0.01 increase on the original V-Dem scale. Table 1 examines the relationship between regime type and log transformed protest count per million population. The data set from which the linear regression is constructed does not separate out revolutions, so more authoritarian countries as measured by the V-Dem scale experiencing a revolution with its concomitant increase in protest events should work in opposition to the hypothesized effect. The economic covariates contain some null values reducing the total available country year samples in the full multivariate regression. These country year examples lacking economic covariate values were separated from the data set and regressed as bivariates separately to check their potential effect, found in Table 7 in the appendix. The economic covariate null value country years showed a larger effect size of polyarchy score than the full multivariate regression, and so their removal from the regression in Table 1 does not impact the findings here. Some of the economic covariates have a significant effect, but importantly we see that regime type remains highly significant with the addition of economic covariates. The effect of regime type on protest counts is not simply a correlate or function of economic growth or development stage. In fact, as shown in the bivariate regression in Table 2, the addition of economic correlates increases the model's goodness of fit without robbing any of the observed effect of regime type. Each unit increase in the regression polyarchy score, which is V-Dem score x 100, is roughly equivalent to a 2% increase in protest count events over a baseline of 1.44 protest events per year per million population. Therefore, we would expect that the most authoritarian country to experience about 1.49 protest events per year per million population, while the least authoritarian country at the original V-Dem scale 0.50 cutoff to experience 5.63 protest events per year per million population. As connected to theory, the emprical data shows a correlation between stronger authoritarian regimes and reduced protest frequency.

Table 1: Model 1: Polyarchy Autocracy with Covariates Log Transformed

	Dependent variable:
	Log Transformed Protest Events Per Million Pop
Polyarchy Index 100	0.020***
	(0.002)
GDP Growth	−0.098***
	(0.013)
GDP Growth Lagged 1 Year	−0.006
	(0.018)
GDP per-capita Growth	0.085***
	(0.013)
GDP per-capita Growth Lagged 1 Year	0.011
	(0.017)
Standardized GDP	−0.121***
	(0.037)
GDP per-capita	0.020***
	(0.002)
Constant	0.891***
	(0.077)
Observations	2,000
R^2	0.107
Adjusted R^2	0.104
Residual Std. Error	1.074 (df = 1992)
F Statistic	34.045*** (df = 7; 1992)
Note:	**p<0.05; ***p<0.01

Table 2: Model 2: Polyarchy Autocracy no Covariates Log Transformed

	Dependent variable:
	Log Transformed Protest Events Per Million Pop
Polyarchy Index 100	0.020***
	(0.002)
Constant	0.809***
	(0.057)
Observations	2,226
R^2	0.049
Adjusted R^2	0.049
Residual Std. Error	1.102 (df = 2224)
F Statistic	115.393*** (df = 1; 2224)
Note:	**p<0.05; ***p<0.01

Looking at a simple bivariate analysis of the polycarchy index in Table 2, we see a similar effect with a roughly halved R^2 value. The economic covariates are accounting for only roughly half the seen effect of regime type. Increases in polyarchy continue to correspond with increasing rates of observed protest counts per year as a function of population size at an identical rate. As a country's democratic score rises, we see increased protest numbers. This remains true even when all countries across the polyarchy scale is included up to and including the strongest democracies, though the effect is reduced considerably as we move into democratic regimes. Full bivariate regression of all countries including democracies is included in the appendix in Table 8.

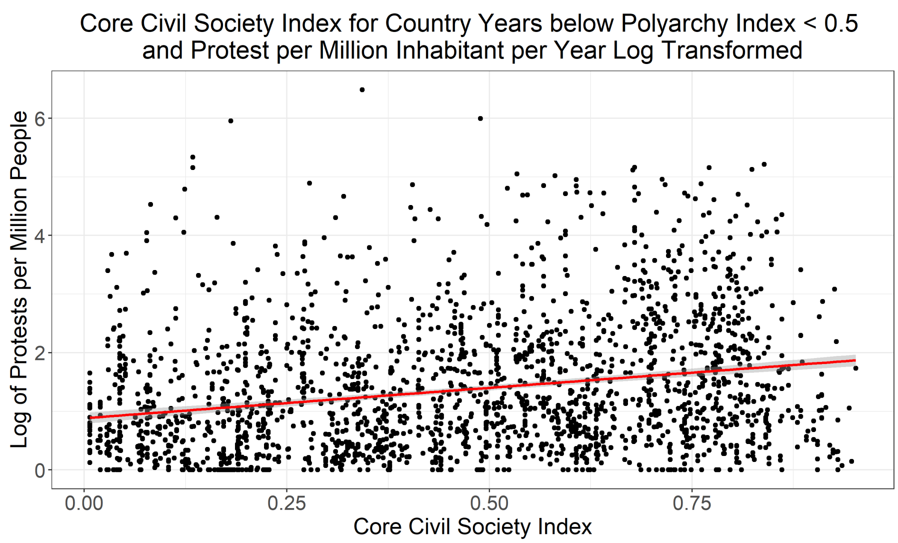

Figure 2: Core Civil Society Index—v2xcs_ccsi for Countries Years below Polyarchy Index < 0.5 and Protest Per Million Inhabitant per Year Log Transformed

Using the polyarchy score of 0.5 as a cutoff separating democracies from authoritarian polities, we are able to drill further down into the question of the information environment more specifically. A secondary data set was constructed from only those county years below the 0.5 cutoff on the polyarchy scale, and regressed against measures of civil society and the information environment. We continue to see a positive correlation between more open regimes and protest events in Figure 2. The **Core Civil Society Index** is a composite index produced by V-Dem to measure the extent to which citizens are able to organize and pursue collective interests and ideals outside the confines of state controlled systems. The CCSI is looks the prevalence of Civil Society Organizations such as interest groups, labor unions, professional associations, charities and other non-governmental organizations. It is a composite measure of CSO entry and exit, CSO repression, and CSO participatory environment intended to reflect the extent of a polity's robust civil society.

The CCSI 3 has a reduced effect in comparison to the full composite score of regime type. The effect of a single point increase in the composite CCSI score has half the effect of regressing against regime type, implying that regime type includes a more holistic connection to the effect we are investigating. A limited civil society environment still maintains a positive and statistically significant effect on the number of expected protest events per year per million inhabitants.

Table 3: Model: Core Civil Society Index—v2xcs_ccsi for Countries Years below Polyarchy Index < 0.5 and Protest Per Million Inhabitant per Year Log Transformed

	Dependent variable:
	Log Transformed Protest Events Per Million Pop
CCSI 100	0.010***
	(0.001)
Constant	0.883***
	(0.049)
Observations	2,226
R^2	0.054
Adjusted R^2	0.054
Residual Std. Error	1.099 (df = 2224)
F Statistic	127.689*** (df = 1; 2224)
Note:	**p<0.05; ***p<0.01

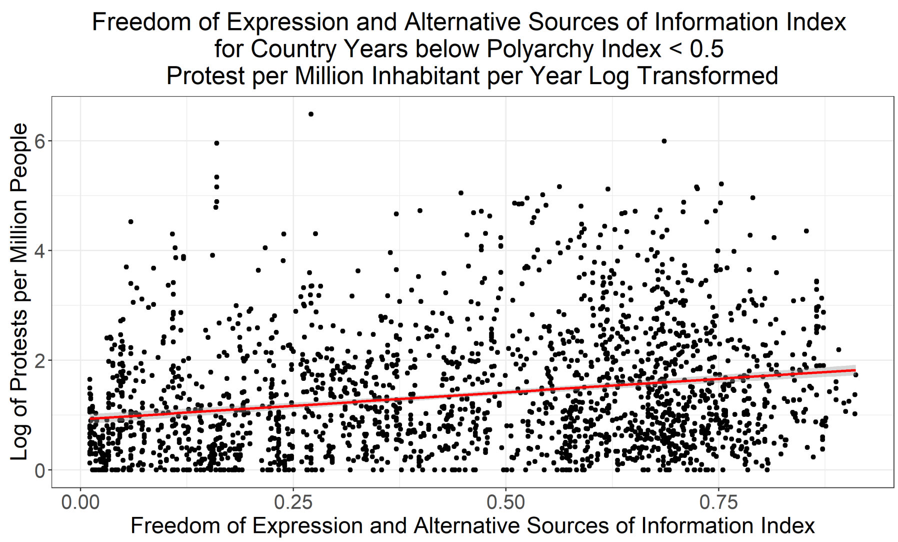

Figure 3: Freedom of Expression and Alternative Sources of Information Index—v2x_freexp_altinf for Countries Years below Polyarchy Index < 0.5 and Protest Per Million Inhabitant per Year Log Transformed

The **Freedom of Expression and Alternative Sources of Information Index**, another composite score produced by V-Dem, examines the information sphere within a country most directly. This metric looks at press and media freedom, academic freedom and cultural expression as well as the ability of ordinary people to discuss politics and express political opinions both in public and at home. FEASII. The sub 0.5 country years from the polyarchy index are plotted in Figure 3 directly against the log transformation of protest counts. In a bivariate linear regression against the log transformed protest count, we see an identical effect size in Table 4 to CCSI, with a slightly reduced R^2 value, implying that it is less explanatory. Additional bivariate regressions of other information sphere related metrics are located in the appendix showing universally positive relationships and broadly similar effect sizes.

In this admittedly coarse evaluation of the connection between regime type and protest frequency, we are able to find some evidence that more authoritarian countries do indeed experience fewer protest events and that the information environment is a factor in this. This analysis is complicated by two important factors worth noting in the original data set. The ICEWS data set has issues with

Table 4: Model: Freedom of Expression and Alternative Sources of Information Index—v2x_freexp_altinf for Countries Years below Polyarchy Index < 0.5 and Protest Per Million Inhabitant per Year Log Transformed

	Dependent variable:
	Log Transformed Protest Events Per Million Pop
FEASI Index 100	0.010***
	(0.001)
Constant	0.925***
	(0.047)
Observations	2,226
R^2	0.050
Adjusted R^2	0.049
Residual Std. Error	1.102 (df = 2224)
F Statistic	116.476*** (df = 1; 2224)
Note:	**p<0.05; ***p<0.01

accuracy due to machine coding of security events. Human based re-coding of events shows about a 70% accuracy rate Boschee et al. (2018) in the methods by which ICEWS compiles data. The most authoritarian regimes also likely have a biased under-reporting and collection of all security events, including protests which complicates any specific claims relating to protest frequency and regime connection in a quantitative sense. Rather it is the breadth and constituency of the effect that is weakly supportive of the hypothesis, though in no way definitive. To better understand this connection, we shall look toward agent based modeling to tease out how individual attributes and decision processes can lead to the empirical data we see here.

3 Model Design Overview

3.1 Overview

3.1.1 Purpose

As discussed previously, though Kuran (1991) and Lohmann (1994)'s adaptation of Hirschman (1970) both speak directly on the issue of individuals' lack of information on general regime support being an impediment to action, neither discuss specifically on individual perceptions of what is and is not acceptable public behavior in information-controlled societies, nor on individual perceptions of regime support as mediating factors in resistance cascades. The agent based modeling of this paper is an attempt to synthesize this previous work with individual perceptions of those two aspects of information-controlled authoritarian societies. Namely, in this model an agent level error term *epsilon* ϵ_i representing uncertainty and information scarcity interacts with each individual agent's perceptions of what constitutes acceptable and unacceptable public behavior, the probability of suffering consequences for unacceptable behavior, and the individual perception of regime support within each agent's local environment. This heterogeneous individual uncertainty and intertwining effect of millions of inter-agent interactions across a single simulation is fully track-able in a synthetic environment and gives us insight into the empirical findings discussed above.

Model design is described via an adaptation of the ODD + D Müller et al., 2013 framework. The model seeks to examine differences in the speed, size, oscillation characteristics and equilibrium changes of resistance cascades in different authoritarian regime types typified by different levels of information control as a function of the individual agent level uncertainty parameter epsilon ϵ_i. The expectation adapted from the above empirical data is that higher uncertainties with regards to state expectations in less strict authoritarian environments causes more frequent but slower spreading resistance cascades with fewer equilibrium flips and more oscillations.

3.1.2 Entities, state variables and scales

The model consists of two types of agent—Citizen agents which are the primary agents of question and are the subject of most macro-scale model measures, and Security agents. While this paper is mainly focused on the complex emergent cascade behaviors of Citizens, the interactions of Citizens and the state is core to understanding the processes. In multiple case studies of resistance movements, Security forces act as the primary foil of Citizens heavily influencing their choices of if, how, when, and where to activate publicly. Even in scenarios where a cascade has obviously begun, Security forces have significant ability to shape events Pearlman, 2017 Lee, 2009.

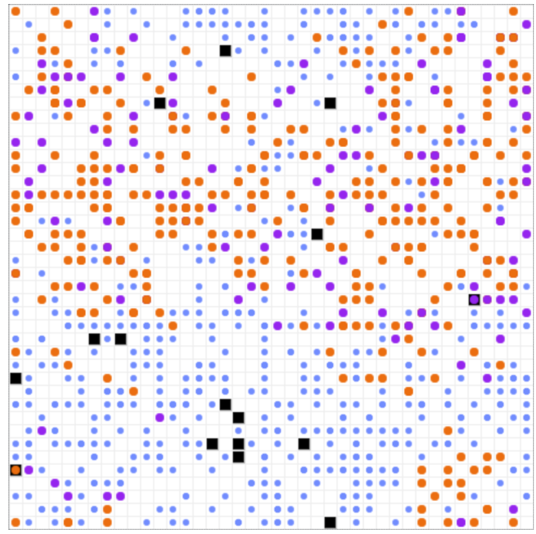

Figure 4: Visualization of resistance cascade model where the circles represent Citizen agents in their three *public preference* states: blue being **Support**, purple being **Oppose**, and red being **Active**. Security are represented by the black squares. In this figure we can see a cascade propagating through the model as more agents flip active as they see other agents flipping active.

The primary agent level attribute of Citizen agents is *private preference* which derives its theoretical basis from "Now out of Never" Kuran, 1991. In the paper, Kuran defines private preference as some internally held opinion on a regime, either for or against the status quo, which at any point in time is essentially fixed. Citizens are also defined by a value *epsilon* which is the primary point of research for this paper. *Epsilon* is the operationalization of uncertainty across various regime types and levels of information control. In different regime types, those of higher or lower information control, how each individual in that society interacts with regime expectations of what is and is not an acceptable public opinion, methods of display of those opinions, and the internal private preferences of one's neighbors, family, and friends carries with it a level of uncertainty. How that individual uncertainty interacts with the environment and the state can describe in part the differences we see in resistance cascades across these varying regime types.

The Citizen agents *public preference* or visible state is a function of the above two exogenous variables. The Citizen agent can occupy one of the three states which is visible to other agents. The *private preference* of each agent is an assigned and unchanging value representing their unspoken privately held opinion on the regime in power, but each agent's *public preference* is self determined in the sense that each individual agent decides their publicly viewable state by incorporating spatial information viewable by all agents, in combination with their internal non-public information. The three states or *public preferences* of Citizen agents are **Support**, **Oppose**, and **Active**. Citizen agents can inhabit a fourth state **Jailed** imposed on it by Security forces where they are removed from the board and await release.

The exogenous factors within the model are the two above variables, *private preference* and *epsilon*, as well as *vision*, *Citizen Density*, *Security Density*, *threshold* T_C, and *maximum jail term* J. *Vision* refers to each agents vision radius. Each agent has a set distance at which they can view other agents. While this can be understood as a literal representation of spatially local information restrictions, it also represents an abstraction of limited information. This variable can be adjusted separately for each class of agent, Citizen or Security, to represent more restrictions on Citizen agents' information access while holding Security constant, but the model default is *vision* radius 7 for both agent classes.

Decisions on whether to change state or *public preference* is based on the exogenous global variable *threshold* T_C. Each agent's individual *epsilon* ϵ_i interacts with the global threshold value to set their own personal threshold for activation. The changes in standard deviation of *epsilon* correspond to the differing regime types, with more information-controlled societies having a lower *epsilon*, aka lower standard deviation in a Gaussian distribution, while lower information-controlled societies have a higher *epsilon*, or a higher standard deviation in a Gaussian distribution.

Citizen Density and *Security Density* determines the number of Citizen and Security agents as a percentage of available space within the spatial grid. Using Epstein's Civil Violence Model Epstein, 2002 as a starting point for understanding the behavior of different agent densities, the *Citizen Density* was set at 0.7 and *Security Density* was allowed to fluctuate between 0.00 and 0.09. *Maximum Jail Term J* is the maximum integer value that a Security agent can impose on an active Citizen agent. This is applied stochastically during an arrest as a uniform distribution between 0 and *Maximum Jail Term*.

The model works on a multilevel grid where multiple agents can occupy the same grid square simultaneously by default. Through the activation of a user parameter the model can operate on a single layer grid where each grid cell is limited to a single agent at a time. The grid itself either in single-layer or multi-layer is a torus where the top, bottom, and sides are connected. Agent vision and movement is able to jump from the bottom to the top, or from one side to another. The grid has a height of 40 squares, and a width of 40 squares for 1,600 total squares. This grid size also defines the agent count via the *Citizen Density* and *Security Density* variables expressed as a proportion of total squares on the grid. Agents move one square per step in their Moore neighborhood if any move is available. A Moore neighborhood includes all squares adjacent to a given square both orthogonally and diagonally, for a total of eight neighbors in a two-dimensional grid, excluding the square the agent inhabits. An agent's spatial vision is defined as the radius of the exogenous *vision* variable in their Moore neighborhood. Thus, an agent with *vision* radius 7, the default value in the model, would be assessing $(2 * v + 1)^2 = (2 * 7 + 1)^2 = 225$ squares in their vicinity, some of which will contain a single agent, some multiple agents, and some no agents in the default multi-layer setup.

The model is split into two temporal stages for each agent's decision and action phase using a simultaneous activation scheduler. The scheduler first loops through each agent in a random sequence where agents decide on their future state in a static environment. The scheduler then loops through the agents a second time in a random sequence who then activate their stored chosen state or *public preference* and then take their actions. With this temporal setup, all states of each individual agent in the step function are predetermined during the first loop in a static environment, and so state declarations by any agent are independent of the evolving state declarations of other agents in the action step.

3.1.3 Process overview and scheduling

A full step for each agent type is outlined below describing which decisions and what actions happen in what order.

<u>Decision Phase</u>

A Citizen agent first moves through their state decision phase encapsulated in the step function. The agent first checks if it is jailed, and if so, terminates their decision phase. If the agent is not in jail, they then look at their neighbors within their *vision* radius and based on their local environment and personal attributes choose which state or *public preference* they will have during the action phase.

A Security agent makes no choices in their decision phase.

<u>Action Phase</u>

Citizen agents checks whether they are jailed, and if so reduces their jail term by 1 increment then terminates their action phase. If the agent is not in jail, they assume their stored state or *public preference* and then moves to a random valid square in their Moore neighborhood of radius 1.

The Security agents looks at only those agents within their Moore neighborhood of radius 1, and if one or more **Active** Citizen agents exist, chooses one at random to arrest. If no **Active** Citizen agents exist, the Security agent then looks at all **Opposed** Citizen agents in radius 1 and if any exceed the global threshold constant T_C for activation, aka they are exceeding "the red line" for acceptable behavior, they arrest a single qualifying Citizen agent at random. The Security agent then moves to a random valid square in their Moore neighborhood of radius 1.

3.2 Design Concepts

3.2.1 Individual Decision-Making

Citizen agents make a single decision, which state or *public preference* they will assume for the next turn. Citizen agents decide their state based on their visible local environment, how many other **Active** agents they see, how many other **Oppose** agents they see, and how many Security agents they see. Citizens agents are not optimizing in pursuit of a goal per se, but *private preference* represents their preferred state absent other factors, and a *public preference* other than their preferred state represents *preference falsification*. Citizens in this model do not have a long-term strategy for escaping *preference falsification* and are simply reactive to their immediate environment, but take the opportunity to reveal that *private preference* when the right scenario arises.

Where Citizen agents have a positive, aka pro-regime, *private preference*, a resistance cascade can sometimes cause them to flip **Active** thus creating an anti-regime *preference falsification*. This makes sense as internal East German Socialist Unity Party of Germany (SED) polling prior to regime collapse showed mediocre internal party support for the regime even in the months immediately prior to the revolution Lohmann, 1994, and low though not non-existent support among sections of the public. Following the dissolution of East Germany and its unification into West Germany, polling showed nearly non-existent support for the prior communist regime. In a cascade model, this is an expected outcome. During a resistance cascade, there comes a point where the potential costs of being the last holdouts wedded to an unpopular regime outweigh the possible costs of defying it. If an individual agent in their limited information environment makes the determination that the regime is unlikely to survive, it makes rational sense to announce public opposition even where one's *private preference* is in support of the regime.

Security agents decide who to arrest in their vicinity if any Citizen agents meet the criteria for detainment. Security decisions to arrest a Citizen agent are two sided. Citizen agents are looking to minimize *preference falsification* and so can choose one of three states, **Active**, **Oppose**, **Support**. Citizen agents have access to and complete knowledge of their own *private preference* with no error term, while Security has access to the un-transformed *threshold* value that represents the "red line" for acceptable public opposition within the information-controlled society. Citizen agent's own threshold values are altered from the global threshold value each by their own individual error term pulled from a stochastic distribution. As Citizen agents are looking to minimize *preference falsification*, based on the above discussed decision model for Citizen states, will choose to **Oppose** at a level they believe to be below where the regime determines to be unacceptable. **Active** opposition is always unacceptable. The Security agent looks at their immediate Moore neighborhood of radius 1, i.e. the 8 squares in their vicinity, and arrests a single **Active** Citizen at random if present. If no **Active** Citizens are present it reviews any **Oppose** Citizens to see if their behavior exceeds acceptable levels, and arrests one of these agents at random if so.

3.2.2 Individual Prediction

Citizen agents have a perception term moderating their understanding of local regime support which interacts with their error term. The Citizen agent assesses the proportion of other Citizen agents in their vision either in the **Oppose** or **Active** *public preference* state, and interacts this ratio with their error term to estimate actual regime support as they understand it. Agents calculate their own state based on this information simultaneously with other agents then store that state decision for the simultaneous reveal phase.

3.2.3 Heterogeneity

Agents are heterogeneous in their private preferences and error term distributions. All Citizen and Security agents at initialization are assigned a *private preference* P_i from a Gaussian distribution of varying user-defined mean and standard deviation of 1. Citizen agents are assigned an error term ϵ_i from a Gaussian distribution of mean 0, and varying user-defined standard deviation. The interaction of this error term means all Citizens have heterogeneous threshold values, probabilities of arrest, in addition to their differing perceptions of regime support. All Citizen agents use the same decision function when assessing their next state.

3.2.4 Stochasticity

Citizen activations are inherently stochastic. The individual Citizen's uncertainty interacted threshold values exist as alterations to a probability calculation. The expanded function is outlined in detail in the submodel's section, but a Citizen's private and environmental information is passed through a sigmoid function to output a probability of activation for each step. The individual assignment of *private preference* P_i and error term *epsilon* ϵ_i values are also stochastic. Agents move at random when not **Active** and Security arrest **Active** and **Oppose** agents exceeding the global threshold constant T_C at random.

3.2.5 Observation

The batch run of the model during data collection is designed to use a user-defined seed as a parameter. A seed, in the context of a random number generator, is an initial value or input that is used to initialize the Psuedo-Random Number Generator (PRNG) and determine the sequence of pseudo-random numbers it will generate. By setting seed as a parameter, although the numbers will generate as if at random, that sequence of random numbers will be the same across every run as long as the same seed is used. With this, each run of the same seed will isolate the effect of our parameters of interest while maintaining randomness.

The batch run process collects model-level aggregate data and agent level individual data. model-level data contains several key metrics and characteristics of the model that help in understanding and analyzing the model's behavior. These metrics include Seed, Citizen Count, Active Count, Oppose Count, Support Count, Speed of Spread, Security Density, Private Preference, Epsilon, and Threshold. These data are collected at each time step during the simulation, providing a detailed view of the model's dynamics over time.

Agent level data contains various attributes of individual agents in the model, such as their position, *public preference*, opinion—the pre-sigmoid *public preference* state calculation, activation—the post-sigmoid *public preference* state calculation, *private preference* P_i, *epsilon* ϵ_i, oppose threshold T_O, active threshold T_A, jail sentence, flip status—Whether the agent has flipped **Active** or **Oppose** on the current step, and whether they have ever flipped. Additionally, it also includes some model-level attributes for ease of analysis when working with the data-frames. These agent-level data are collected at each time step as well, allowing us to analyze how agent behavior leads to cascade events, or the lack thereof.

3.3 Details

3.3.1 Implementation Details

The model was written in python using the mesa package Kazil et al., 2020.

The full source code for the model can be found at https://github.com/JoeHelbing/cascade

3.3.2 Initialisation

The model is initialized with either a set or random seed that can be pulled from the log files, from the terminal, or from the data-collector files. The model then creates a grid 40 cells wide and 40 cells high for 1600 total cells. The model defaults to initializing as a multi-layer grid where multiple agents can stack on a single cell at a time, but can initialize as a single layer grid where each grid cell is exclusionary and only a single agent can inhabit any particular grid cell. Using the global model parameters, Citizen density, Security density, the number of agents of each type is calculated as a proportion of total cells in the grid, where Citizen density of $0.7 * 1600 = 1120$ Citizen agents.

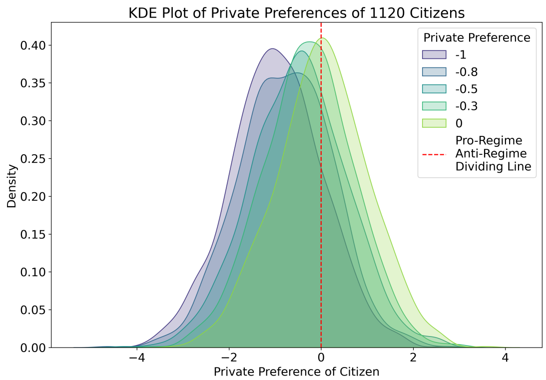

Figure 5: Example KDE Plot of Private Preference Distributions with Varying Means)

Private Preference	% Pro-Regime
-1.0	12.9%
-0.8	20.0%
-0.5	30.5%
-0.3	36.3%
0.0	50.2%

Figure 6: Binary pro-regime breakdown of Citizen agents as a percentage of total agents. A synthetic poll, essentially.

Given the agent counts, the model initializes the agents. Citizen agents are first assigned an agent specific *private preference* P_i from a Gaussian distribution of user-defined distribution mean and standard deviation 1. The *private preference* corresponds to an agent's opinion of the regime in power, where negative values are to varying degrees anti-regime, and positive is correspondingly varying pro-regime, and distance from 0 is strength of that preference. Because of the stochastic nature of the construction of the distributions, the number and strength of pro and anti-regime Citizen agents will vary between each run but the values found in Figure 4 are roughly the expected levels of regime support for each *private preference* parameter.

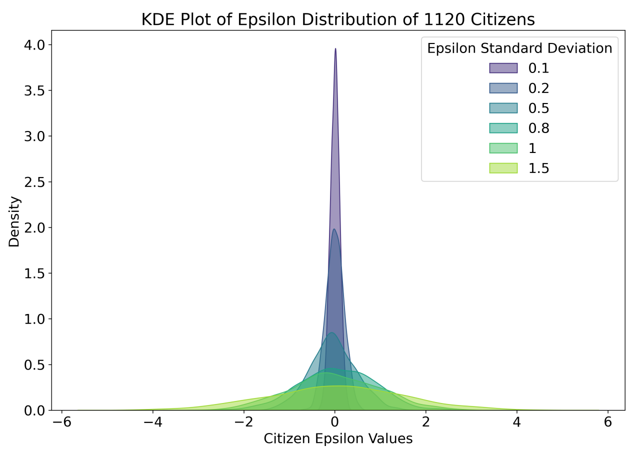

Figure 7: Example KDE Plot of Epsilon Distribution with Varying Standard Deviations

The model assigns epsilon ϵ_i values for each Citizen agent from a Gaussian distribution with mean 0 and varying user-defined standard deviation. The threshold values for **Oppose** and **Active** *public preference* states are then calculated from the global threshold constant T_C and the agents individual epsilon ϵ_i. A Citizen agent's oppose threshold T_O is assigned as the lower value from a 2 value pair pulled from a Gaussian distribution with mean of the global threshold parameter T_C and standard deviation of the individual uncertainty parameter *epsilon* ϵ_i. An agent's activation threshold T_A is the larger of these two values. A breakdown of *epsilon* 0.1 and 1.0 is seen in Figure 8.

thresholds = [random.gauss(T_C, ϵ_i) for _ in range(0, 2)]

$$T_O = min(thresholds) \qquad T_A = max(thresholds)$$

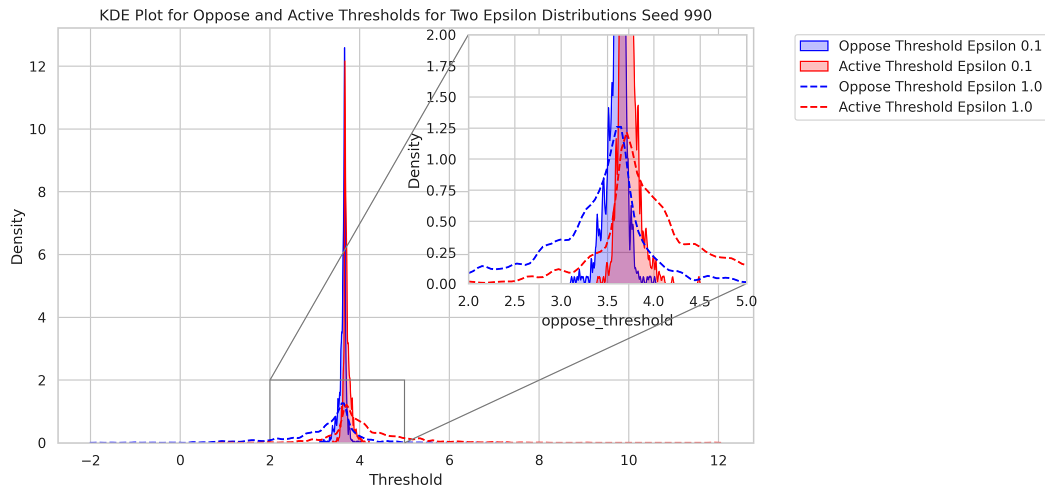

Figure 8: Distribution of Activation Thresholds for Seed 990 for Two Different Epsilon Values Overlapped with Inset using Global Threshold Constant T_C of 3.66356

After assigning epsilon ϵ_i values, calculating thresholds, and assigning *private preferences*, the model then places Citizens and Secruity on the grid at random, and adds them to the simultaneous scheduler.

3.3.3 Submodels

After initialization the model initiates the first step. Agents enter their decision phase which looks at all agents in their vision radius and calculates their state decision. Citizen agent's *public preference* state decision calculation is represented mathematically in the following form. In the model, the pre-sigmoid value of the activation function is referred to as Citizen opinion.

$$Active = \sigma(-P_i + R * L_i - T_O) - A_i \qquad Opposed = \sigma(-P_i + R * L_i - T_A) - A_i$$

Which when expanded...

$$Opposed = \sigma(-P_i + \frac{A+O}{S} * (A + O * \sigma(\epsilon_i))^{\frac{1}{\epsilon_i^2+1}} - T_O) - A_i$$

$$Active = \sigma(-P_i + \frac{A+O}{S} * (A + O * \sigma(\epsilon_i))^{\frac{1}{\epsilon_i^2+1}} - T_A) - A_i$$

- P_i—Private Preference
- R—Ratio of **A**ctives and **O**pposed to **S**upport Citizen agents in vision.
- L_i—Moderator of perception of spatially local regime support.
- A_i—Calculated arrest probability.
- T_O—Individual Oppose Threshold.
- T_A—Individual Active Threshold.

P_i is multiplied by -1 so that a negative *private preference* corresponds to a positive move up the probability of activation in the simoid σ function. In isolation of other factors, very few agents have a low enough *private preference* to exceed the model's global threshold constant for activation. For any cascade to start though, there must be at least one agent with a low enough *private preference* that that in combination with their epsilon ϵ_i parameter that they are willing to change to the **Oppose** or **Active** state when the probability of arrest and expected cost are 0. The analog of this would be that at least one agent must be dissatisfied enough with the regime, uncertain enough about where "the red line" is in terms of acceptable behavior, and believes they will get away with whatever action they decide to take that they would be willing to engage in some form of resistance to alleviate their *preference falsification*.

$$R = \frac{A + O}{S}$$

R is the ratio of **A**ctive and **O**pposed Citizens to **S**upport Citizens within a the Citizen agent's vision radius. The agent always counts themselves as **Active** and when interacted L_i, the Citizen's perception of that ratio, this produces the variable pair that can create resistance cascade events.

$$L_i = (A + O * \sigma(\epsilon_i))^{\frac{1}{\epsilon_i^2+1}}$$

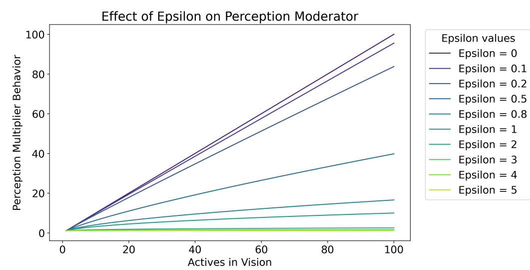

Figure 9: Graph of the Behavior of the Perception Moderator L_i for Individual Citizen agent's ϵ_i epsilon values by active agents in vision

L_i is the encapsulation of a Citizen agent's uncertainty of the true *private preferences* of their neighbors. This interacts with their accurate view of the ratio of who in their vicinity is **Active** or **Opposed** to create their perception of regime support. As discussed earlier in the theoretical discussion of the model, the more information control in a society, the more accurately Citizens are able to understand where "the red line" falls, but the less able they are to understand the closely held true opinions of their neighbors. As epsilon ϵ_i in L_i gets smaller, it exaggerates the effect of each **Active** Citizen in view. And while it may certainly feel this way sometimes, no person has infinite uncertainty, so a constant 1 is added in the denominator to tamper the effect of ϵ_i at extremely low values. When ϵ_i is 0, this acts as a linear moderator to the ratio of **Actives** and **Opposed** to **Support** as seen in Figure 9. As this is passed through a sigmoid function, the moderator pushes agents higher toward certain activation with a probability of activation value of 1. Figure 20 in the appendix shows an example breakdown of different epsilon effects for varying **Active** agent counts.

$$A_i = 1 - e^{-2.3\frac{S}{A}2\sigma(\epsilon_i)}$$

A_i is the individual perception of the probability of arrest if **Active**. This function, adapted from Epstein, 2002 "Modeling Civil Violence", is designed so that an agent always counts themselves as **A**ctive, which is an agent's calculation of "if I were to activate this turn, what is the probability that I would be arrested". In a scenario where a Citizen sees no other **A**ctive agents in their vision, and see one **S**ecurity agent, the ratio $\frac{S}{A} = \frac{1}{1}$ and the constant $k = -2.3$ then computes to $A_i = 0.9$ absent the interaction with error term *epsilon* ϵ_i. The function has a maximum value of 1 and a minimum value of 0. The example given by Epstein is that of a person with a Molotov cocktail looking at a storefront. If that person is standing alone on the street looking at 9 **S**ecurity officers arrayed around the storefront, the person would calculate that throwing a Molotov cocktail would produce an extremely high likelihood of being arrested. If on the other hand, that person was standing in a crowd of 200 other rioters facing the same 9 **S**ecurity officers, then that person would reasonably calculate that throwing a Molotov cocktail would produce a much lower chance of being arrested. A_i is an approximation of the safety of crowds, the smaller the ratio of **S**ecurity officers to **A**ctive Citizens, the lower the value of the probability of being arrested. In the Resistance Cascade model, this function is altered to account for uncertainty. A Citizen with epsilon $\epsilon_i = 0$ $\sigma(\epsilon_i) = 0.5$ $2 * \sigma(\epsilon_i) = 1$ has no error in calculating their probability of arrest, but those with more uncertainty over or underestimate their chances.

The final probability for activation and is compared versus a randomly chosen float value pulled from a uniform distribution between 0 and 1, first checking if an agent will choose an **Active** state, and if not then checks the **Oppose** probability, and if neither defaults to a **Support** state.

4 Model Results

Two simulations were run, one simulation run was conducted with the below variables using a multi-layer grid for a total of 7,200 model runs to examine the effect of various parameters to hone the analysis simulation run. A second simulation of 10,000 model runs was conducted across a smaller set of parameters and a larger number of random seed values to examine the behavior of the model at scale across random conditions. All model runs were capped at 500 steps for both sweeps.

Table 5: Model Parameter Examination Sweep

Parameter	**Values**
Seed	213490, 213491, 213492, 213493, 213494
Private Preference	-1, -0.8, -0.5, -0.3, 0
Security Density	0.00, 0.01, 0.02, 0.04, 0.07, 0.09
Epsilon	0.1, 0.2, 0.5, 0.8, 1, 1.5
Threshold	0, 1.38629, 1.7346, 2.19722, 2.94444, 3.66356, 4.18459, 4.59512

The effect of Security and the position of "the red line", aka the global threshold constant T_C are both crucially important theoretically and practically to the operation of the model. The position of the global threshold constant is highly determinant of how many full equilibrium flips, revolutions, and Citizen activations occur through model runs. Low thresholds produced high numbers of activations, while high threshold values nearly excluded activations entirely. To produce a reasonable subset for analysis the threshold value 2.94444 was chosen and security densities above 0.04 were excluded.

Table 6: Seed and Parameter Examination Sweep

Parameter	**Values**
Seed	344000 to 344099
Private Preference	-1, -0.8, -0.5, -0.2, 0
Security Density	0.01, 0.02, 0.03, 0.04
Epsilon	0.1, 0.2, 0.5, 0.8, 1
Threshold	2.94444

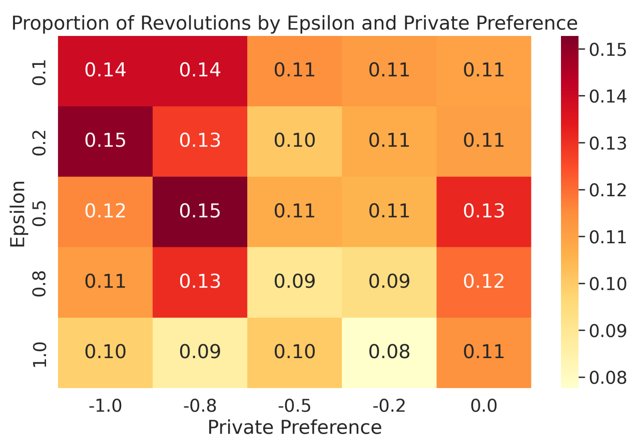

Figure 10: Heatmap of Epsilon vs Private Preference for Proportion of Each Epsilon and Private Preference Value that Ended in Successful Revolution

Let us begin by analysing successful revolutions, full equilibrium flips, where more than 95% of agents are either **Active** or **Jailed**, . The full simulation run had a total of 3278 revolutions out of 10,000 model runs. The above discussed threshold was chosen to provide enough samples to analyse. Figure 10 is a heatmap of epsilon and private preference values across all seeds and security distributions. The private preference values are the mean of a Gaussian distribution with standard deviation of 1, which as discussed in Figure 6 a value of -1.0 equates to roughly 13% regime support while 0.0 equates to roughly 50% regime support. As expected the model produces a higher proportion of full equilibrium flips, successful revolutions at lower private preferences. Nothing about this is surprising, as one would expect extremely unpopular regimes to suffer regime collapse or revolutions at a much higher rate than merely very divided regimes. Importantly, we see epsilon, the uncertainty parameter, and private preference interacting in the top left corner to produce a higher percentage of model runs that end in full blown revolutions. This is not to say that high epsilon, aka low information control polities do not suffer revolutions and this is purely a behavior of low epsilon, aka high information control polities. We can see that there is some interaction with certain private preference distributions and epsilon values that also produce revolutions at higher rates. The mechanism is a consequence of the speed of the spread between activating agents within high information control, low epsilon, regimes.

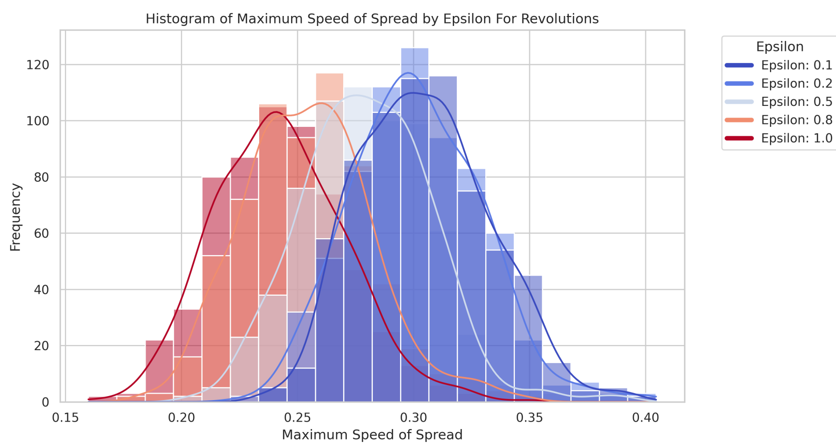

Figure 11: Speed of Spread of Activations Between Agents for Models with Successful Revolutions.

In Figure 11 we see the breakdown of the maximum value transmission speed of activations between agents as a proportion of total agents activating in each model per step. A value of 0.40 would correspond to 40% of Citizen agents going from either **Support** or **Oppose** to **Active** during their simultaneous and independent decision phase in a single step. Lower epsilon values, aka high information control polities have a higher transmission speed between agents when a cascade begins, quickly overpowering the deactivation pressure of Security forces. The action is uncoordinated and atomized in the model, but as the information value of an agent going active in a low epsilon environment carries high information content to it's surrounding agents. Even when the region for activation is more compressed than in low information-control regimes, when a cascade happens, it precedes much more quickly. Higher epsilon values progress much slower, with each level of information-controlled polity still producing cascades, but each at a slower speed than the last.

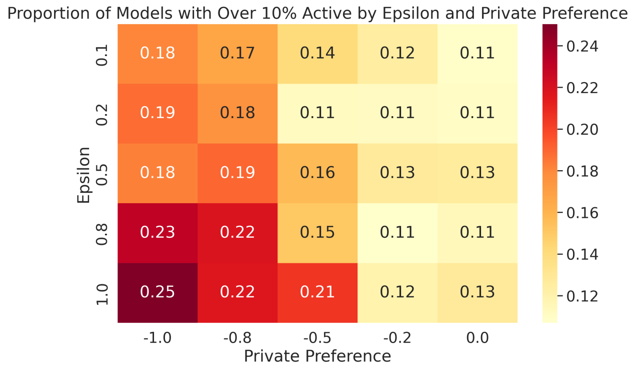

Figure 12: Heatmap of Epsilon vs Private Preference for Proportion of Each Epsilon and Private Preference Value with Over 10% of Agents Activating

Higher epsilon, aka low information control polities exhibit an important behavior. At lower private preferences, they produce more model runs where over 10% of Citizen agents activate, but stop shy of full blown revolutions. In Figure 12 the bottom right corner shows a hotspot where high epsilon low private preference model runs exhibit a high proportion of over 10% active agents per model. When compared to Figure 10 we see a proportion difference of 0.15, meaning 15% of Private Preference -1.0 Epsilon 1.0 models reach 10% active without achieving a successful revolution. The cascade process is highly abortive, stopping short of a full equilibrium flips more often than it achieves them. This is true for all of the bottom left corner values, where there is a significant proportion of model runs with low private preference and high epsilon that fail to achieve full cascades and instead fall into cyclical explosions of **Active** Citizens who are arrested, removed from the grid and put downward pressure on other agents to deactivate only to then cascade again. Other high epsilon models fall in to a scenario of consistent but low levels of activation, with a small pocket of agents remaining active while the rest of agents remain un-activated. The low epsilon, high information control polities on the other hand have very few models that exceed 10% activations that do not complete a full cascade, totalling only 4% of model runs.

Partially the ability of higher epsilon models to achieve full revolutions is a consequence of the higher number of opposed agents. Higher epsilon polities in the model have greater room for agents to find methods of dissent below the threshold that would lead to arrest. Agents do not know where this line is, and so if they have a *preference falsification* will guess. The higher epsilon models have a high level of error of where this region is, and so will sometimes in their decision phase choose to enter an **Opposed** state above the threshold value, and thus be arrested. Some though are able to guess properly where that line is, and the model is systematically designed to have more room for that state within the higher epsilon runs, representing those polities where that dissenting behavior is less clearly defined and less clearly punished.

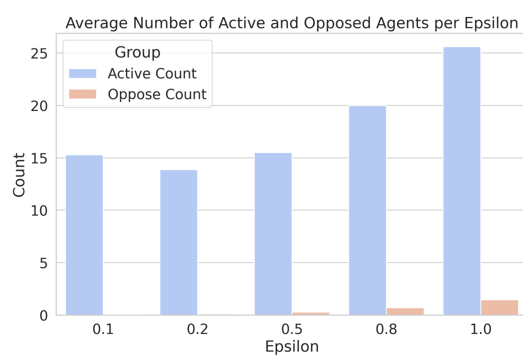

Figure 13: Average Number of Active and Opposed Agents by Epsilon Value Across All Model Runs For All Steps

Other agents see opposed agents, but apply their own error terms to calculating the value of that opposition. Thus **Oppose** is a much more frequent state in higher epsilon polities, which transfer some information to other agents at a much more frequent rate, but lower informational value than low epsilon models. On any average step, there are still few **Opposed** agents in any epsilon parameter, but they are able to spur other agents to **Oppose** or go **Active** and so high epsilon models have both higher numbers of **Active** and **Opposed** agents in any given step.

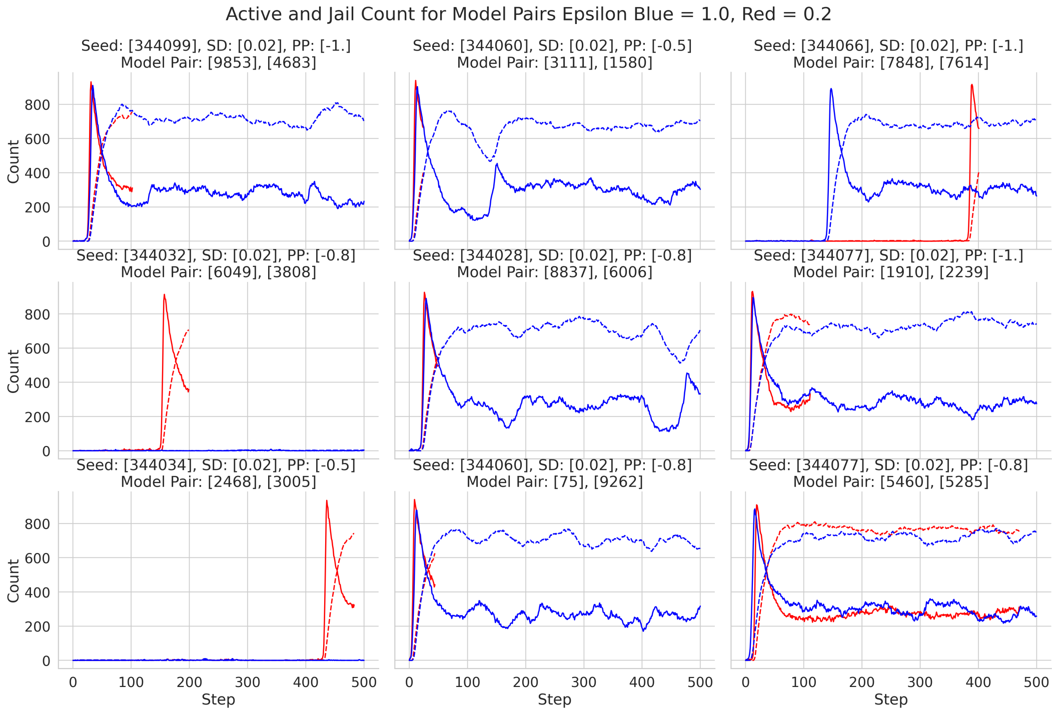

Figure 14: Facet Graph of Randomly Chosen Subset of Paired Model Runs with Same Seed, Security Density and Private Preference Values Varying only on Epsilon. Red is Epsilon 0.2 and Blue is Epsilon 1.0. Solid Lines Represent Active Count while Dashed Lines Represent Jail Counts

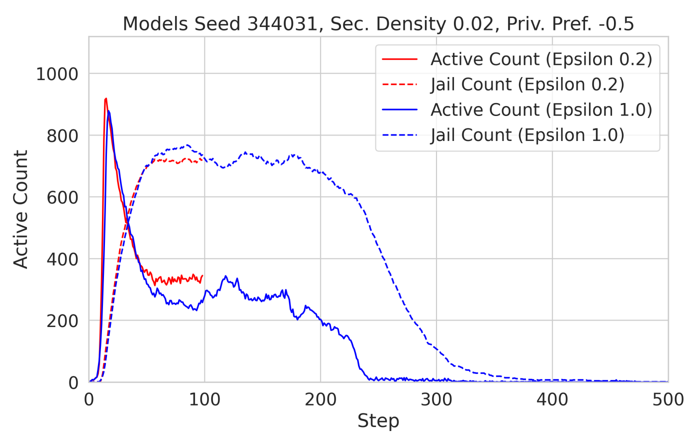

Figure 15: Model Pair 253 and 3648 Active and Jail Count for 500 Step Model Run

The highly complex nature of the model acts much like a double pendulum, where small changes in initial conditions produce completely different outcomes. To further drill down on what is happening within the model, we will pick several individual representative runs and look at agent level behavioral data to tease out the mechanisms at work. In Figure 14 model pairs were constructed of same seed, same security density and same private preference distributions varying only on epsilon where low epsilon parameters produced revolutions and high epsilon parameters did not. In some models, behavior is fairly similar, such as model pair 9853 and 4683, where a cascade started for both models at roughly the same time and preceded in roughly the same manner, but the low epsilon parameter cascaded faster than security was able to react and tamp down activations, while the high epsilon parameter was too slow to overcome the back pressure and instead maintained a moderately high level of activation for the rest of the model run. Other model pairs show a quickly unfolding cascade resulting in a full equilibrium flip for the low epsilon parameter while the high epsilon parameter never produces a cascade at all. Still others produce cascades at different times and some fall into a cyclical activation scenario with receding waves of **Active** agents producing repeating but aborted cascades as previously arrested agents are released back into the model.

Let us look more deeply at a pair of model runs shown in Figure 15 using nearly identical parameters varying only on epsilon. Model 253 and 3648 both use Seed 344031, Security Density 0.02, Private Preference -0.5 and the Global Threshold Constant T_C 2.94444. Model 253 uses epsilon 0.2 while model 3648 uses epsilon 1.0. Both models begin a cascade early in the run, and model 253 quickly cascades to over 95% of agents either **Active** or **Jailed** and reaches the stop condition of a full equilibrium flip. Model 3648 barely misses that threshold as Security is able to arrest activating agents just quicker than the cascade is able to progress, and that slightly higher back-pressure on activation causes some agents to deactivate in the higher epsilon model. This proceeds to a continuous moderately high activation level before Security is able to repress the resistance cascade fully and the model returns to a near zero baseline by step 250.

Figure 16 shows the epsilon and private preference distributions for the models in question. The two models were created using the same seed value so the private preference distribution of agents within the model is identical as seen in the top left inset. The epsilon distributions on the other hand are very dissimilar. Model 253 uses the epsilon parameter 0.2 as the standard deviation in the creation of its epsilon distribution for agents, creating a tight distribution, while model 3648 uses epsilon parameter 1.0 leading to a more dispersed distribution. Thus in model 253, agent level uncertainty about state expectations for behavior and the probability of consequences for exceeding those limits is quite clear, but not perfectly so. Model 3648 on the other hand has high levels of uncertainty, but even so a significant proportion of agents still have a fairly clear understanding of behavioral limits.

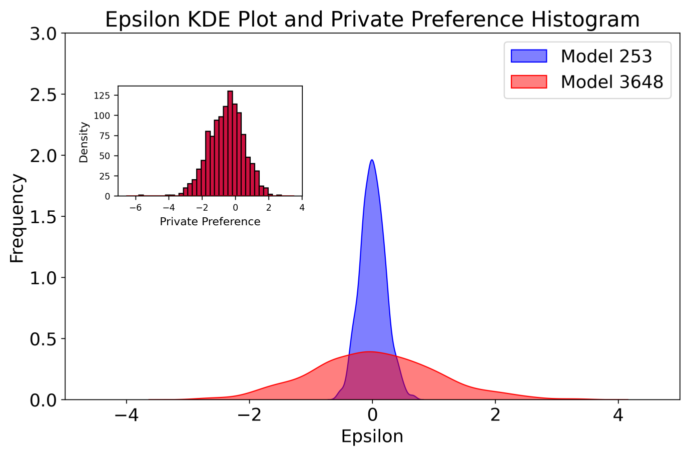

Figure 16: KDE Plot of Epsilon Distribution for Models 253 and 3648. 253 is Epsilon 0.2 and 3648 is Epsilon 1.0. Both models use Seed 344031 and with different Epsilon parameters create different Epsilon distributions but identical Private Preference Distributions

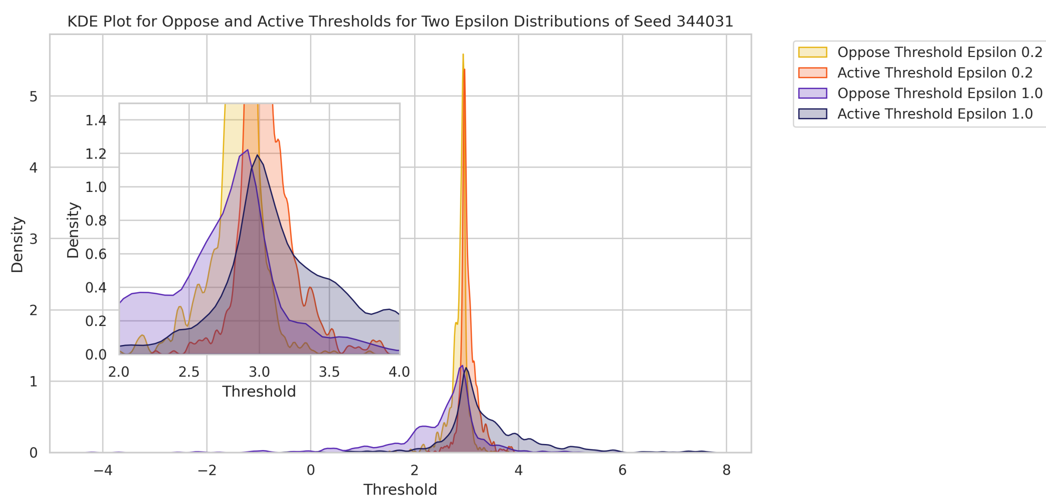

Figure 17: KDE plot of Oppose and Active Thresholds for Models 253 and 3648 with Inset to More Clearly See the Overlap at Center

Using agent level uncertainty parameters epsilon ϵ_i the model constructs the agent level **Oppose** and **Active** thresholds. Again we see the effect of the uncertainty parameter epsilon on the threshold distributions, but the effect is even more pronounced as seen in Figure 17. The individually assigned epsilon value for each agent ϵ_i is used as the standard deviation to draw two values from a Gaussian distribution centered on the model's global threshold constant T_C and the smaller value is used as the agents **Oppose** threshold, while the higher value is used as the agents **Active** threshold. The effect this creates is that agents in the low uncertainty, low epsilon parameter model have a tightly distributed set of thresholds. This reflects two theoretical concepts, regime clarity on the location of "the red line" or what is and is not acceptable behavior, and also thus a smaller societal window for acts of dissension short of open resistance. The high epsilon, or high uncertainty society lacks clear information on what is and is not acceptable behavior, and thus each individuals placement of what they personally estimate to be where they could dissent without consequence and what might constitute unacceptable dissension is much wider. Agents in this high uncertainty model are more likely to enter an **Oppose** state, but also significantly more likely to be arrested within that state than in model 253. This transfers information to other agents about regime preference more broadly, but in an error prone way.

Figure 18: Scatterplot of Active and Oppose Activation Levels by Step and Individual Agent for Model 3648 with an Overlay of Avg Arrest Probability

Figure 18 is a scatterplot of the individual activation levels of each 1120 Citizen agents in the model, with the blue and red lines showing average activation levels and the green line showing average arrest probability. Activation levels are calculated by passing the agents individual private preference and to the ratio of **Active** and **Opposed** agents in vision multiplied by their perception of regime support L_i, subtracting their individually calculated threshold values and passing that value though a sigmoid where probability of arrest is subtracted from that sigmoid output. In Figure 18 we can see a cascade beginning by step 6 of model 3648. The cascade propagates quickly but at its peak, only 85% of agents reach an activation level exceeding 0.75. Even at the height of positive pressure, there is still a wide distribution in activation levels owing to the uncertainty parameter epsilon's interactions with the global threshold constant and a muted perception moderator for high epsilon agents.

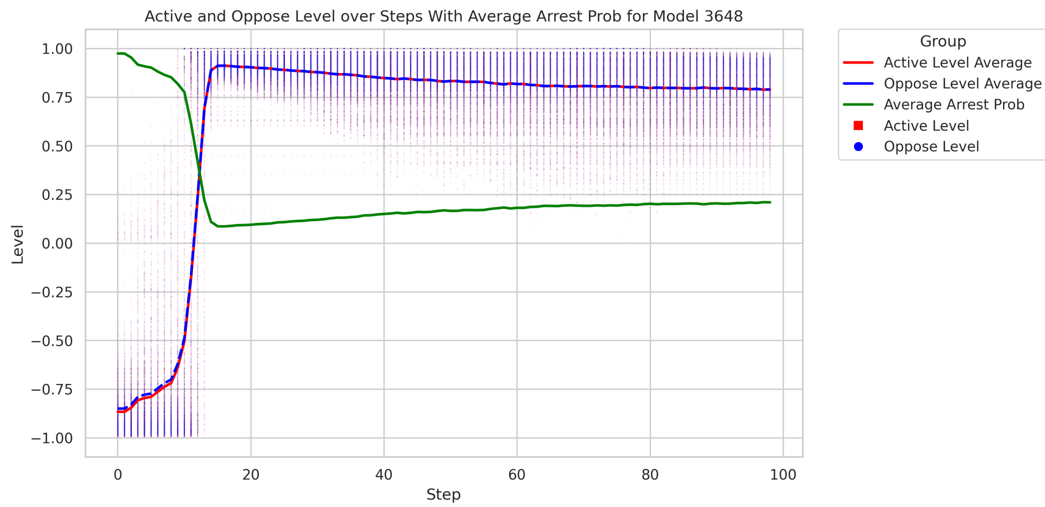

Figure 19: Scatterplot of Active and Oppose Activation Levels by Step and Individual Agent for Model 253 with an Overlay of Avg Arrest Probability

Looking at the per step activation levels of individual agents on a per step basis in model 253 in Figure 19, there is a high level of coherence between agents as the cascade progresses. The tight threshold distribution of low epsilon parameters means that agents are all have a fairly good understanding of where "the red line" exists for expected behavior, and so activating agents move en masse across that threshold into a positive probability range for activation. The cascade begins where at model initialization, a small number of agents have a positive probability of activation, and the clustering of those agents create the chain cascade effect pushing other agents to activate, further reducing the probability of arrest which quickly results in a model wide high probability of activation. Some holdout agents with high regime private preference are hard to push into activation but as the cascade reaches it's zenith, security is unable to arrest agents quickly enough to stop the cascade from slowly pushing those last holdouts to activate and the model reaches the stop condition at Step 99. At the cascades's height, over 95% of Citizen agents had an acitvation level of over 0.75. The high certainty of state expectations for behavior, and thus the increased uncertainty of the true opinions of the neighbors in each agent's vicinity increases the strength of activations as agents see other agents activate and thus activate themselves.

5 Conclusion

The Resistance Cascade model is an attempt to understand how individual level uncertainty affects resistance movements. The cascade mechanics of resistance movements are shown, at least theoretically through the model, to be affected by the strictness of the information environment within a regime, where more controlled information environments suffer fewer resistance events, but those events that do happen are more likely to progress to a full revolution. Higher uncertainty, aka less information-controlled societies suffer more frequent resistance events, but are less likely to suffer full blown revolutions. This is a consequence of a slower inter-agent spread of active resistance in lower information-control environments because of the increased range and willingness of agents to participate in milder forms of dissension transmitting information about regime popularity through the society. The lower level of uncertainty about each individual Citizen agent's local neighbor's regime opinions, gained as a consequence of higher levels of uncertainty about state limitations on public behavior, combine to create the more frequent, slower spreading, and smaller resistance cascades we see in the model. This supports the empirical findings that less authoritarian countries with lower levels of information control experience more frequent protest events, as recorded in the ICEWS data set and compared via various metrics of regime type contained in the V-Dem classifications.

The model represents a snapshot in time at the moment a cascade process begins, or fails to begin. While it is able to replicate some of the behavior we see in differing regime types, it is not an examination of the longer term evolving social context of an authoritarian society and whether those changes may affect the probability of outbreaks of resistance. There are many factors to consider when examining longer term socio-political changes that may increase or decrease the possibility of resistance cascades. Whether the announcement of Gorbachev's "Sinatra Doctrine" where the Soviet Union promised not to interfere in the politics of Eastern European satellite countries could be potentially approximated as a changing global threshold value T_C, or whether the Arab Spring could possibly relate to a shift in general *private preference* or local perception L_i, the model design is agnostic. It represent a snapshot of time given a set of conditions and does not contend with the admittedly important evolving context of longer time frames. The model does manage to capture fairly well the role of uncertainty in the context of resistance cascades and its effects on how those cascades evolve in short time-frames with fixed conditions.

The model is limited in its treatment of agent coordination and movement. An important aspect of resistance or protest movements is the use of public spaces to gather and express discontent. The model has no distinction between one grid square and another, and so there is no symbolic public gathering place that combines the aspects of agent level coordination and movement. Implementing some method of coordinated movement such as a gravity model, where agents are able to independently decide on a directional relevant movement choice based either purely on their own limited environmental information, or even inter-agent communication, would more closely approximate realistic resistance behavior. This would be an important addition to the model, as concentrations of active Citizen agents, or agents with low *private preferences* is core to the creation of cascade events.

The model also does not deal with certain behavioral patterns, such as the day/night cycle, or partisanship. There is reason to believe that the addition of a day/night cycle might alter cascade behavior. Some, though not all, protest events are temporary gatherings where the location is abandoned between protests, while others keep a permanent presence. The day/night cycle is still an important factor in both types. Partisanship has been shown to be important in the development of protest movements (Aytaç et al. (2018)) and the public's reaction to those movements as well. Each of these represent areas of model expansion that may further increase its explanatory power.

The ability to compare model outputs to more granular empirical data on resistance events and revolutions might allow for further fine tuning of the model's behavior as well. The addition of accurate revolution counts and more fine-grained person counts for those events on a day by day or even hour by hour level with position data could be used to increase the accuracy and improve model behavior. Future research using satellite imagery and computational approaches to record locations and numbers of people, potentially with the ability to distinguish security from non-security could be immensely useful information in a re-design of the model's dynamics. This would be no easy task, and in many instances could be polluted with counts of plain-clothed security officers. However, tt is hypothetically possible with computational image analysis techniques and potentially very instructive.

6 Bibliography

References

Hellbeck, J. (1996). Fashioning the stalinist soul: The diary of stepan podlubnyi (1931-1939) [Publisher: Franz Steiner Verlag]. *Jahrbücher für Geschichte Osteuropas*, *44*(3), 344–373. Retrieved April 21, 2023, from https://www.jstor.org/stable/41052987

Kuran, T. (1991). Now out of never: The element of surprise in the east european revolution of 1989. *World Politics*, *44*(1), 7–48. https://doi.org/10.2307/2010422

Kuran, T. (1989). Sparks and prairie fires: A theory of unanticipated political revolution [Publisher: Springer]. *Public Choice*, *61*(1), 41–74. Retrieved February 7, 2023, from https://www.jstor.org/stable/30025019

Kuran, T. (1995). The inevitability of future revolutionary surprises. *American Journal of Sociology*, *100*(6), 1528–1551. https://doi.org/10.1086/230671

Acknowledgements
I want to thank Dr. Jean Clipperton for her significant assistance in the design of the model and her multiple rounds of readings and edits on this paper. I would also like to thank Thiyaghesson Poongundranar and Loizos Bitsikokos for their consultations on the design of the Citizen activation function. Finally, my greatest appreciation is reserved for my fiance Kara Andersen and the infinite patience she maintained during the months spent finishing this project.

Lohmann, S. (1994). The dynamics of informational cascades: The monday demonstrations in leipzig, east germany, 1989–91 [Publisher: Cambridge University Press]. *World Politics*, *47*(1), 42–101. https://doi.org/10.2307/2950679

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states* [Google-Books-ID: vYO6sDvjvcgC]. Harvard University Press.

Havel, V., & Keane, J. (2016, September 16). *The power of the powerless: Citizens against the state in central eastern europe* [Google-Books-ID: sscYDQAAQBAJ]. Routledge.

Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., & Starz, J. (2018). ICEWS weekly event data. https://doi.org/10.7910/DVN/QI2T9A

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Alizada, N., Altman, D., Bernhard, M., Cornell, A., Fish, M. S., Gastaldi, L., Gjerløw, H., Glynn, A., Grahn, S., Hicken, A., Hindle, G., Ilchenko, N., Kinzelbach, K., Krusell, J., … Ziblatt, D. (2022). V-dem country-year dataset v12. https://doi.org/https://doi.org/10.23696/vdemds22

Bank, T. W. (2023). *World bank open data* [World bank open data]. Retrieved April 20, 2023, from https://data.worldbank.org

Dahl, R. A. (1971). *Polyarchy: Participation and opposition* [Google-Books-ID: LRG7ngEACAAJ]. Yale University Press.

Hlavac, M. (2022). *Stargazer: Well-formatted regression and summary statistics tables*. Social Policy Institute. https://CRAN.R-project.org/package=stargazer

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., Schulze, J., Weise, H., & Schwarz, N. (2013). Describing human decisions in agent-based models – ODD + d, an extension of the ODD protocol. *Environmental Modelling & Software*, *48*, 37–48. https://doi.org/10.1016/j.envsoft.2013.06.003

Pearlman, W. (2017). *We crossed a bridge and it trembled: Voices from syria*. Custom House.

Lee, T. (2009). The armed forces and transitions from authoritarian rule: Explaining the role of the military in 1986 philippines and 1998 indonesia. *Comparative Political Studies*, *42*(5), 640–669. https://doi.org/10.1177/0010414008328640

Epstein, J. M. (2002). Modeling civil violence: An agent-based computational approach [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *99*, 7243–7250. https://doi.org/10.1073/pnas.092080199

Kazil, J., Masad, D., & Crooks, A. (2020). Utilizing python for agent-based modeling: The mesa framework. In R. Thomson, H. Bisgin, C. Dancy, A. Hyder, & M. Hussain (Eds.), *Social, cultural, and behavioral modeling* (pp. 308–317). Springer International Publishing.

Aytaç, S. E., Schiumerini, L., & Stokes, S. (2018). Why do people join backlash protests? lessons from turkey. *Journal of Conflict Resolution*, *62*(6), 1205–1228. https://doi.org/10.1177/0022002716686828

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. https://www.jstatsoft.org/v40/i03/

7 Appendix

The original ICEWS data set is split by year, with evolving standards across that time-frame. Combining the 26 separated .tab files across the full ICEWS data set creates a 1,974,817x20 matrix comprised of coded news stories from 1995 to 2021 from across the world. It is a continually updated data set of worldwide events pulled from news stories using an adapted version of the Conflict and Mediation Event Observations (CAMEO) coding system which classifies stories into various categories ("Make Public Statement", "Demand", "Protest", "Engage in unconventional mass violence" etc.). Using the CAMEO coding system, the event information was condensed to only protest coded events, then Lubridate Grolemund and Wickham, 2011 was used to create a unified date format to extract the year information and apply that for each event. The data was then further condensed to a count of events per country per year.

Figure 20 shows that small *epsilon* values have much tighter coherence in their changing perception of regime support moderator L_i as the number of **Active** agents increases while higher *epsilon* values have a much wider distribution.

Table 7: Model 3: Polyarchy Autocracy no Covariates Log Transformed Using Only Dropped Econ Covariates

	Dependent variable:
	log_per_mil
v2x_polyarchy_100	0.031***
	(0.006)
Constant	0.668***
	(0.141)
Observations	230
R^2	0.116
Adjusted R^2	0.112
Residual Std. Error	1.044 (df = 228)
F Statistic	29.774*** (df = 1; 228)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Model 4: Polyarchy All Countries No Covariates Log Transformed

	Dependent variable:
	log_per_mil
v2x_polyarchy_100	0.004***
	(0.001)
Constant	1.206***
	(0.034)
Observations	4,648
R^2	0.009
Adjusted R^2	0.009
Residual Std. Error	1.051 (df = 4646)
F Statistic	41.430*** (df = 1; 4646)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 9: Model: v2x civlib Index of Polyarchy Index < 0.5 Country Years on Protest Per Million Inhabitant per Year Log Transformed

	Dependent variable:
	log_per_mil
v2x_civlib_100	0.014***
	(0.001)
Constant	0.663***
	(0.058)
Observations	2,226
R^2	0.073
Adjusted R^2	0.073
Residual Std. Error	1.088 (df = 2224)
F Statistic	176.217*** (df = 1; 2224)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 10: Model: v2x clpol Index of Polyarchy Index < 0.5 Country Years on Protest Per Million Inhabitant per Year Log Transformed

	Dependent variable:
	log_per_mil
v2x_clpol_100	0.012***
	(0.001)
Constant	0.839***
	(0.048)
Observations	2,226
R^2	0.066
Adjusted R^2	0.066
Residual Std. Error	1.092 (df = 2224)
F Statistic	158.111*** (df = 1; 2224)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11: Model: v2x clpriv Index of Polyarchy Index < 0.5 Country Years on Protest Per Million Inhabitant per Year Log Transformed

	Dependent variable:
	log_per_mil
v2x_clpriv_100	0.010***
	(0.001)
Constant	0.827***
	(0.057)
Observations	2,226
R^2	0.046
Adjusted R^2	0.046
Residual Std. Error	1.104 (df = 2224)
F Statistic	107.360*** (df = 1; 2224)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 12: Model: v2x frassoc thick Index of Polyarchy Index < 0.5 Country Years on Protest Per Million Inhabitant per Year Log Transformed

	Dependent variable:
	log_per_mil
v2x_frassoc_thick_100	0.012***
	(0.001)
Constant	0.848***
	(0.045)
Observations	2,226
R^2	0.073
Adjusted R^2	0.072
Residual Std. Error	1.089 (df = 2224)
F Statistic	174.100*** (df = 1; 2224)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 13: Model: v2x freexp Index of Polyarchy Index < 0.5 Country Years on Protest Per Million Inhabitant per Year Log Transformed

	Dependent variable:
	log_per_mil
v2x_freexp_100	0.010***
	(0.001)
Constant	0.913***
	(0.048)
Observations	2,226
R^2	0.050
Adjusted R^2	0.050
Residual Std. Error	1.102 (df = 2224)
F Statistic	117.195*** (df = 1; 2224)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 14: Model: v2xcl disc Index of Polyarchy Index < 0.5 Country Years on Protest Per Million Inhabitant per Year Log Transformed

	Dependent variable:
	log_per_mil
v2xcl_disc_100	0.011***
	(0.001)
Constant	0.853***
	(0.050)
Observations	2,226
R^2	0.057
Adjusted R^2	0.057
Residual Std. Error	1.098 (df = 2224)
F Statistic	134.510*** (df = 1; 2224)
Note:	*p<0.1; **p<0.05; ***p<0.01

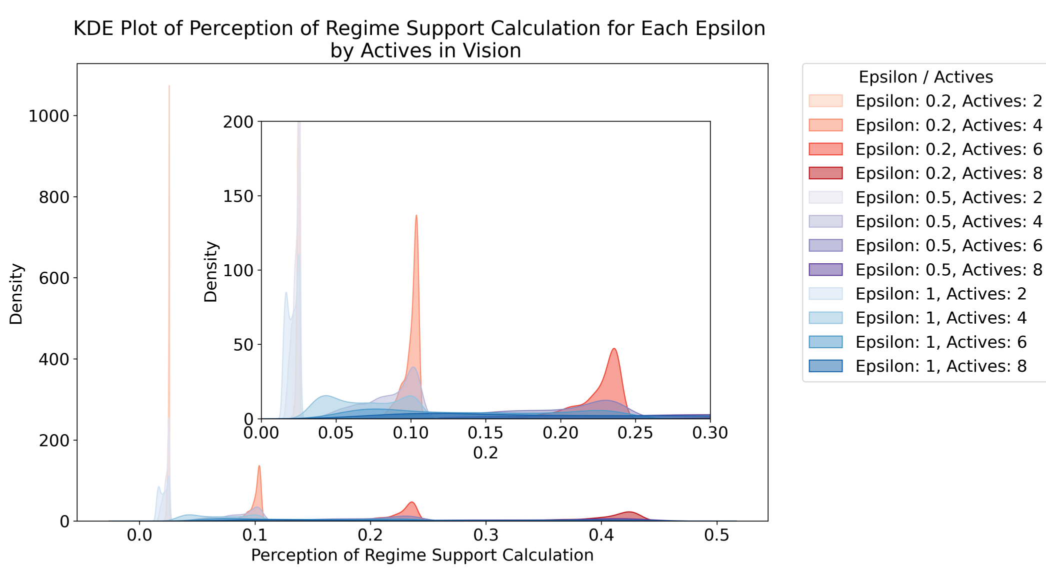

Figure 20: Distribution of Perception of Regime Support L_i multiplied by $R = \frac{A+O}{S}$ where Citizens of state Support is (157 - Actives) split by epsilon values and varying number of Active agents