Authoritarian Dependence on Agents of Repression and Regime Survival: a Transitional Justice Perspective

Steven Boyd and Monika Nalepa

The University of Chicago

Prepared for Presentation at the *APSA*, August 26, 2023

Abstract

We investigate the dilemma of rank-and-file members of the authoritarian regime apparatus when facing orders of repression. Such agents not only incur the moral cost of carrying out repressive orders but may be punished for human rights violations by the new democratic elite if the autocracy collapses. Yet if they refuse the order, they may be sanctioned by the autocrat for refusing to carry it out. Most existing models of transitional justice focus on holding accountable leaders of authoritarian regimes leaving transitional justice targeting low-ranking agents of repression out of the analysis. Accounting for low-ranking agents of repression exposes under-theorized aspects regime transitions and transitional justice. We illustrate our findings with descriptive statistics from the Global Transitional Justice Dataset, which reports prosecutions over time and includes information on the rank of defendants both before and after the regime transition. This information allows us to construct an original independent variable: the proportion of rank-and-file held accountable as a percentage of all criminal trials against former perpetrators (the relative severity of justice).

The authors are grateful to Ipek Cinar, David Crabtree, Nicole Martinez, Moksha Sharma, Jacob Delgado, Tatum McCormick, Julian Santesteban, and Alma Moskowitz for stellar research assistance and for funding their work to the National Science Foundation, the Center for International Social Science Research, and the Pearson Institute for Global Conflict at the University of Chicago. All mistakes are the authors' responsibility.

> *Bear in mind that they would not have been able to do what they did without others to assist them. Nor would they have attempted to come here, except in the hope of being rescued by those same people, who have come not to help them, but in belief that if you acquit those who are responsible for the greatest evils, they themselves will have virtual impunity for what they did and might wish to do so in the future*

Lysias in *Against Erastothenes* (Todd et al., 2000*a*)

1 Introduction

Following the rule of the Thirty Tyrants (a pro-Spartan oligarchy installed in Athens after its defeat in the Peloponnesian War), the Athenians took several measures against the Tyrants and their collaborators. *Dokimasia* was a screening procedure aimed at vetting who among the Athenian citizens had collaborated with the Thirty. Citizens proven to have collaborated could not hold public office (Todd et al., 2000*b*). A similar fate awaited those who had served in the cavalry of the Thirty. To verify if a citizen had been a member of the cavalry, Athenians consulted the *sainidion*, a register of the cavalry (Todd et al., 2000*b*). Finally, there were criminal prosecutions. The Thirty themselves and their supporting council of Eleven were prosecuted and, in most cases, sentenced to death, but all 3000 of their supporters were amnestied and allowed to seek refuge outside of Athens, with one exception: *If they had killed another man with their own hands* (Todd et al., 2000*a*).

The choice made by the Athenians to prosecute most severely those fulfilling orders seems highly unusual. For one, it is hard to prosecute rank-and-file agents of repression because of the principle of non-retroactivity (*nullum crimen sine lege* or "no crime without a law"), a rule of law standard that constrains new democracies and prevents prosecutions for conduct that was not only legal under the previous constitutional framework, but indeed, encouraged. At least from a legalistic point of view, it is easier to prosecute order-givers than order-takers.

In certain instances, such agents should not be punished because they contributed to the fall of the authoritarian leader, as happened in Tunisia in 2011. According to Masri (2017), Ben Ali "relied heavily on his police forces to quell the populace" because he could not

depend on the army that "did not have the power nor political will to intervene on behalf of the regime". On numerous occasions, Rachid Ammar-Alis chief of staff refused orders to fire on protesters during the Arab Spring. Ultimately, the army developed a reputation for being an ally and supporter of the demonstrators.[1]

Tunisia's example illustrates that rank-and-file agents of repression may be spared accountability if they refuse to execute orders to perform repression. Such agents know that if they are asked to carry out repression and refuse, this refusal may contribute to the regime's downfall and allow them to avoid accountability from the new democratic elites.

Yet another case, also from the recent Arab Spring in the Middle East, calls into question the prudence of punishing the rank-and-file by the new democratic regime. In 2011, the Arab Spring put an end to the nearly 30-year rule of Hosni Mubarak. The transitional government directly succeeding Mubarak largely ignored the rank-and-file when holding the authoritarian government to account. However, Morsi's government, which won the free elections of 2012, changed course and after purging the military leadership started to to prosecute and even convict rank-and-file perpetrators. Just months after the first rank-and-file conviction, the military launched a successful coup, creating a military government (Hoyle, 2019).

Even when the new democracy pulls its transitional justice punches, it cannot guarantee rank-and-file immunity from prosecutions. Consider the case of Augusto Pinochet's regime. Years before Chile's democratic transition, in 1980, the regime arrested twenty police officers, who were charged for the kidnapping and beating (resulting in at least one death) of member of the far left opposition to the the Pinochet government. In a rare admission of the officer's responsibility, the Pinochet administration publicly admitted that the officers may have committed human rights violations (AP, 1980).

Beginning with the Nuremberg trials following WW2, transitional justice—the holding

[1]Masri (2017) claimed that "resentment had built up among the forces after Ben Ali directed leadership purges in the 1990s, replacing both senior and junior military officers whom he accused of having Islamist leanings and allowing the police to attempt to exert authority over the armed forces." Consequently, all the violent crackdowns on protesters were carried out by security forces. The security forces were also better rewarded and respected than the military by the dictator.

to account of members and collaborators of former authoritarian regimes for human rights violations that took place before the democratic transition—has focused on prosecutions of former authoritarian leaders, that is, those issuing orders of repressions (Escribà-Folch and Wright, 2015; Krcmaric, 2018, 2020; Bates, 2021).[2] Prosecutions of the rank-and-file agents of repression have at most received attention in the form of single country studies (Duursma and Müller, 2019; McAllister, 2019). Beyond the literature on coup proofing (Harkness, 2018; Greitens, 2016; GREWAL, 2023), political scientists have not theorized about them, nor have they been studied globally from a comparative perspective, in large part due to the lack of available data.

What has also evaded scholars is that, in contrast to the leaders issuing orders, prosecution of agents of repression can occur prior to a country's transition to democracy, as the Chilean example illustrates. Why would an authoritarian regime punish its own agents of repression? Would that not decrease their incentives to support the regime? On the surface, it seems that it would. We theorize, however, that given the threat of post-transition accountability that agents of repression face, authoritarian sanctions can be designed in a way to counteract those incentives.

We propose a novel approach to analyzing how states deal with agents of repression. The analysis takes into account both the dynamics that emerge in the authoritarian setting, when these agents of repression are asked to engage in human rights violations to prop up the regime, as well as the dynamics of transitional justice that develop in the post-authoritarian, democratic period.

We begin by reconstructing the conventional wisdom about how a rank-and file member's readiness to execute orders of repression is secured in a regime where this support of the rank-and-file is critical. That is, the regime depends on rank-and-file support for its survival. This exercise exposes gaps in this conventional understanding of the rank-and-file members' calculus. In light of these shortcomings we next develop two formal models that stochastically

[2]Additionally, much of this literature focuses on accountability from the hands of international tribunals of various sorts (Krcmaric, 2020; Bates, 2021; Mallinder, 2008)

accounts for the regime's ability to survive or fall with and without rank-and-file members' support.

We end the paper by proposing a research design for testing the implications of our theory with a new dataset of criminal trials of authoritarian perpetrators before and in the aftermath of democratic transitions. The dataset, which is an extension of the Global Transitional Justice Dataset (Bates, Cinar and Nalepa, 2020), not only codes criminal trials of perpetrators over time, but in addition to the volume of trials, distinguishes between the prosecutions of order-givers and order-takers.

2 The dilemma of criminal prosecutions

This section uses the literature to present the dilemma facing rank-and-file members of the repressive apparatus and formulates the conventional wisdom about how these agents make decisions in a simple model. We then use Varieties of Democracies Data to illustrate why what we seem to know about the calculus of rank-and-file, as reconstructed in a simple model, fails to pass the test of real world data.

The dilemma of criminal prosecutions against agents of authoritarian repression is: how to punish perpetrators of atrocities under authoritarian regimes harshly enough to deter them from obeying cruel orders but mildly enough for them to be an asset in the democratic transition rather than a hindrance to democratic stability.

According to a broad section of the transitional justice literature, agents of authoritarian repression should be punished following transition to democracy so they do not obey cruel orders from autocrats (Trejo, Albarracín and Tiscornia, 2018; Vinjamuri and Snyder, 2004). The reason to prosecute agents of repression is that without sanctions for carrying out the cruelest orders, agents of repression, unscathed, can be recruited by future autocrats (Vinjamuri and Snyder, 2015; Kim and Sikkink, 2010). Transitional justice targeting the rank-and-file decreases their willingness to repress authoritarian resisters and thus can bring

about the end of authoritarian rule, contributing to democratization. At the same time, democracy carries with it constraints that make punishing former agents of repression difficult. These constraints are normative (retroactive justice violates key rule of law principles) and legal (punishing rank-and -file for carrying out orders that were legal at the time they were issued violates the principle of "Nullum crimen sine lege" [*lat.* No crime without a law]. Yet retroactive justice, no matter how legally complicated, acts as a deterrent for future autocrats.

The practical reason to avoid transitional justice against the rank-and-file is that such punishment may induce them to undermine democracy. This undermining could take the form of rank-and-file members joining their former leaders' attempts at thwart democracy in a military coup, as happened in Egypt in 2012. We know, however, that coups are a very unlikely way for democracies to fall (Singh, 2014). It is more likely for the rank-and-file to engage in the gradual and not necessarily purposeful undermining of democratic stability as described by Varese (2001); Volkov (2016); Bates et al. (2020*a*). According to these authors agents of repression who are let go from working for the state (either through purges or criminal prosecutions) may find reemployment in criminal organizations and make the establishment of rule of law close to impossible.

While the first body of literature described above addresses agents' incentives under autocracy, abstracting from what happens in democracy, the second merely focuses on their decision calculus under democratic institutions. These two decisions are clearly interrelated, as the threat of transitional justice can both reduce incentives for repression and increase incentives for the rank-and-file to support the regime. Describing the interplay of incentives under democracy and autocracy appears to be a good opportunity to use formal modeling.

Although to our knowledge, there are no theories of what these conflicting incentives imply for democratic stability,[3] two models exploring the relationship between authoritarian leaders and their agents under autocracy have been recently published.

[3]But see non-formal paper by Bates et al. (2020*b*) on how purges of former uniformed members of the ancien régime contribute to crime and especially violent crime

The first, by Dragu and Lupu (2018), uses a formal model with incomplete information to show that repression is most likely to hinge on expectations and is the result of a coordination game. Paradoxically, when authoritarian leaders need it most—when dissent against them is at its highest—mobilizing agents of repression is most difficult. While in Dragu and Lupu (2018), the autocrat's dilemma is limited to the authoritarian period, Tyson (2016) extends the consequences of the autocrat's actions and the consequences of actions of his agents of repression into the post-authoritarian period, allowing for transitional justice. He models the interaction between a leader and his repressive apparatus in circumstances where the stability of the authoritarian regime is uncertain. The autocrat in these circumstances must compensate his agents of repression to offset their potential of being punished should the regime collapse. Tyson's model uses the prospect of transitional justice to model repressive agents' incentives but his theory does not distinguish between agents in different ranks.

Hence, the literature cited above either fails to distinguish between order-givers and order-takers in the context of transitional justice or it models autocrats and their agents of repression as passive recipients of transitional justice. In real-world settings, former autocrats may defend themselves from transitional justice by any of the three actions:

1. throwing under the bus their own agents of repression (holding them accountable while the authoritarian regime is ongoing)

2. bargaining for amnesty while negotiating the transition (Bates, 2020) and, where necessary, resorting to skeletons in the opposition's closet(Nalepa, 2010*b*) and

3. engaging in anti-democratic backlash once the democratic transition is completed. (Nino, 1996; Sikkink, 2011) or sabotaging the consolidation of democracy (Helmke, 2012).

In this version of the paper, we will consider the first possibility from this list. The goal of our modeling exercise is to understand the dynamics that result from the agent's desire not to be punished and the autocrat's desire to remain in power when faced with a

threat by a democratic challenger. We present several possible formulations of the game these actors face, presented from least to most complex. We begin with a simple formulation that accounts for several of the factors entering the calculus of both players. The scope conditions matching this model best are ones where the authoritarian regime is so heavily dependent on the rank-and-file for survival: with their support, it survives but without it, it collapses. Such conditions obtain most starkly in non-personalist regimes (Geddes, Wright and Frantz, 2014), where the ruler engages in a fair deal of power-sharing (Meng and Paine, 2022). Hence the rank-and-file's decision to execute orders of repression is both *necessary and sufficient* for the regime's survival.

2.1 Model 1: support of rank-and-file necessary and sufficient for regime survival

First, consider a two-period, two-player game of complete and perfect information. The players are the autocrat or *Leader* represented with L and a representative member of the rank-and-file or *Follower* represented with F.

The game starts with L's choice of the *level of repression* $R \in [0, 1]$, with R closer to 1 representing more severe measures ordered by the autocrat. F moves in the second period and chooses to either obey or disobey ($o = 1$ or $o = 1$) the order. In this case, let us suppose that the agent's obedience is both necessary and sufficient for the regime's survival: if the agent follows the order, then the regime survives and if the agent disobeys the order, the regime transitions to democracy.

We assume that the autocrat will face transitional justice proportional to the level of his regime's brutality if there is a democratic transition, *and* that repressing citizens makes him a less legitimate leader. Hence, the order to repress is costly whether it is followed or not. L pays a cost R if the regime survives and RT if it transitions to democracy. T here represents the cost of transitional justice to the autocrat. For now we allow both $T > 1$ to represent harsh transitional justice and $T \leq 1$ to represent mild transitional justice. We normalize

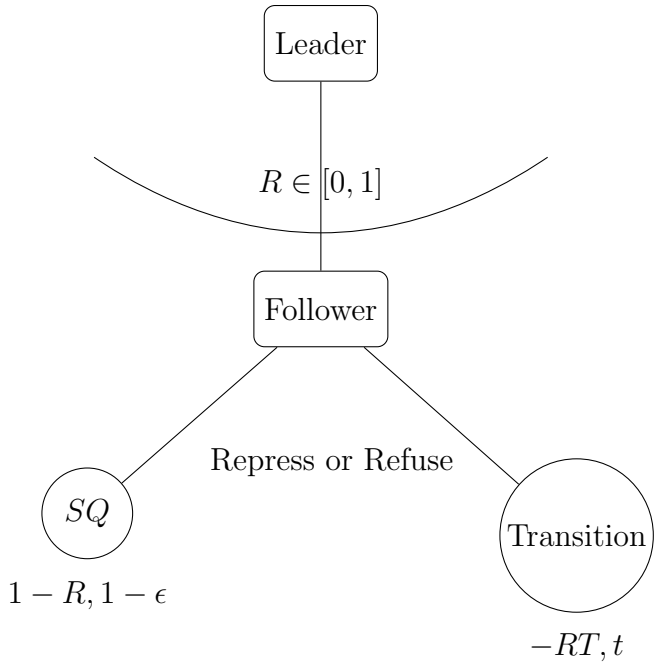


Figure 1: Simple Game reflecting interaction between rank-and-file and leader issuing orders

the benefit of holding office to 1, so if the order is obeyed, the autocrat's payoff is $1 - R$ and $-RT$ otherwise. Following the order to repress may also be costly. The agent suffers a disutility of $-\epsilon \leq 0$ for following the order, but 0 otherwise. Recall, that in this first model, the regime never falls in the event that the order is followed, hence $-\epsilon$ represents purely ethical considerations of the agent. If the agent refuses the order, he receives $t < 1$, which is the value of being a former rank-and-file agent following the transition to democracy.

This interaction is represented with the game tree in Figure 1.

This simple model can be solved for Subgame Perfect Equilibrium by backward induction. First note, that in the final stage, the agent will choose to execute the order as long as $1-\epsilon \geq t$ and choose to refuse it otherwise. When solving for the L's optimal action, we consider both cases. First suppose that the agent pays a very low moral cost for executing the order (low ϵ) or that the agent cannot count on a career in the democratic enforcement apparatus (low t). In this instance, L receives a payoff $1 - R$ if he issues an order of repression at the level R. Hence, he gets his highest highest payoff from issuing an order for the lowest possible level of repression, that is $R = 0$.

Now consider the second case, when ϵ is high because the agent suffers high moral costs from following repressive orders or when t is high because his career prospects following the democratic transition are promising. In this instance, L obtains $-RT$ from setting the level of repression at R. Hence, again, he maximizes his payoff by choosing the lowest level of repression possible, that is $R = 0$.

This simple formulation of the game has a unique subgame perfect equilibrium in which the agent obeys of disobeys the order (depending of the respective magnitudes of ϵ and t) and the autocrat orders the minimum level of repression ($R = 0$). This is due to the pivotality of the agent's decision. Even a very minor preference to *not* repress the uprising is sufficient to deter an agent who knows their action is decisive from following the order. Knowing that their order will never be followed, a rational autocrat will minimize their order to repress to reduce the severity of transitional justice meted out by the new regime.

The key finding from solving this simple model is that regardless of whether harsh or mild transitional justice is applied to the Leader if regime fails and regardless of whether the rank-and-file have strong moral reasons to avoid implementing an order to repress, and finally, regardless of whether the agent of repression has promising or poor career prospects following the democratic transition, the leader has unique optimal strategy to repress as little as possible.

This model is clearly falsified by what we observe in authoritarian regimes, regardless of whether they undergo transition. Leaders issue orders to repress; sometimes these orders are fulfilled and sometimes they are not. In Figure 2, we use data from the Varieties of Democracy (V-Dem) dataset to demonstrate this point (Coppedge, 2021). Specifically, we use a re-coded version of V-Dem's *physical violence index* (`v2x_clphy`).[4] We have re-coded this index so higher values represent higher levels of political violence carried out by the state. [5] We plot the political violence index for the subset of countries which appear in our

[4]The physical violence index itself is the average of the freedom from torture and freedom from political killings measures

[5]The original version of this variable uses higher values to represent greater respect for physical integrity

Global Transitional Justice Dataset (Bates, Cinar and Nalepa, 2020), but in the years that correspond to the decade preceding the democratic transition. [6]. It is clear from the plot that although the level of repression often drops in the years preceding a transition, in many cases it does not change or increases. In many instances, the decline in repression does not occur until the year of the transition, so it is possible that the decreases that we see in many cases reflects the behavior of the *new* regime, not the former autocratic one.

In light of these findings, in the next subsection, we relax one of the assumptions of the simple model: that the the support of the Follower is necessary for regime survival.

2.2 Model 2: support of rank-and-file is sufficient for regime survival

In the simple formulation above, the Follower's ability to determine the regime's survival implied that even a minor moral cost associated with repression was sufficient to dissuade them from executing the repressive order. For this reason, the Leader had no incentive to issue more than the minimum level of repression. As we observe, leaders in autocratic regimes *do* sometimes order harsh repression and the rank-and-file *do* sometimes follow these orders. To understand these dynamics, we need a more complex model that captures the *uncertainty* that leaders and followers face.

To this end, we introduce a stochastic element to the outcome of the game. Suppose that the Follower knows that they are sufficient, but not necessary for the regime's survival. As in the simple model, if the Follower obeys the order to repress, the regime survives with certainty. The Leader then pays the cost R (for a total payoff of $1-R$), while the Follower, as in the baseline model receives $1-\epsilon$, with ϵ reflecting the moral cost of executing a repressive order.

The crucial difference relative to the simple model is that if the Follower does not obey

[6] We omit the following countries from the figure because of missing data: Estonia, Bosnia and Herzegovina, Latvia, Lithuania, Moldova, Montenegro, North Macedonia, Slovakia, Slovenia, and Ukraine

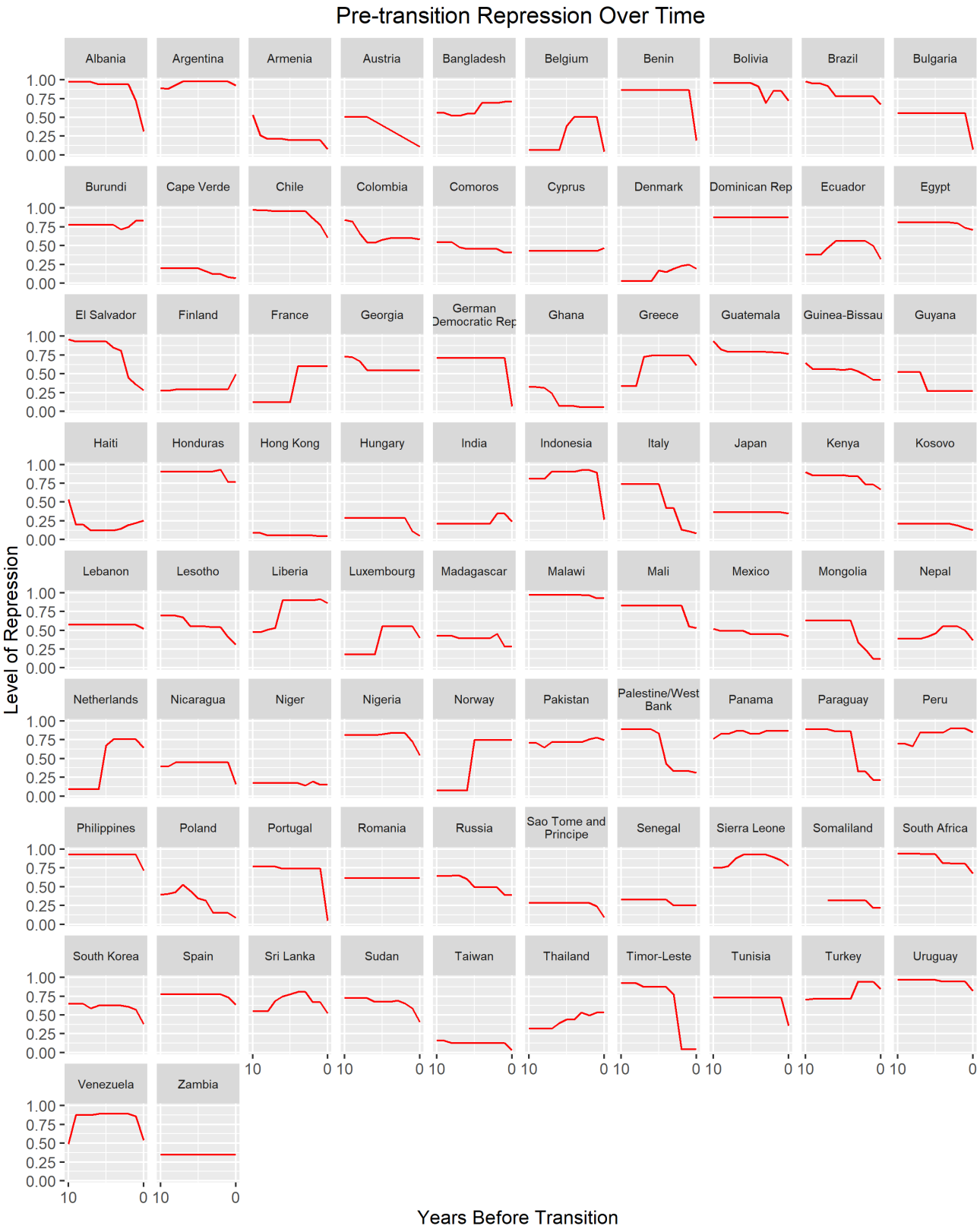


Figure 2: Level of violent repression in years leading up to regime transition

the order, the regime may still survive with probability $p \in (0, 1)$, in which case the Leader receives a payoff of $1 - R$ and the Follower receives $-\gamma R$. Here, γ represents the punishment the Leader doles out to the rank-and-file if they disobey his orders of repression and he is in a position to do so. We remain agnostic about the form or pretext this punishment takes. For instance, as the Chilean example from the introduction illustrates, the Leader may publicly claim that he is implementing transitional justice against the rank-and-file, only in reality, to sanction them. Hence, the Follower has to weigh the possibility of being punished for disobedience against their distaste for repression.

In the event that the regime does not survive, which happens with complimentary probability $1 - p$, the Leader receives a payoff of $-RT$, where T still stands for transitional justice processes targeting authoritarian order-givers. Intuitively, this punishment is proportional to the level of repression they ordered, R. The Follower, in turn, gets t, which reflects his career prospects under the new democratic regime.

2.2.1 Equilibrium

Since this is a game of complete, though imperfect information, we can use Backward Induction to find the he Subgame Perfect Equilibrium. Nature is not a strategic player.

We begin by computing F's expected payoffs for obeying and ignoring the order to repress. The expected payoff to the Follower for refusing is given by

$$\mathbb{E}U_F(o = 0) = -\gamma Rp + (1 - p)(t)$$

while the expected payoff to the Follower for obeying is

$$\mathbb{E}U_F(o = 1) = 1 - \epsilon$$

Thus, F weakly prefers to refuse the order to repress if:

$$R \leq \bar{R} = \frac{t(1-p) - 1 + \epsilon}{\gamma p}$$

$\bar{R}$ here represents the critical level of repression below which the rank-and-file refuse to repress. $\bar{R}$ is increasing in ϵ, the moral reservations of the rank-and-file, increasing in t the democratic career prospects of the rank-and-file, and decreasing in γ, the punishment the Leader can apply to the insubordinate rank-and-file.

To find the SPE, we consider two cases:

1. $R > \bar{R}$

2. $R \leq \bar{R}$

In case 2, L's expected payoff is

$$\mathbb{E}U_L(R, o = 0) = p + R(pt - p - t) \tag{1}$$

Since the parenthetical expression on the right hand side (RHS) of 1 is negative, L maximizes his payoff selecting the lowest R that satisfies the constraint 2, that is $R^* = 0$.

Consequently, G's expected utility of choosing this R is p.

Consider next, case 1. Here, L's expected payoff is $1 - R$, subject to $R \leq \bar{R}$ and it is straightforward to see that that the level of repression maximizing this payoff is

$$R^* = \bar{R} = \frac{t(1-p) - 1 + \epsilon}{\gamma p}$$

Thus, in equilibrium, L will either choose 0 or $\frac{t(1-p)-1+\epsilon}{\gamma p}$, depending on what yields him a higher payoff. If he chooses $R = 0$, his expected payoff is p. If he chooses $\frac{t(1-p)-1+\epsilon}{\gamma p}$, his expected payoff will be $1 - \frac{t(1-p)-1+\epsilon}{\gamma p}$.

Hence, the L will choose a positive level of repression, $R = \bar{R}$ as opposed to $R = 0$, if

and only if

$$p < 1 - \frac{t(1-p) - 1 + \epsilon}{\gamma p}$$

The above inequality reduces to a condition that can be expressed in terms of easy to interpret parameters:

$$(1-p)(t - \gamma p) \leq 1 - \epsilon \tag{2}$$

The RHS of expression 2 represents the value of the status quo to the rank-and-file, that is, the value of holding office minus the ethical cost of repression. The left hand side (LHS) of 2 contains two terms. One represents the probability the regime will collapse if the agent refuses to carry out orders, while the other is the difference between career prospects for the rank-and-file under democracy and the punishment under autocracy.

Intuitively, the appeal of the status quo makes it easier for the leader to order repression, as does the probability of regime survival despite the order to repress and the magnitude of punishment the autocrat can impose. Improving career prospects for rank-and-file under the democratic regime has the opposite effect.

2.3 Model 3: support of rank-and-file neither sufficient nor necessary for regime survival

In this case, the obedience of the rank-and-file is neither necessary nor sufficient for regime survival. There are still just two players, L, who moves in first period, F, who moves in the second period and a move of Nature, which determines regime survival. In the first period, L orders the level of repression, $R \in [0, 1]$. In the second period, F decides whether to obey the order ($o = 1$) or to refuse ($o = 0$). The part of the game following refusal to implement the order is exactly the same as in Model 2.

In the event that F obeys, we introduce an additional stochastic element. Instead of assuming the regime survives with certainty, as in Model 2, we assume that it survives

with probability q and transitions with complimentary probability $1-q$, where $q \in (p, 1)$. Transition is always possible, though the regime is more likely to survive with obedience than without it ($q > p$). The added uncertainty compounds the Follower's dilemma, as now he can be punished both for disobeying an order to repress (by the Leader if the regime survives) and punished (by the new democratic elite) for obeying the Leader if it collapses.

L's payoffs in this exchange depend on the regime's survival. When the regime survives, L receives $1-R$ and when the regime transitions, L gets $-RT$. F's payoffs, in turn, are conditional both on the regime's survival and on their decision to repress.

If F disobeys the order but the regime survives, he gets $-\gamma R$. If F obeys the order and the regime collapses, he gets $-\alpha R$. $\gamma \in (0, 1)$ represents the severity of punishment by a surviving autocrat, while $\alpha \in (0, 1)$ represents the severity of punishment by a nascent democratic regime. As in Model 2, T represents transitional justice against the Leader and t represents the career prospects of the Follower in the event of a transition.

When their actions influence (but do not fully determine) the regime's survival, F must weigh the severity of punishment they expect from both sides. The structure of the payoffs, jointly with the stochastic elements of the game, ensure that the Follower prefers to be on the winning side of any challenge to the autocrat.

2.3.1 Equilibrium

Again, this is a game of complete, albeit imperfect, information and so a Subgame Perfect Equilibrium can be found via Backward Induction. Since the regime's survival in the final period is determined stochastically, we begin by considering F's decision to obey or ignore the order to repress. The condition for F to follow the order can be written as:

$$\mathbb{E}U_F(R, o = 0) \geq \mathbb{E}U_F(R, o = 1)$$

which is true if and only if

$$p(-\gamma R) + (1-p)(t) \geq q(1-\epsilon) + (1-q)(-\alpha R - \epsilon)$$

This, in turn, reduces to

$$R \geq \bar{R} = \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha} \tag{3}$$

We refer to equation 3 as the **"obedience condition"** as $\bar{R}$ represents the critical level of repression that L has to order to induce obedience from the rank-and-file.

$\bar{R}$ is decreasing in the probabilities of regime survival (whether following F's refusal or obedience) and increasing in F's moral reservations regarding repression. It is also decreasing in the severity of punishment from the hands of a surviving autocratic regime, γ (for refusing the order) but increasing in the the severity of punishment from a the nascent democratic regime, α (for obeying the order).

The obedience condition provides us with a cutoff for determining whether F will obey or disobey the order in period 2. It partitions all possible values of R into two regions: above and below the $\bar{R}$ cutoff. Recall, that since L's payoffs are decreasing in R we only need to focus attention on lower limits of these two regions. Hence, the only two possible choices of R^* in equilibrium are $R = 0$ and $R = \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha}$. To find the ultimate best response of L, we need to determine in each of these two cases, which provides him with the higher payoff.

1. $R \geq \bar{R}$

2. $R < \bar{R}$

To determine under which conditions each possible value of R^* is a best response for L, we calculate L's expected utility in two cases defined by the $\bar{R}$ cutoff.

1.

In case 1 (where the obedience condition is satisfied), L's expected payoff is:

$$\begin{aligned}\mathbb{E}U_G(R = \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha}, o = 1) &= q(1-R) + (1-q)(-RT) \\ &= q - R(q + T - qT) \\ &= q - \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha}(q + T - qT)\end{aligned}$$

When the obedience condition is not satisfied (case 2), L's expected payoff is:

$$\begin{aligned}\mathbb{E}U_G(R = 0, o = 0) &= p(1-R) + (1-p)(-RT) \\ &= p(1-0) + (1-p)(0) \\ &= p\end{aligned}$$

Directly comparing these expected payoffs, we see that L's best response is

$$R^* = \begin{cases} 0 & \text{if} \quad p > q - \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha}(q + T - qT) \\ 0, \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha} & \text{if} \quad p = q - \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha}(q + T - qT) \\ \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha} & \text{if} \quad p < q - \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha}(q + T - qT) \end{cases} \tag{4}$$

We can rearrange $p > q - \frac{\epsilon - q + t(1-p)}{\gamma p + \alpha q - \alpha}(q + T - qT)$ to isolate γ and α on opposite sides of the expression to arrive at:

$$\gamma < \frac{(\epsilon - q + t - tp)(q + T - qT)}{p(q-p)} - \frac{\alpha(q-1)}{p} \tag{5}$$

This provides us with a **"zero repression condition"** indicating when the leader is better off ordering $R = 0$ rather than $R = \bar{R}$. We summarize our finding in the only proposition of our paper, as follows:

Proposition 1 *For an authoritarian Leader to order repression against his citizens, his ability to punish agents who fail to execute orders must exceed a threshold* $\bar{\gamma} \equiv \frac{(\epsilon - q + t - tp)(q + T - qT)}{p(q-p)} -$

$\frac{\alpha(q-1)}{p}$*). The size of this threshold decreases in the extent to which the regime depends on rank-and-file for survival.*

To see the intuition behind this result, consider the decision facing each player. The Follower must balance the expected punishment by the Leader (i.e. γ) for disobedience against their moral concerns, their prospects under a new regime, and the severity of TJ against them if they obey and the regime transitions. If the threat of punishment is low, then to induce obedience, the Leader must compensate by ordering harsher repression (because the Follower expects punishment proportional to the harshness of the order). Recall that ordering harsher repression is costly for the leader, so there is a point at which he is better off ordering no repression. The more dependent the regime is on the rank-and-file, the less concerned they are about the threat of punishment for disobedience, and the less likely it is that the regime will be able to induce obedience.

2.3.2 Comparative statics

Presenting the zero repression condition in terms of γ allows us to see for what values of α and γ the condition is satisfied, after fixing other parameters of model. In this section we focus on visualizing comparative statics this way.

In all of the plots that follow, γ is on the vertical axis and α is on the horizontal axis. The shaded regions represent those combinations of α and γ that satisfy the zero repression condition, that is when it is optimal for the Leader to set $R = 0$. In Figure 3, we fixed all remaining parameters of the model as follows: the magnitude of the Follower's ethical concerns $\epsilon = .5$, the career prospects of the Follower under democratic rule, $t = .5$, the transitional justice sanction to the Leader if the regime collapses, $T = 1.25$, and the chances the regime survives despite refusing the order, $p = .5$. The three shaded regions represent different values of q, the probability that the regime survives following the Follower's decision to execute the order. These three values of q are also labeled on the plot.

Note that as the distance between p and q increases, the shaded region shifts upward.

The distance between p and q has an intuitive interpretation as the regime's dependence on the rank-and-file for survival. The upwards shift can thus be interpreted as follows: As the regime becomes more dependent on the rank-and-file for survival, the higher F's ability to influence the survival of the regime, the less salient the fear of punishment for disobedience is. As outlined in Proposition 1, inducing obedience requires the Leader to order harsher repression, but because this is costly, it becomes more likely that L prefers to order no repression.

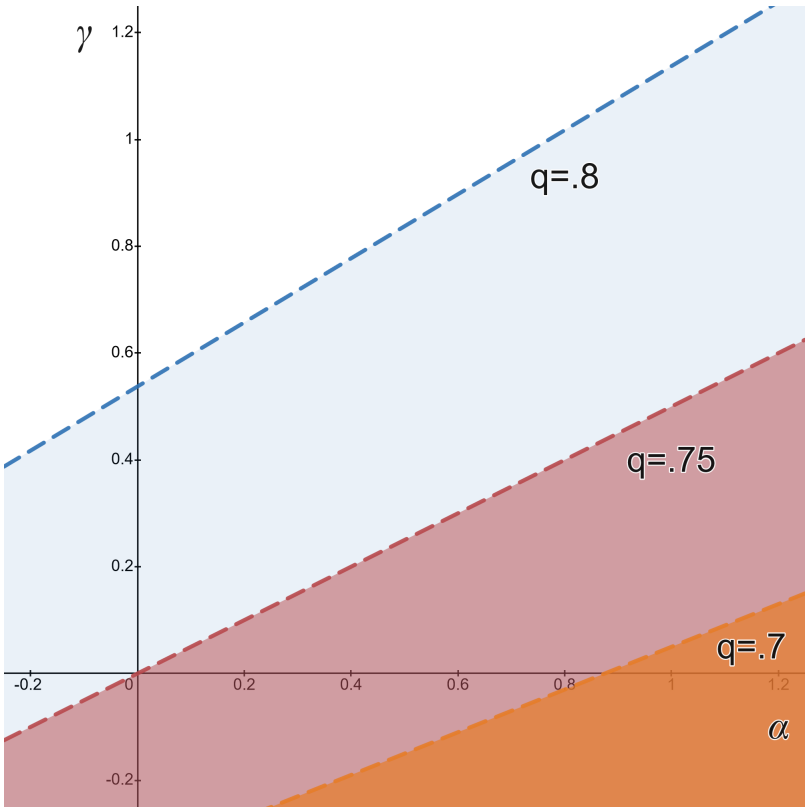


Figure 3: Shifting q, probability of regime survival with rank-and-file obedience

Figure 4 shows how the area of no repression (the shaded region) expands in response to increasing T, transitional justice against L. The values of previously fixed parameters remain unchanged (including three regions representing different values of q), except that T is now set higher relative to Figure 3.

In Figure 4a, we set $T = 2$, causing the orange region (corresponding to less dependency of the regime on the rank-an-file because of the lower q of .7) to shift downward and the blue

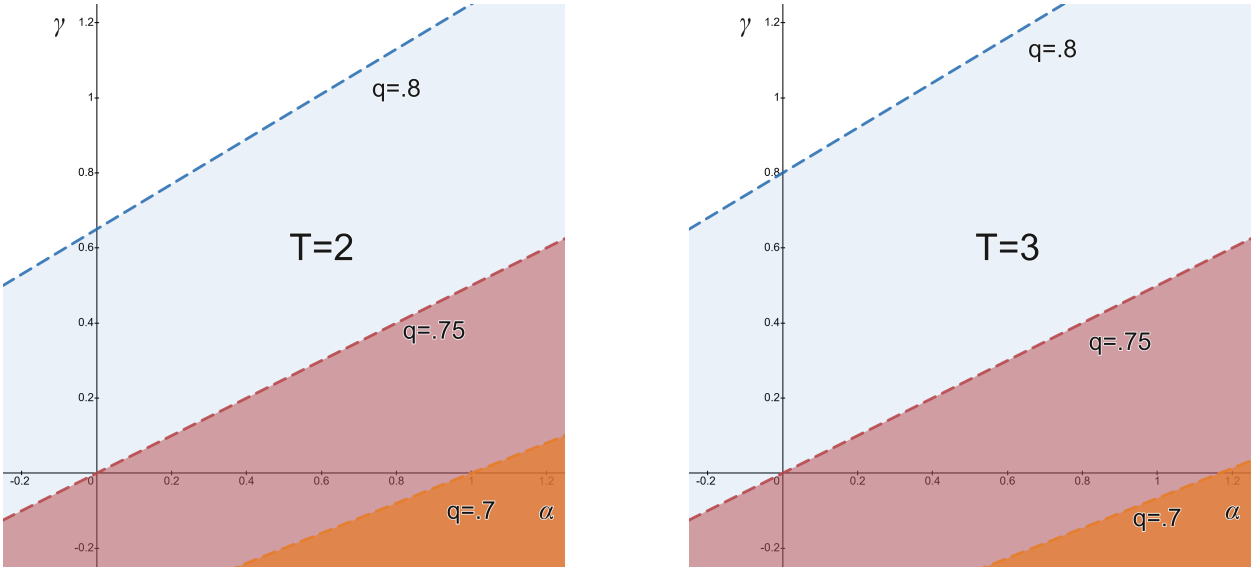
region (corresponding to more dependency on the rank-an-file because of the higher q of .8) to shift upward. The red region (corresponding to moderate dependency on the rank-and-file with $q = .75$) appears unchanged.

Figure 4b shows that when T increases to 3, this divergence becomes even more pronounced. Substantively: as the severity of transitional justice against leadership increases, the zero repression condition can become more *or* less likely to be satisfied, depending on $q - p$, how reliant the regime is on the rank-and-file. As T increases, the lower dependency region (in orange) shrinks, while the higher dependency region expands. Since these regions correspond to no repression being ordered, the divergence observed in the graphs has a substantive empirical interpretation. Namely, increasing transitional justice against authoritarian leaders makes regimes less dependent on rank-and-file order less repressive, while regimes less dependent on the rank-and-file become more repressive.

Moreover, the regions' divergence suggests that there is a level of dependence (represented as a value of $q-p$) at which the corresponding region is stationary in T, that is, changing T does not affect it. What is now but a conjecture, is corroborated by the fact that the high dependence (blue) and low dependence (orange) regions move in opposite directions while the red (moderate dependence) seems unchanged. The specific values of $q - p$ at which T is stationary remain to be derived analytically.

Next, we illustrate comparative statics of t, the career prospects of the rank-and-file. Figure 5 fixes parameters of the model at $\epsilon = .5$, $T = 1.25$, $p = .5$, and $q = .75$. The shaded no repression regions correspond to three different values of t ($t = .6, t = .5$ and $t = .4$, from lightest to darkest in the plot). We see that t increases, the no repression region expands. Substantively: as the rank-and-file's prospects under a new regime improve, the more likely it is that the leader's best response is to order no repression.

Finally, we look at the effect of moral reservations of the rank-and-file. In Figure 6 we fix the parameters of the model at values $T = 1.5$, $t = .5$, $p = .5$, and $q = .75$. The various shades of no repression regions correspond to three different values of ϵ (.45, .5 and



(a) $T = 2$ (b) $T = 3$

Figure 4: Shifting T, the severity of punishment against L by a new democratic regime

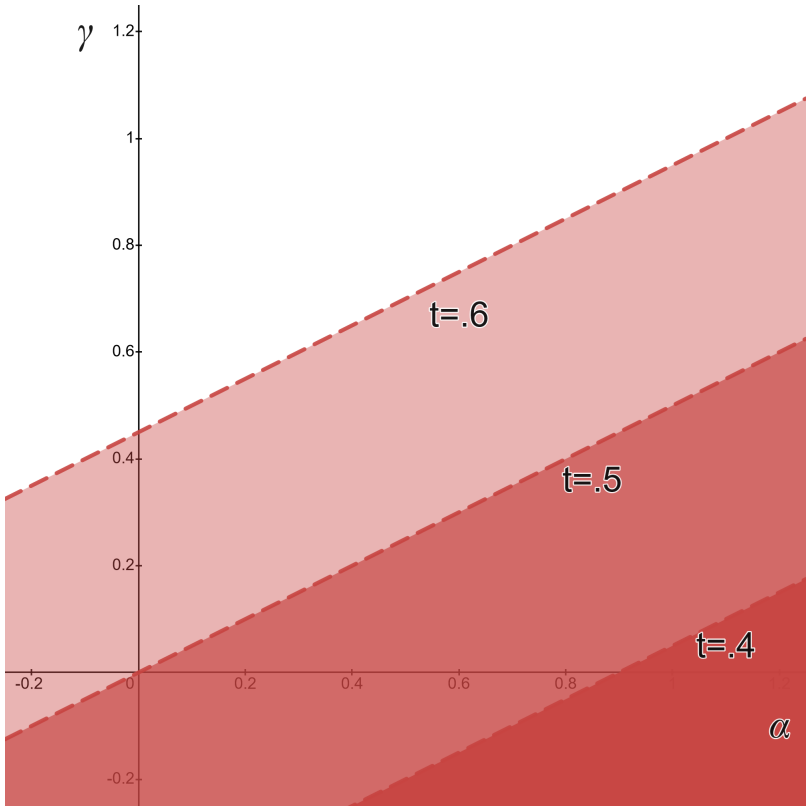


Figure 5: Shifting t, F's prospects under a new democratic regime

.55 from darkest to lightest). As ϵ increases, the no-repression region expands. This effect in similar to the effect of increasing t. Substantively: as the rank-and-file's moral reservations

about executing orders of repression increase, the Leader's best response is more likely to be no repression. This result is supported by recent empirical work on protesters' use of "nonviolence and fraternization" with the military's rank-and-file during Hirak protests in Algeria (GREWAL, 2023).

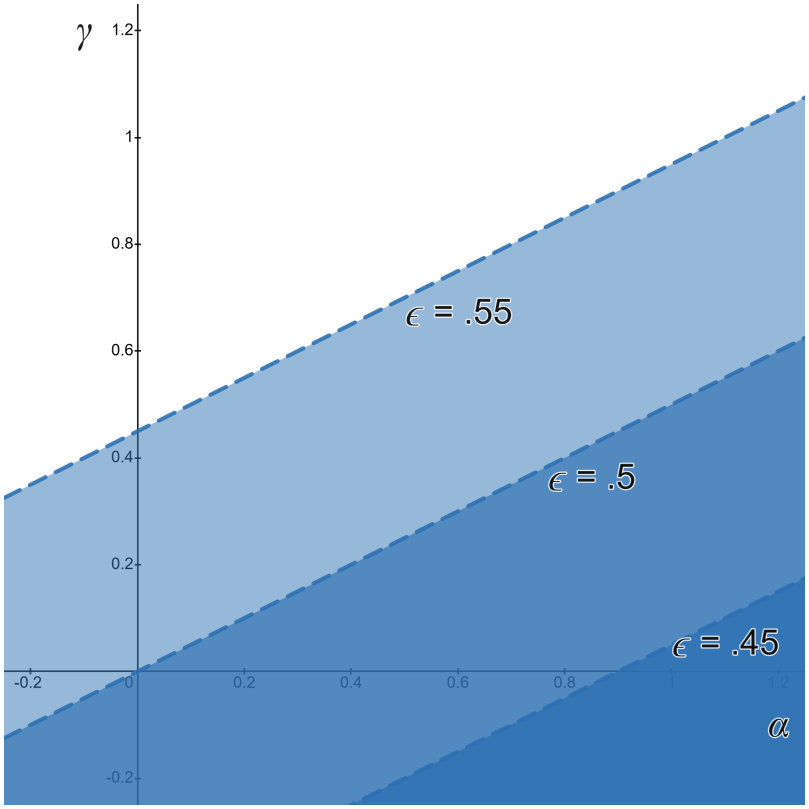


Figure 6: Shifting ϵ, F's moral cost of following the order to repress

2.4 Empirical Predictions

The goal of our theory is to reconstruct the decision calculus of agents of repression in an authoritarian regime facing the prospect of transitional justice. The final model systematically and rigorously presented the trade-offs facing the rank-and-file, who want to be on the winning side of a contest for power, and the leadership, which wants to ensure the regime's survival. It is safe to assume that truly authoritarian elites—elites that cannot reinvent themselves to flourish in democratic conditions (Grzymala-Busse, 2002)—prefer the regime

survives. But without the participation of the rank-and-file, their attempt to stave off a democratic challenger is less likely to succeed.

The main conclusions from our theory can be summarized as follows:

1. As the regime's dependence on the rank-and-file increases, repression becomes less likely. Recall, that dependence on the rank-and-file is interpreted as the difference in the probability of regime survival when the rank-and-file obeys the order to repress relative to when they disobey the order. Hence, dependence itself increases when the rank-and-file's repression becomes more effective, when the regime becomes weaker, or both.

2. The effect of severity of transitional justice against Leader increases on the likelihood of repression depends on how reliant the regime is on the rank-and file. Increasing transitional justice against Leaders makes repression less likely in highly dependent regimes, but more likely in less dependent regimes.

3. As the rank-and-file's career prospects under a new democratic regime improve, repression becomes less likely.

2.5 Empirical data

To test the empirical implications of the our theory, we plan to use forthcoming data from the the Global Transitional Justice Database (GTJD). Produced by the Transitional Justice and Democratic Stability Lab, the second release of the database will add criminal trial events to the existing data on personnel transitional justice events (Bates, Cinar and Nalepa, 2020).

The criminal trials data, which is scheduled for release in the fall of 2023, will cover 98 countries whose authoritarian period ended between 1918 and 2020. For each country, lab members use a combination of electronic databases of primary sources and numerous secondary sources, ranging from monographs and chapters in edited volumes to articles in peer-reviewed journals in social science, to create a chronology document. The documents

include each transitional justice event in chronological order, noting the date, a brief identification of the event, the relevant state and non-state actors, a more detailed description of the event, and the data source. After the chronologies are created, each criminal trial event is independently coded by two lab members.

Each event is coded along two dimensions that are relevant to this project. First, each event is coded as *positive* or *negative*, reflecting whether it advances or hinders accountability for perpetrators of human rights violations and atrocities. Positive events include indictments, arrests, convictions, and the upholding of convictions on appeal. Negative events include amnesties/immunities, acquittals, appeals, and overturning of convictions on appeal. Steps moving legislation enabling criminal prosecutions forward also count as positive events. These include considering such legislation on the floor of the legislature, the passage of such legislation by an assembly vote, supreme court decisions upholding the constitutionality of such legislation, or the overturning a presidential veto against such legislation. Negative events, in contrast, include voting down, vetoing, or declaring a bill enabling prosecutions unconstitutional. The fate of amnesty bills is coded the exact opposite way: legislation moving amnesties forward is considered negative, while events halting the legislative process of amnesties is positive. The lifting of a criminal statute of limitations is positive, while preventing the lifting of such a statute is negative.

An innovation of the expansion of the GTJD to criminal trials is a second coding dimension that classifies whether the individuals involved were *leadership* or *rank-and-file* (which we may interpret as Followers vs. Leaders). These categories are not exclusive, as *both* leaders and rank-and-file can be implicated in the same event. Events that concern the legislative process or events where the rank of individuals involved is unclear, are coded *neither* (but still classified as positive or negative). The number of positive and negative transitional justice events is then counted to create an annual panel, with countries as the cross section and time relative to the transition year as the temporal dimension. A panel assembled in this way allows for the creation of many different measures of transitional justice.

At the country level, the data allows us to examine trends in positive and negative events over time. Figure 7 presents preliminary criminal trials data for a single country (Poland) beginning in the year of its transition. The first positive events include the reopening of cases against perpetrators of two egregious crimes in the late 80's: the murder of a Catholic priest, Father Popieluszko, and the pacification of the coal miner strike in Lower Silesia during Martial Law. The first trial ended in conviction, the second came to a halt: those who ordered the killings of coal miners were excused due to failing health of the defendants. The rank-and-file members of the special-ops unit maintained a code of silence familiar to students of police brutality in the US. Although the deadly shots could be linked to specific weapons, the servicemen insisted that out of urgency, they had not used their own weapons. Since it was impossible to assign individual responsibility for the murders, the average sentences were only for 3.5 prison years. Another fraught trial was that of Stalinist torturer Adam Humer, stayed for health reasons in 1994 though eventually resulting in a sentence in 1996. In 1995, in the Polish version of Nuremberg, 12 order-givers responsible for shooting workers of the Gdansk shipyard in 1970 stood trial. After four years in court devoted to 18,000 pages of evidence, nobody was convicted because the trial had been postponed so many times on account of defendants' poor health or because like Czeslaw Kiszczak, former Secret Police chief, they were simultaneously standing trial for the implementation of Martial Law. Kisczak was ultimately acquitted of that last crime in 1998. Eventually, after 12 years of trying, prosecutors learned to build cases that would survive doctors' notes and codes of silence. The first swiftly prosecuted case, in 2001 was against Jozef Mania and was prepared individually by the Institute of National Remembrance even though Mania was one of eight living perpetrators being charged for their role in an extermination program.

As mentioned above, the panel structure also allows the construction of interesting cross-sectional measures. For example, we can construct a measure of criminal trial event "severity" similar to the measure of personnel transitional justice severity in Bates, Cinar, and Nalepa (2020). Figure 8 plots the severity of rank-and-file and leadership criminal trial

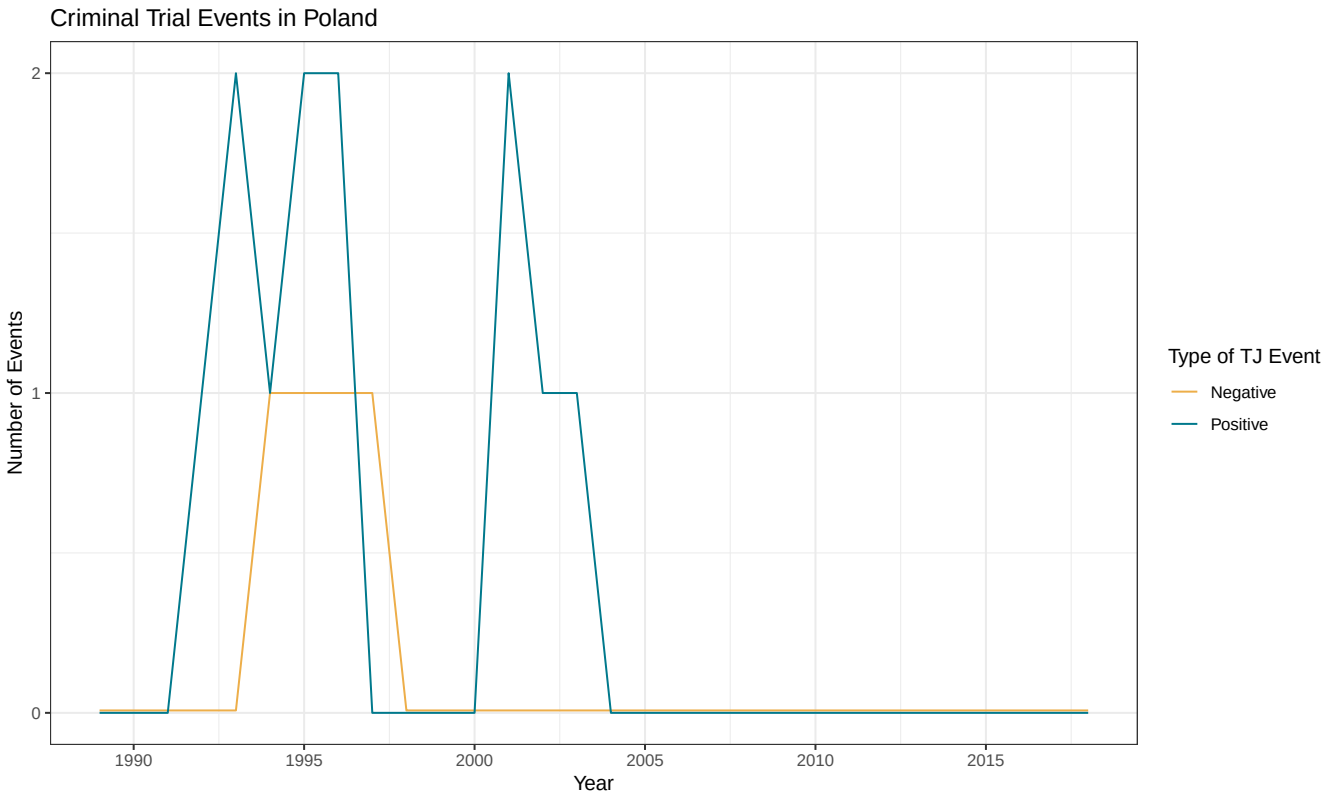


Figure 7: An example of preliminary criminal trial events data

events for a selection of countries. In this case, severity represents the ratio of positive events to total (positive and negative) events aggregated across all years in the country's most recent democratic period.

Our other ongoing data collection effort will expand the time period of *both* the existing personnel TJ data and the criminal trials data to the 10 years *preceding* a country's most recent transition to democracy.

The impetus for expanding the data to include a pre-transition phase is to try to quantify the extent to which regimes punish their own agents of repression prior to transition, a crucial element of the theory presented above. The combination of these two updates to the Global Transitional Justice Database will enable empirical testing of some of the model's empirical implications, including the relationship between severity of punishment against rank-and-file before and after transition, repression, and dependence on the regime. We have not yet devised specific empirical tests or statistical models, but welcome suggestions using the

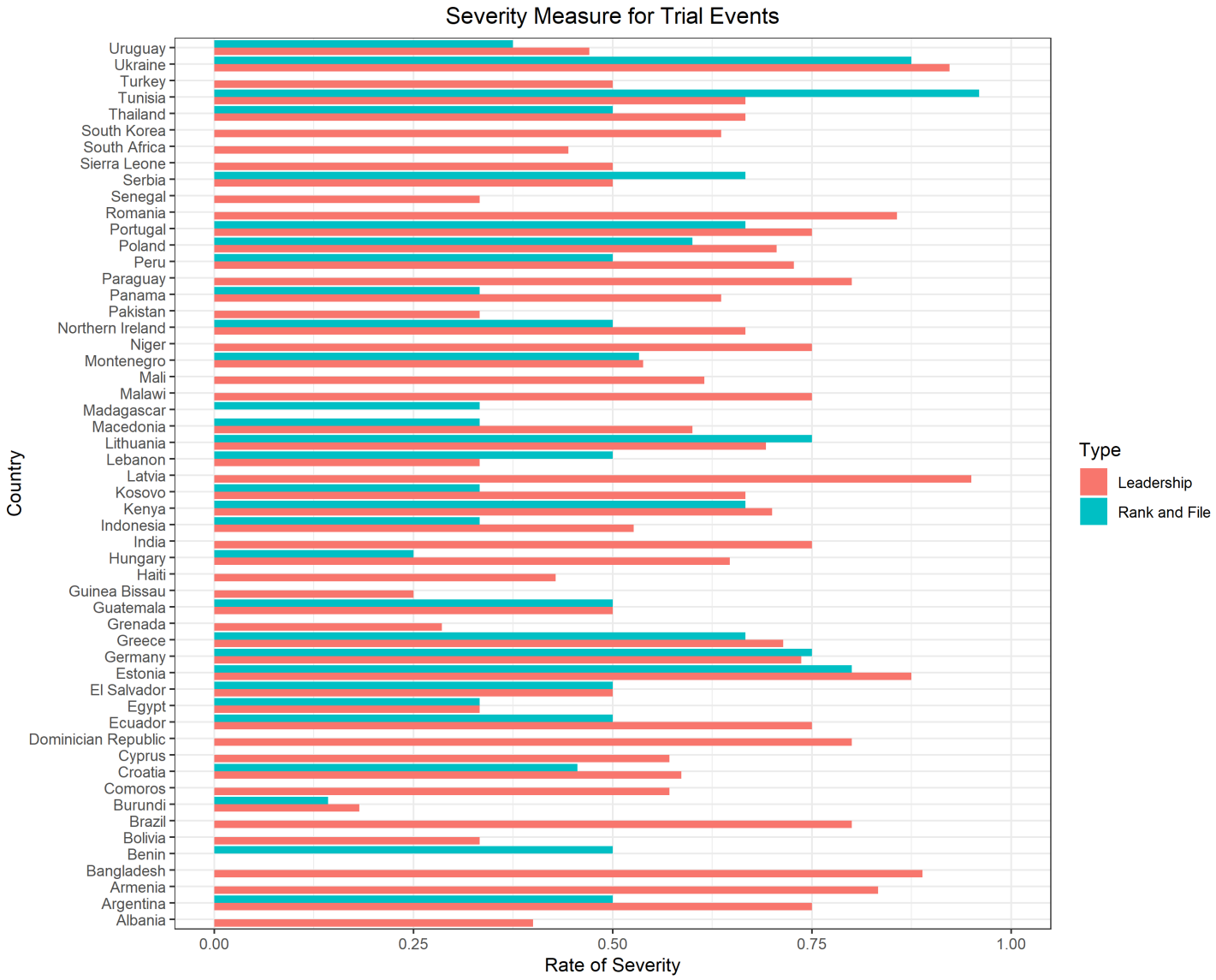


Figure 8: An example of a cross-sectional measure of TJ using preliminary criminal trials data

forthcoming data described here or alternative sources of data.

3 Conclusion

Societies transitioning from authoritarianism and civil war to a democratic order have to reckon with members and leaders of their repressive apparatus. Former agents of repression can either be integrated back into the new democratic state, following extensive vetting

(Nalepa, 2010*a*; Stan, 2013) or they can be prosecuted.[7] Yet even the decision to prosecute is not a straightforward one. In this paper, we have focused on the dynamics that emerge when leaders and rank-and-file members of a regime's authoritarian regime face the prospect of regime change and transitional justice.

We have proposed a formal model that captures the decision calculus of leadership and rank-and-file members of the authoritarian repressive apparatus. Leaders' repressive orders and followers' obedience impact their fates, but also influence regime survival. Our models' parameters correspond to key arguments that have been made in support of wide-spread prosecutions of former perpetrators of human rights abuses, as well as the key arguments made against such prosecution.

The reason to prosecute agents of repression is straightforward: it discourages followers from obeying such orders, which in turn discourages leaders from issuing them in the first place.

But there might be reasons for new democracies to exercise some restraint when it comes to punishing former rank-and-file members of the regime. The rank-and-file have to weigh the expected costs of following repressive orders against their ethical concerns and (crucially) potential punishment by an autocrat if the regime survives without the rank-and-file's obedience. Because the rank-and-file face punishment on two fronts, conditional on their own behavior and who holds power, harsher transitional justice against them can make it more likely they will follow the regime's order to repress.

The solution to our first model, in which the rank-and-file's obedience is necessary and sufficient for regime survival, suggests that repressive orders will never be followed and therefore never be issued. As scholars of transitional justice, we know that this is untrue: orders to repress are issued and sometimes followed, suggesting that we need a more sophisticated version of the model. In the second and third models, we relax the assumption that the rank-and-file's obedience fully determines the regime's survival. Introducing a stochastic el-

[7] Technically, they can also be purged and forced to seek employment elsewhere, but this strategy comes with other risks, such as increased levels of organized crime (Bates et al., 2020*b*; Lessing, 2017).

ement to the game yields equilibria in which a Follower only obeys orders of repression above a critical value of harshness. The Leader only issues the order to repress in equilibrium if their ability to punish disobedience is sufficiently high. This is less likely to happen as the severity of transitional justice against the rank-and-file increases.

This finding microfounds theoretically what many scholars of democratic transitions, from Huntington (1993) to Vinjamuri and Snyder (2004) have argued all along: that harsh transitional justice can play into the hands of autocrats. At a practical level, it also explains why justice NGOs and INGOS will always remain dissatisfied with the level of criminal transitional justice that is observed around the world. While we share the sentiment behind their lamenting of impunity for those who "killed with their own hands" [p. 114] (Todd et al., 2000*a*) in the name of the dictatorship they worked for, to insist on harsher prosecutions may threaten the possibility of rank-and-file support for democratic challengers. This does not mean we have to throw the baby out with the bathwater and eschew all forms of transitional justice. Even in this framework where punishing Followers can push them towards cooperation with the autocrat rather than the democratic challenger, harsh transitional justice of Leaders remains an effective deterrent.

References

AP. 1980. "Chile Holds Policemen for Nine Kidnappings and a Beating Death.".

Bates, Genevieve. 2020. "Negotiated Justice and International Accountability: The ICC and Transitional Justice During Peace Negotiations.".

Bates, Genevieve. 2021. Holding Their Feet to the Fire: Negotiated Accountability in the Shadow of the International Community PhD thesis The University of Chicago.

Bates, Genevieve, Ipek Cinar and Monika Nalepa. 2020. "Accountability by numbers: A new global transitional justice dataset (1946–2016)." *Perspectives on politics* 18(1):161–184.

Bates, Genevieve, Ipek Cinar, Monika Nalepa and Evgenia Olimpieva. 2020*a*. "What is the effect of Personnel Transitional Justice on Crime?".

Bates, Genevieve, Ipek Cinar, Monika Nalepa and Evgenia Olimpieva. 2020*b*. "What is the effect of Personnel Transitional Justice on Crime?".

Coppedge, Michael. 2021. "V-Dem Dataset 2021.".
URL: *https://www.v-dem.net/dsarchive.html*

Dragu, Tiberiu and Yonatan Lupu. 2018. "Collective action and constraints on repression at the endgame." *Comparative Political Studies* 51(8):1042–1073.

Duursma, Allard and Tanja R Müller. 2019. "The ICC indictment against Al-Bashir and its repercussions for peacekeeping and humanitarian operations in Darfur." *Third World Quarterly* 40(5):890–907.

Escribà-Folch, Abel and Joseph Wright. 2015. "Human rights prosecutions and autocratic survival." *International Organization* 69(2):343–373.

Geddes, Barbara, Joseph Wright and Erica Frantz. 2014. "Autocratic breakdown and regime transitions: A new data set." *Perspectives on Politics* pp. 313–331.

Greitens, Sheena Chestnut. 2016. *Dictators and their secret police: Coercive institutions and state violence*. Cambridge University Press.

GREWAL, SHARAN. 2023. "Military Repression and Restraint in Algeria." *American Political Science Review* p. 1–16.

Grzymala-Busse, Anna M. 2002. *Redeeming the communist past: The regeneration of communist parties in East Central Europe*. Cambridge University Press.

Harkness, Kristen A. 2018. *When soldiers rebel: Ethnic armies and political instability in Africa*. Cambridge University Press.

Helmke, Gretchen. 2012. *Courts under constraints: judges, generals, and presidents in Argentina*. Cambridge University Press.

Hoyle, Justin A. 2019. "To govern, or not to govern? Opportunity and post-coup military behaviour in Egypt 2011–2014." *Democratization* 26(6):993–1010.

Huntington, Samuel P. 1993. *The third wave: Democratization in the late twentieth century*. Vol. 4 University of Oklahoma press.

Kim, Hunjoon and Kathryn Sikkink. 2010. "Explaining the deterrence effect of human rights prosecutions for transitional countries." *International Studies Quarterly* 54(4):939–963.

Krcmaric, Daniel. 2018. "Should I stay or should I go? Leaders, exile, and the dilemmas of international justice." *American Journal of Political Science* 62(2):486–498.

Krcmaric, Daniel. 2020. *The justice dilemma: Leaders and exile in an era of accountability*. Cornell University Press.

Lessing, Benjamin. 2017. "Counterproductive punishment: How prison gangs undermine state authority." *Rationality and Society* 29(3):257–297.

Mallinder, Louise. 2008. *Amnesty, human rights and political transitions: bridging the peace and justice divide*. Bloomsbury Publishing.

Masri, Safwan M. 2017. *Tunisia: an Arab anomaly*. Columbia University Press.

McAllister, Jacqueline R. 2019. "Deterring wartime atrocities: Hard lessons from the Yugoslav tribunal." *International Security* 44(3):84–128.

Meng, Anne and Jack Paine. 2022. "Power sharing and authoritarian stability: How rebel regimes solve the guardianship dilemma." *American Political Science Review* 116(4):1208–1225.

Nalepa, Monika. 2010*a*. "Captured commitments: an analytic narrative of transitions with transitional justice." *World Politics* 62(2):341–380.

Nalepa, Monika. 2010*b*. *Skeletons in the closet: Transitional justice in post-communist Europe*. Cambridge Studies in Comparative Politics Cambridge University Press.

Nino, Carlos Santiago. 1996. *Radical evil on trial*. Yale University Press.

Sikkink, Kathryn. 2011. *The Justice Cascade: How Human Rights Prosecutions Are Changing World Politics (The Norton Series in World Politics)*. WW Norton & Company.

Singh, Naunihal. 2014. *Seizing power: The strategic logic of military coups*. JHU Press.

Stan, Lavinia. 2013. "Reckoning with the Communist past in Romania: A scorecard." *Europe-Asia Studies* 65(1):127–146.

Todd, Stephen Charles et al. 2000*a*. *Lysias: Against Erastothenes*. Vol. 2 University of Texas Press.

Todd, Stephen Charles et al. 2000*b*. *Lysias: Against Manthitheus*. Vol. 2 University of Texas Press.

Trejo, Guillermo, Juan Albarracín and Lucía Tiscornia. 2018. "Breaking state impunity in post-authoritarian regimes: Why transitional justice processes deter criminal violence in new democracies." *Journal of Peace Research* 55(6):787–809.

Tyson, Scott A. 2016. "The agency problem underlying the use of repression.".

Varese, Federico. 2001. *The Russian Mafia: private protection in a new market economy.* OUP Oxford.

Vinjamuri, Leslie and Jack Snyder. 2004. "Advocacy and scholarship in the study of international war crime tribunals and transitional justice." *Annu. Rev. Polit. Sci.* 7:345–362.

Vinjamuri, Leslie and Jack Snyder. 2015. "Law and politics in transitional justice." *Annual Review of Political Science* 18:303–327.

Volkov, Vadim. 2016. *Violent entrepreneurs: The use of force in the making of Russian capitalism.* Cornell University Press.

A Formal Appendix

A.1 Formal Description of "Sufficiency" Game

- **Players**: Leader (L); Follower (F); nature (N)
- **Timing**: L moves in period 1; F moves in period 2; N moves in period 3 (provided the Follower refused)
- **Actions**: L orders a level of repression $R \in [0, 1]$; F chooses to disobey ($o = 0$) or obey ($o = 1$); if F chooses $o = 1$, the regime survives and the game ends; if T chooses $o = 0$, N determines whether the regime survives according to the following probabilities:
  1. $Pr(survive|o = 0) = p$
  2. $Pr(transition|o = 0) = 1 - p$
- **L's payoffs**: L's payoff only depends on the regime's survival, not on F's decision to obey:
  1. If the regime survives, L receives $1 - R$
  2. If the regime transitions, L receives $-RT$
- **F's payoffs**: F's payoff is conditional on the regime's survival *and* F's decision to obey the order:
  1. If F disobeys the order to repress and the regime survives, F receives $-\gamma R$
  2. If F disobeys the order to repress and the regime transitions, T receives t
  3. If F obeys the order to repress and the regime survives, F receives $1 - \epsilon$

A.2 Formal Description of "Sufficiency and Necessity" Game

A.2.1 The Game

- **Players**: Leader (L); follower (F); nature (N)
- **Timing**: L moves in period 1; F moves in period 2; N moves in period 3
- **Actions**: L orders a level of repression $R \in [0, 1]$; F chooses to disobey ($o = 0$) or obey ($o = 1$); N determines whether the regime survives or transitions, according to the following probabilities:
  1. $Pr(survive|o = 0) = p$
  2. $Pr(transition|o = 0) = 1 - p$
  3. $Pr(survive|o = 1) = q$
  4. $Pr(transition|o = 1) = 1 - q$

where $q > p$ (i.e. the regime is more likely to survive if F obeys the order to repress than if F ignores the order)

- **L's payoffs**: L's payoffs only depend on the regime's survival, not on F's decision to obey:
  1. If the regime survives, L receives $1 - R$
  2. If the regime transitions, G receives $-RT$
- **F's payoffs**: F's payoffs is conditional on the regime's survival *and* F's decision to repress or not:
  1. If F disobeys the order to repress and the regime survives, T receives $-\gamma R$
  2. If F disobeys the order to repress and the regime transitions, T receives t
  3. If F obeys the order to repress and the regime survives, T receives $1 - \epsilon$
  4. If F obeys the order to repress and the regime transitions, T receives $-\alpha R - \epsilon$