Journal Title
XX(X):1–13
©The Author(s) 2021
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

If There is No Help for It They Will Fight Hard: How External Punishers Affect the Principal Agent Problems for Repressive Autocrats

Timothy Liptrot[1]

Abstract

This article explores the problem of punishing the agents of autocrats for their crimes. Pro-democracy actors cannot punish the agents of a surviving regime, but can punish agents after a regime collapse or a defection. The expectation of punishment can perversely incentivize more loyalty and greater repressive violence. [Results here]

Keywords

Class file, LaTeX 2$_\varepsilon$, *SAGE Publications*

> Throw your soldiers into positions whence there is no escape, and they will prefer death to flight. If they will face death, there is nothing they may not achieve. Officers and men alike will put forth their uttermost strength. Soldiers in desperate straits lose the sense of fear. If there is no place of refuge, they will stand firm. If they are in the heart of a hostile country, they will show a stubborn front. If there is no help for it, they will fight hard. *Thus, without waiting to be marshaled, the soldiers will be constantly on the alert, and without waiting to be asked, they will do your will; without restrictions, they will be faithful; without giving orders, they can be trusted.* - Sun Tzu

Introduction

In 2012 Anwar Raslan was an operative of the Syrian secret police. He directed an office which imprisoned and tortured dissidents. In 2011, the regime responded to popular opposition by massively escalating state-citizen violence through a series of massacres. In one massacre 100 people from Raslan's home village were killed, many of them children. Raslan learned that the regime had launched false-flag "terrorist" attacks against its own side to motivate greater violence. These two events caused Raslan to defect. Knowing his relatives would be killed if he defected, he waited for an opportunity then smuggled

[1] Sunrise Setting Ltd, UK
[2] SAGE Publications Ltd, UK

Corresponding author:

Timothy Liptrot, Georgetown University Georgetown, DC, 20007, UK.

Email: tl790@georgetown.edu

his immediate family into a rebel region and then into self-imposed exile.

Raslan was initially welcomed by the Syrian opposition, who hoped that Raslan's defection would trigger a cascade. But these hopes slowly faded as the war dragged on with no defection cascade. In 2019 Raslan and another defector were arrested in Germany. The German government has since tried and convicted them for crimes against humanity. Advocates of the case have argued that crimes should be punished unconditionally, regardless of their effects on the conflict.

These trials raise the question of how the punishment of individual repressive agents changes the outcomes of the politics of repression. In this paper, I relates the most advanced current models of autocratic repression to assess how external punishers affect the decision to violate human rights and the degree of effort that agents supply.

Recent research has shown that autocracies are not unitary actors, contrary. National policies are too complicated and costly for any single actor to carry out. As a result the individual responsible for a policy (the principal) must contract others with specialized skills and information (the agents). Every regime is comprised of a network of principal-agent relationships. From autocrat to foot soldier, superiors contract subordinates to perform services. However, each agent has her own motivations, interests and information distinct form their principal. Moral hazard is the problem that agents know more about themselves and their issue area. Agents can exploit this informational advantage. Agents may also act without their principals knowledge, a problem called hidden action. This simple model has been used to explain outcomes from election fraud (Little) to civilian killings to purges.

I focus on a critical problem facing external punishers. When the regime holds power it is, by design, impossible to apply strong selective punishments outside of assassinations and minor sanctions. An external punisher can only reach those agents who defect (escape) or wait for the regime to collapse. The folk wisdom is that punishing defectors discourages defection, and punishment after regime change discourages the shirking and mutinies that allow democratization.

I model when external punishment influences autocrats to increase or moderate their repression and influences agents to give greater or lesser effort. I distinguish between several situations. In the first the dictator commands and observes a repressive act prior to the first defection opportunity, while the agent chooses to comply, and later to defect or give a some degree of repressive effort. In the second the dictator only observes repressive effort after the first defection opportunity. In the third the agent provides some effort which is visible to both observers, and a second invisible effort. I model two types of external punishers, an unconditional punisher and a consequentialist punisher.

I find <add findings>.

This articles proceeds as follows. The second section reviews the existing literature on principal agent problems in autocratic repression. The third section builds three formal models and describes selected equilibria. The third section explores three case studies which illustrate the applicability of the models.

Literature Review

Repression: What, when and why?

The purpose of this essay is to understand how outsiders can influence the level of repression that autocrats demand and the level actually enacted. Any plausible answer to that question must build on a firm grasp of repression from the inside. In this section I first ask why incumbent leaders command repression, and then ask why their agents comply with such commands.

Most leaders have substantive policy preferences, but leaders must remain in power to enact them (Ames, 2020). All authoritarian regimes face two existential challenges (Svolik, 2012). The regime must counter challenges from the society they rule, in the form of protests, revolutionary movements and elections. Deterring challenges is too complex for the ruler alone, so leaders rely on allies such as party bosses, traditional leaders, and generals for assistance. But those allies create a new threat. The regime must divide the benefits of ruling and manage rivalry within itself, to prevent coups and motivate effort. These two imperative shape political behavior across authoritarian politics, from repression to policy selection to power sharing (Svolik, 2012). Most studies argue that repression is rational response to threats with the goal of keeping power (DeMeritt, 2016).

Repression is violence or coercion for the purpose of shaping subjects political behavior. Repression includes restrictions on political freedoms like bans on political parties and censorship, or the selective denial of state benefits (Albertus and Svolik). It also includes overt violence against the opposition, through protest policing, imprisonment, torture and assassination. Repression does not include coercion to deter non-opposition violence, like theft or non-coercive responses to challenges like cooptation, toleration and meeting demands.

Leaders sometimes do not explicitly "command" repression, while letting it be known that repression will be rewarded. Hassan and O'Mealia found that the government provided promotions to officials whose districts saw anti-opposition violence, while officials whose districts saw opposition-instigated violence were more likely to be fired.

Public displays of dissent are dangerous for leaders because dissent can spread. The costs of dissent decrease as more citizens join as crowds bring anonymity and strength in numbers. A fallen regime can no longer punish rebels, so deterrence loses its power once actors expect a regime failure (Kuran, 1992). Repression punishes early joiners, degrades opposition organization and signals the regimes strength and resolve. Repression is also cheaper than accommodation, because the regime need only deploy its existing retinue of violence-specialists.

Decades of quantitative research finds that dissent incentivizes repression. Rises in opposition behavior are associated with a rise in repression (deMeritt, 2016; Davenport, 2012). The result is so consistent that it is sometimes called "the law of coercive responsiveness". Consider the iconic image of repression as a solider beating protesters with a truncheon. The regime targets protesters because they are resisting its rule and challenging its legitimacy; the regime has many ways to display its strength, but only attacking protesters will deter future challenges. This fact is obvious but highly important.

An extensive literature has studied the effects of regime types on repressive forms. There appears to be a threshold of democracy after which repression is too costly for incumbents (De Mesquita et al, 2005). Meanwhile autocrats struggle to accomodate the opposition because they have no credible commitment devices (Acemoglu and Robinson). Once the protestors go home, there is nothing to prevent the autocrat from betraying past promises. Opposition leaders know this, making accommodation difficult relative to repression.

Repression is highest in states in the middle of the polity democracy index (Davenport, 2012). In the most autocratic states the threat of violence is so strong that citizens stay home and less repression is observed.

Does repression achieve its goal of stopping dissent and prolonging leader tenure? Studying this topic is difficult due to simultaneity; dissent affects repression at the same time that repression affects dissent. The findings on repression and opposition are mixed, with some studies finding repression increases dissent and some finding the reverse. A study in Guatemala found that repression decreased dissent only if the repressive actors successfully targeted opposition organizers, rather than activists.

However, repression does extend the leaders time in office. A study by Escriba-Folch which accounts for simltaneity finds that both political terror and civil liberties restrictions decrease the probability of exit. The only exception to this rule is that political terror does not protect leaders from violent exits. One explanation is that leaders who do violence against the opposition are more likely to be killed if they fall, leading to a null net affect on violent exits. In any case, repression is an effective tool for staying in office.

Why do agents comply?

The previous section explained why rules want repression, but not if they get what they want. In 1989 Romanian dictator Nicolae Ceausescu called for violence against an opposition protest. But the Romanian military mutinied after the order, and Ceausescu was arrested, tried and executed just days later. An order given is no guarantee of an order carried out.

Leaders cannot personally carry out repression, and must rely on a group of agents to enact their policies. We refer to these problems as principal-agent problems because the principal (leader) contracts to the agent (supporters) to carry out tasks in exchange for compensation or policy control. These agents might be generals, police chiefs, district bosses, traditional leaders or party members. This process of delegation is imperfect because the leader may select the wrong agents for the job (hidden information) and those agents may choose to act against the leaders intent (hidden action). Hidden information is dangerous when leaders choose agents who claim to prefer the incumbent but secretely prefer the incumbents fall, or who secretly are unwilling to repress. Hidden action can be dangerous in several ways: the agent may repress more or less than the desired amount, they may mislead the leatder about the severity of threats (little, kompromat), they may simply refuse orders, and they may escape without carrying orders out. Agents may even mutiny and overthrow the leader themselves. Indeed, such internal coups are more common than revolutionary overthrows of dictators (Svolik, 2012).

All governance problems involve delegation, but repression has uniquely challenging characteristics. There is no external enforcer to ensure that agreements are kept between the autocrat and his agents. This means the autocrat may reneg on promises to reward hardworking agents (Gallagher and Hanson). Repression provides many opportunities for agents to shirk undetectably. Christopher Sullivan use Guatemalan police archives to show that repression was effective at reducing opposition only if opposition leaders were identified, while targeting footsoldiers was counterproductive (2016). Shirking agents may comply with requests for documentable acts of violence, while simply targeting the wrong groups. In less violent endeavors citizen feedback is vital to policing shirking or abusive agents, but in repression citizens are usually unwilling or unable to provide that information. Finally, political terror is unpleasant for most agents, so the incentives to shirk, escape and mutiny are strong.

If the autocrat loses power they also lose the ability to punish. The probability that the autocrat will punish shirkers is weakly lower than the probability the autocrat will survive. Few agents will give their lives and health for an incumbent they expect to fail. In many PA models performance is related to the probability of leader survival (Fruge, 2019; Little, 2014; Tyson, 2019). Fruge (2019) argues that a request for repression is itself a signal of weakness (depending on the equilibrium played) and that autocrats may respond by requesting repression during peacetime.

Expectations are doubly important to an autocrats survival. If everyone expects the incumbent will win her supporters work harder and his opponents are more scared (Dragu and Lupu, 2018). The resulting herd dynamics explain rapid and contagious collapses such as the 1989-91 collapse wave in the Warsaw Pact states. Autocrats respond by closely controlling public signals of weakness, such as elite defections, admissions of error or opposition protests.

Often leaders do not explicitly command each act of repression, either due to administrative costs or desire for secrecy. They instead simply reward agents that enact repression or deliver obedience. Hassan and O'Mealia found that the government provided promotions to officials whose districts saw anti-opposition violence, while officials whose districts saw opposition-instigated violence were more likely to be fired.

These decentralized orders actually create a unique control problem. Andrew Little showed that the pervasive and obvious fraud in autocratic elections occurs because leaders reward district chiefs that deliver votes. District chiefs attempt to maximize their vote count conditional on the autocrat winning. When agents expect a victory they supply generous fraud to seem more popular and competent, but when they expect defeat they supply no fraud (exactly when the leader most desires fraud).

Despite this explosion in modeling of internal regime dynamics, few studies have asked how outsiders can influence repression. Foreign military interventions in a repressive state do affect civilian killings (DeMeritt, 2014). Supportive military interventions decrease the ordering of civilian killings and oppositional interventions limit the enforcement of such orders. However, deMeritt only examines cases in which outsiders intervene directly and provide immediate targeted punishments to the atuocrats agents. The effectiveness of such punishment is unsurprising and does not answer the thorny question of how to influence killings when outsiders cannot punish regiem insiders.

A brief history of defection

Talk about exit costs and early democracy

talk about current defectors. How often and what motivates them.

Note the mutineer

A Simple Model of Atrocities

In this section I Develop three simple models to show how externally punishing repressive agents affects repression.. In the first model the autocrat compels some act which is instantly observed and punished by the autocrat, before the agent can defect. In the second model both the autocrat and the external punisher only observe the agents action after the crisis is resolved. In the third model the external punisher observes some unit of effort which the autocrat cannot observe. From each model I derive predictoins about how punishment affects both the degree of repression ordered and the level executed

Model 1: Public Acts

In this model, the autocrat can choose to command some initial repressive act γ. The incumbent perfectly observes the act γ and the agent has no chance to defect before committing γ. A commitment to punish

Model Specification

The model describes the interaction of three agents; an incumbent I (pronoun: she), a group of agents A (pronoun: he) and an external punisher J (pronoun: he). The incumbent wishes to remain in power. The agents wish to avoid punishments and supply less repressive effort. The external punisher wishes to punish actors who do certain acts. However, the punisher cannot punish actors who remain inside a functioning regime.

The game proceeds in the following order. First, the autocrat selects some act γ from a set of acts Γ and commands their agents to commit it. The agents then commit the act or refuse. If they refuse, the autocrat observes the refusal immediately and applies some punishment p_i. This reflects the autocrats limited surveillance resources; She cannot check all the work of her agents, because administering nation-wide repression is too sophisticated for one human. Instead she can compel and monitor a small number of discrete tasks. γ might refer to some a signature on an order for atrocities, the signing of death warrants, or a public endorsement of some crime, all of which are easily commanded and observed.

The model also assumes that the agent cannot defect before this punishment. In a consolidated autocracy the leader may simply call a meeting with no announced agenda, then demand γ as a fait accompli. But an autocrat need not such secrecy. Defectors from consolidated autocracies must bring their immediate family with them into exile, to avoid indirect punishment. NKVD officer Genrikh Lyushkov attempted to smuggle his wife and child into Poland before crossing the Soviet-Manchukuo border. Unfortunately most of his family was captured, tortured and executed. Iranian Air Force Colonel Jawad Hussein escaped in 1981 with his wife having already sent his three children to school in the US. Raslan was able to extract hsi immediate family only with the help of rebel forces in the capital suburbs. Because defection requires careful planning and fortuitous opportunities, the autocrat need only observe compliance before the next opportunity to smuggle the agents family into exile.

In the next round the agent chooses some repressive effort $x \in \mathbb{R}_+$ to prevent regime change, with cost $c(x)$. No amount of effort makes the regime perfectly safe constrain, $c(1 - q - f(\gamma) = \infty$. Effort has diminishing marginal returns, such that $c'(x) > 0$ and $c''(x) > 0$. If the agent defects $x = 0$. I do not include a defection move because any defecting agent would exert 0 effort, which has the same effect on all other actors in the model.

Nature then determines if the regime survives or not.. The regime survives with probability

$$Pr(Survival) = q + r(x) + f(\gamma)$$

where $f(\gamma)$ is some effect of the act γ on regime survival, which may be positive or negative.

Finally, if the regime falls the external punisher decides to punish the agent or not. The punisher observes γ and the cost of punishment c_J but does not observe the effort x.

Table 1. Timeline

t=0	Incumbent selects act $gamma$
t=1	The agents comply with $gamma$ or are punished
t=2	The agents select some effort x
t=3	The regime survives or fails based on strength and effort
t=4	Punisher J chooses to punish if and only if defection or regime failure

Payoffs

If the agent defects or the regime collapses, the puncher's utilty is

$$U_j = P_e(V(\gamma) - c_e)$$

where P_e is 1 if J punishes and 0 else. $V(\gamma)$ is some utility that the punisher receives from punishing agents that commit act γ and c_e is the cost of punishing. Without defection or collapse the punisher's utility is irrelevant.

If the agent refuses γ the receive some punishment $p_I > p_e$, reflecting that the incumbent can inflict more severe punishments than the external punisher. If the agent defects they receive an outside option $\bar{u}$ and pays p_e if punished. When the leader remains in power the agent receives some utility $\beta_{rem} \geq 0$. If the regime fails the agent takes their outside option and faces the threat of punishment. The agent's utilities are

$$U_A = \begin{cases} U_A(R) = -P_I \\ U_A(\gamma,\ D, x) = (q + x + f(\gamma))(\beta) + \\ (1 - q - x - f(\gamma)(\bar{u} - p_e) - c(x) \end{cases} \tag{1}$$

For simplicity, the incumbent receives utility 1 if the regime survives and 0 if it fails. Because this model focuses on the external punishment, we neglect any costs of repression or monitoring for the incumbent.

$$U_I = q + x(\gamma) + f(\gamma) \tag{2}$$

Results

I solve by backward induction, starting with the external punisher. If the incumbent falls, they punish when $V(\gamma) > c_e$.

Proceeding backward, the agent solves

$$\max_{x>0}(q + x + f(\gamma))(\beta_r em) + (1 - q - x - f(\gamma)(\bar{u} - p_e) - c(x)$$

. He selects some $x*$ such that $c'(x*) = \beta_{rem} + P_e(\gamma) - \bar{u}$. Because $c''(x) > 0$, x^* is greater when $P_e(\gamma) = 1$. Label X_H is the effort during punishment and X_L is the optimal effort when punishment does not occur.

If γ is forbidden, then the agent complies if

$$-p_I < (q+x_H+f(\gamma))(\beta_r em)+(1-q-x_H-f(\gamma)(\bar{u}-p_e)-c(x)$$

Any forbidden act gamma is less appealing both because the agent anticipates higher effort and the possibility of punishment. When the agent is confident the regime will survive, so q is large, this constraint is weak because the effort differential is small and external punishment is unlikely. This constraint never binds for any q or γ if $p_I > \bar{u} - p_e - c(x_H)$.

Knowing the effects on the other players strategies, the incumbent commands γ as long as $f(\gamma) + (X_H - X_L) > 0$ and the agent will comply. Assuming compliance, the incumbent will always command γ if its direct effect on survival is positive or negative and smaller than the effort gain. If the agent will comply, the decision to externally punish agents always encourages the dictator to command crime.

This is bad news for external punishers. The commitment to punish actually increases the amount of repression demanded and supplied. If $f(\gamma) < -(X_H - X_L)$ then punishment has no effect on repression commanded or provided. External punishment only deters γ if it induces agents to prefer the incumbent punishments.

When would we expect p_I to be low? Autocracies actually vary widely in the ability of agents to check the autocrat (Svolik, 2009). Certain autocrats such as Stalin and Saddam Hussein so effectively consolidated power and terrorized their agents that the agents could not credible threaten to sanction them. Such autocrats can, and do, use punishments much greater than plausible punishments in democracies, such as the execution of the agents family. Fortnately, many other autocracies feature formal institutions that enable collective action among regime members, such as modern Russia's Duma. Forbidding discrete observable acts in consolidated autocracies has negative effects, but while in consolidated autocracies the effect depends on internal regime characteristics.

Model 2: Punishing Repressive Effort

This second model examines the direct punishment of repressive effort, rather than some discrete observable act. The game proceeds as follows.

Prior to the game, all actors observe q, the probability of regime failure with minimum repressive effort. The game starts with the incumbent makes a binary decision to order repression or not, $r \in 0, 1$. The agents then choose a level of repression visible repression $x \in x_H, x_L$ and a level of invisible effort $x \in y_H, y_L$. For simplicity I neglect defection once again.

X here represents all the components of repression that the autocrat can observe. It could include counts of people killed, tortured and imprisoned, as the Stalin and Assad regimes practiced (Kalyagin et al.). It might include the extent of opposition activity in the agents zone of responsibility (Hubert and Little, 2020; Hassan and O'Mealia). Y includes those components that the dictator will not observe. It might include whether targets were actually involved in the opposition. More prosaically, it includes whether the agent misrepresents the cost of repression to receive more funding for less work done.

Nature then decides if the incumbent remains in power with probability $q + \eta + \phi$ where η is 1 if the $x = x_H$ and ϕ is 1 if $x = x_H$.

If the regime false, the external punisher chooses a punishment $_{e,x} \in 1, 0$ for visible effort and a punishment $_{e,y} \in 1, 0$ for hidden effort. If the regime remains the leader rewards agents that gave high effort with $\beta_H > \beta_L$.

The utility of the agent is given by

$$U_A = \begin{cases} U_A(X_H, Y_H) = (q + \eta + \phi) * \beta_H + (1 - q - \eta - \phi)(\bar{u} + p_{e,x}) \\ + p_{e,y} - c_x - c_y \\ U_A(X_L, Y_H) = (q + \phi) * \beta_H + (1 - q - \phi)(\bar{u} + p_{e,y}) - c_y \\ U_A(X_H, Y_L) = (q + \eta) * \beta_H + (1 - q - \eta)(\bar{u} - p_{e,x}) - c_x \\ U_A(X_L, Y_L) = q * \beta_L + (1 - q)(\bar{u}) \end{cases} \tag{3}$$

The effect of external punishment on observable effort is highly dependent on q. However, the herd dynamics of repressive effort have been well explored by Dragu and Lupu so I do not reprise them here. I therefore assume that q is sufficiently high that agents choose high effort regardless of punishment.

I focus on the interesting case, in which the agent plays $s_A = X_H, Y_L$ in the absence of external punishment. This implies that q is large enough that the reward motivates high effort and that $c(y) > \phi(\beta_H - \bar{u})$ so hidden effort is not worthwhile. When the external punisher conditions punishment on visible effort, the high effort condition becomes $c(y) > (\beta_H - \bar{u} + p_e)$. Thus high hidden effort becomes more appealing to avoid the punishment. This effect is dependent only on the agents contribution to regime survival, not the base rate q.

Punishment on hidden effort does not have this problem. Under no punishment the agent again plays y_H if $c(y) < \phi(\beta_H - \bar{u})$. Punishment on hidden effort changes the y_H condition to $c(y) + P_{e_y}(1 - q - \eta) > \phi(\beta_H - \bar{u})$. Thus conditioning on hidden effort strictly decreases the amount of hidden effort.

Observable Implications

1. Incumbents can induce greater effort by reducing an agents exit options. We should observe incumbents preventing spontaneous defections during repressive effort.

2. Punishments are effective against non-observable actions. Autocrats should therefore invest heavily in observing that agents carry out forbidden repressive acts.

3. Agents that publically commit forbidden acts are more incentive aligned with the incumbent. When incumbents are safe from internal coups but in external danger, they should promote agents who have publicly committed atrocities.

Case Studies

This section provides brief case studies that suggest the observable implications. The first subsection argues that autocrats try to make forbidden acts more like the γ acts, i.e. impossible to defect before committing. It relies on several very short descriptions of defections. The second section discusses compelling agents to violate norms.

Ultimately, no cases were identified that unambiguously supported commanding repression to increase agents exit costs. This could reflect that outsiders already possess such detailed information about repression that autocrats do not need to bait retaliation. Unambiguous evidence of the intentions of such secretive organizations is rare, so the failure to find strong cases does not prove the null, but does update downward on my model. Strong evidence was found that consolidated autocrats use collective punishment to deter defection and shirking, and that exit costs do motivate higher effort.

Defectors

If the model in part 1 is accurate, autocrats should "spring" repressive tasks on actors who are unable to defect.

Matei Pavel Haiducu was a Romanian secret agent charged with industrial espianoge in France in 1981. While abroad he received an order from his superiors to assassinate two dissident writers in exile in France. Haiducu chose to defect and cooperate with the French secret service. With the French, Haiducu staged the assassinations of both writers to satisfy his superior. He then returned to Romania and retrieved his family. The elaborate plot to fake the murders suggests that Haiducu felt his families safety was a precondition for defection.

That same year, Iranian Air Force colonel Javad Hussain hijacked a C-130 airplane and flew to Turkey with his wife. He claimed that he would be executed in his home country. It was no coincidence that his three children were at school in the United States at the time.

In 2016 North Korean diplomat Thae Yong-ho defected after he and his family were recalled to Pyongyang. Thae’s wife and children were in the UK at the time, his brother and sister were in country and he believes they are now imprisoned. Thae commented "In North Korea defection itself is a great offence to the system and to the leadership (...) the families would be heavily punished, especially the family of high level defectors like me".

These stories all suggest that defection requires careful preparation and a fortunate opportunity. Also note that two cases occured in consolidated autocracies where the threat of a coup no longer constrains the dictators punishments (Hussain likely fled during a consolidation episode). This validates the assumption that consolidated autocrats can compel discrete acts with enormous sanctions, while suggesting that autocrats know how to use this leverage.

Burning Bridges to Show Loyalty

Model one also suggests that autocrats should compel forbidden acts to align the fate of key agents with the regime. Unfortunately proving that such acts are strategic displays of "bridge burning" is difficult because a true bridge burner must feign sincerity. If an agent commits the repressive act in such a way as to emphasize his coercion, he avoids the blame and makes the act pointless. Also, repressive agents have career incentives to display high willingness to violence (Hassan and O'Mealia). There is also the prosaic problem that agents are less famous than autocrats.

In consolidated autocracies, key supporters who are internationally wanted are common. In the Assad regime, Maher al-Assad, Assef Shawkat and Muhammad Keirbek are wanted in connection with the assassination of Rafic Harriri. They are the commander of the Republican Guard, the Chief-of-Staff of the army and the secret police chief, respectively (Black, 2011). Mohamed Hamdan Dagalo was a key supporter of the Sudanese regime until 2019, and is part of the current civilian-military government. Dagalo's peripheral origins in Darfur and lack of education made him a surprising candidate. But he orchestrated the 2004 genocide in Sudan, where is crimes are well documented and common knowledge (Reeves, 2019). The after the genocide autocrat Omar Al-Bashir overlooked his lack of qualifications and admitted Dagalo to the junta. Al-Bashir was himself an internationally wanted man at the time, giving him an acute need for natural allies.

Repression in the Syrian Civil War

As the Assad regime watched Tunisian and Egyptian regimes collapse, they resolved to meet any dissent with extreme violence as a deterrent. This strategy required high effort from agents as yet unprovoked by opposition violence. The regime badly needed to motivate effort.

As one of many solutions for motivating effort, the Assad regime divided responsibility for killing civilians between the different army divisions. In 2014 a military policeman responsible for photographing executed bodies defected, providing a detailed account of how killings are documented (Guardian, 2015). People are tortured and killed by four intelligence agencies, the army intelligence, air force intelligence, political intelligence and state intelligence at separate facilities. A smaller number are killed by a separate "palestinian" military arm. The bodies are taken to a military hospital next to the presidential guards barracks, where they are careful catalogued and tallied by branch. The branches are not allowed to coordinate with each other.

The decision to divide responsibility for killings between branches is surprising. Ordering civilian killings is a dangerous action for most autocrats, as it places targeted costs on the agent. Autocrats cannot observe agents individual ressistance to extreme crimes, so each order risks inducing a mutiny or defection. These behaviors make more sense if intended to spread complicity across the regimes to increase the exit costs for each arm. The inclusion of the air force in particular makes sense; they have no special skills in political terror, but are crucial to state violence. The air force has the the lowest exit costs because pilots can and do simply fly away from illegitimate regimes. This behavior is consistent with model 1; the leader structures violence to be immediately survelable to prevent defection, and builds a record for punishment afterward or release in the event of defection.

The one inconsistency is that this evidence was never intentionally leaked, by all appearances. The information did leak after several high profile defectors, including Anwar Raslan from the introductory vignette. The information was eventually used to convict and imprison defectors. However, the details of the leak are inconsistent with any state intention. The photographers families were threatened to deter leaks. The lax surveillance of photographers is explainable given their superiors were concerned with delivering body quotas on time to protect themselves. The defectors also openly described gratuitous crimes by anonymous low-level officials which only delegitimize the regime. The regime either held the information as kompromat on higher level officials (Hubert, 2020) or to punish shirking.

The Murder of Jamal Khashoggi

Agents also desire commitment from the leader to protect the regime. As a result, agents may compel the leader to sabotage her own exit options in contested autocracies. This is a plausible explanation for the killing of Jamal Khasoggi at the request of Saudi Crown Prince Mohammad Bin Salman (MBS). MBS became de facto ruler of Saudi Arabia in 2017. His first act was to publicly alienate himself from the royal family, the traditional base of Saudi support, by torturing and expropriating several relatives in the same hotel where he holds press conferences. This committed MBS to his current supporters, but did not reduce his exit options. MBS was popular in the west, particularly the US. On a tour in 2018 he met President Trump as well as celibrities Oprah Winfrey, Jeff Bezos and Bill Gates.

Later that year MBS publically destroyed his goodwill in the west. He commanded the execution of Jamal Khashoggi, a defector who wrote for the Washington Post. The execution was carried out inside the Saudi Embassy in Turkey, where Khashoggi's entrance to the embassy was filmed, leaving no doubt about the intent. The assassination was of course a deterrent against dissent within the Saudi diaspora community. But this hardly seems worth alienating Saudi Arabia's most powerful allies, on whom the kingdom depends for military protection. MBS's need to credibly signal loyalty to his core supporters is a more plausible account. His supporters could have requested that he authorize the brazen operation to prove his loyalty, because a disloyal MBS would refuse it. With no outside enforcer, credibility is so precious for autocrat that the price is plausible. Alternatively, MBS is so bloodthirsty and illiberal that he chose the operation freely. But he would want us to think that.

Prisoners in the Second World War

The starkest illustration of this model comes not from repression but from open warfare. By the 20th cenutury the world's militaries consistently played a equilibrium in which POW's were not executed. Since keeping POWs alive is an expensive burden on embattled states, it required enforcement by reciprocity and post-hoc punishment. States that executed POWs were sanctioned with the killing of their captured soldiers and the execution of officers post war. The Empire of Japan broke this equilibrium by declining to ratify the Geneva conventions, directing soldiers to disregard the treaty, and mistreating their citizens (Dower, 1986). This policy greatly increased the exit costs on Japanese soldiers and officers. Allied soldiers often killed surrendering Japanese soldiers in retaliation, and Japanese officers faced execution as war criminals. To mislead soldiers about their exit options, Japans foreign ministry refused to alert Japanese families when members became POWs (Dower, 1986).

Despite the differences, the POW case illustrates model 1 in great detail. The Japanese regime called for crimes to intentionally manipulate their agents exit costs. These higher exit costs resulted in greater effort to preserve the incumbent, in the form of refusal to surrender (Dower, 1986). Interestingly, German officers western prisoners were treated well and as result German soldiers and officers had lower exit costs and many strategically surrendered to the allies. 19 German soldiers surrenedered to the western allies for every one battle death, while 19 Japanese soldiers died for every one who surrendered. This indicates that higher exit costs do motivate agent effort. The American forces did believe exit costs motivated higher effort, and launched a propaganda campaign to induce more surrendering from Japanese and acceptance from American troops (Gilber, 1995).

Discussion

Externally punishing autocratic agents presents severe challenges because punishment can only occur if the regime collapses. As a result any punishment induces greater effort to maintain power, binding the agent and autocrat more tightly together. Even worse, I show that a commitment to punish agents for a repressive act can actually make the act more appealing. Given these problems, should external punishment be used, and if so when? Here I give some practical but speculative advice.

Punishments delivered during the regimes tenure are likely to succeed. The most common opportunity is during military interventions and peacekeeping missions. If agents know that foreign militaries can resist their repressive efforts, they are much less likely to cause them. Because unheeded commands embarrass the autocrat, the autocrat simply refrains from commanding repression in any Nash equilibrium.

Punishing easily observable acts in consolidated autocracies has the worst outlook. Dictators with unconstrained punishment ability can deter defection in the short term and force any compliance. Such easily observable acts include the signing of orders, the execution of prisoners and public statements of support for repression. Punishing agents for these acts increases both repressive orders and effort regardless of the expected stability of the regime. The result is strongest in consolidated autocracies, and when agents can constrain the autocrat the effects are contingent on expectations and power-sharing arrangements.

Punishing post-crisis observable acts is also dubious. Dictators may observe effort post-crisis through formal records of terrorized persons, through informants or by measuring regional opposition activity. External punishments for these acts may increase or decrease effort depending on the agents outcome expectations. Unfortunately, expectations already strongly influence political survival. Autocracies whose agents expect defeat are already vulnerable to defection cascades because the regime cannot promise rewards or sanction defection. Furthermore, such punishments could cause selection in which only the confident join the regime.

In our model repressive effort is weakly decreasing in punishments for non-observable effort. If external actors commit to punish effort which the autocrat will not observe, there is no arrangement where punishment increases effort and some where punishment decreases effort. Autocrats cannot perfectly monitor agents, so non-observable efforts do exist. But the practical challenges of identifying repressors who go "above and beyond" are steep.

In general, punishing repressive agents is a difficult and risky strategy. In many cases it is counterproductive to democratization regionally. Fortunately, leaders in democratizing states are well aware of this problem. A resent instrumental variables study design by Pearce Edwards finds that greater human rights organization activity increases amnesties granted to repressive agents (2020). This occurs because democratizing regimes realize that with greater incentives to re neg on their promises of non-punishment, they need a stronger commitment to amnesty.

Finally, there is now a diverse set of strategies that external actors employ to reduce repression and encourage democratization. Regimes can be collectively punished with sanctions or canceling lines of credit. States and private donors can provide aid conditional on political reform, as long as the threat to revoke funding is credible. Military interventions are effective at democratizing non-personalist regimes. Economic development has a positive effect on both democratic transition and consolidation.

Future interventions should consider reducing the exit costs for repressive agents. The efforts of consolidated autocrats to prevent defection defection suggest it does damage their survival. Decreasing exit costs would counter those efforts. In general defectors require protection from assassination by loyalists and prosecution by victims. This could take the form of witness protection for former repressors, similar to what the Syrian opposition attempted in 2012. Relocating agents would also decrease their ability to launch political comebacks which destabilize democratic transitions.

References

Kopka H and Daly PW (2003) *A Guide to LaTeX*, 4th edn. Addison-Wesley.

Lamport L (1994) *LaTeX: a Document Preparation System*, 2nd edn. Addison-Wesley.

Mittelbach F and Goossens M (2004) *The LaTeX Companion*, 2nd edn. Addison-Wesley.