Mass Purges: Top-down Accountability in Autocracy^{*}

B. Pablo Montagnes[†] Stephane Wolton[‡]

This version: November 7, 2017 Most recent version

Abstract

This paper proposes a novel theoretical framework to study mass purges in authoritarian regimes. We contend that mass purges are an instrument of top-down accountability meant to motivate and screen a multitude of agents (e.g., single-party members, state bureaucrats). We establish that the nature of purges depends on the intensity of violence. Mild purges are discriminate (the set of purged agents is well delineated), whereas violent purges are semiindiscriminate (all subordinates risk being purged). The breadth of the purge, in turn, is non-monotonic in violence. We further uncover that the autocrat faces a trade-off between fear and love as greater intensity of violence increases performance, but worsens selection. Even absent de jure checks, the autocrat is de facto constrained in her actions by her subordinates' strategic behavior. We use our theoretical findings to reassess historical (i.a., the Soviet purges, the Cultural Revolution) and recent (the Erdogan purge) events.

[†]Department of Political Science, Emory University. Email: pablo.montagnes@emory.edu [‡]Department of Government, London School of Economics. Email: S.Wolton@lse.ac.uk

^{*}We thank Scott Ashworth, Brett Benson, Dan Bernhardt, Alessandra Casella, Ali Cirone, Torun Dewan, Tiberiu Dragu, Scott Gehlbach, Thomas Groll, Haifeng Huang, Navin Kartik, Roger Myerson, Salvatore Nunnari, Carlo Prato, Arturas Rozenas, Milan Svolik, Scott Tyson, and conference and seminar participants at the Sixth LSE-NYU Conference, Priorat Workshop in Theoretical Political Science, the Third Formal Theory & Comparative Politics Conference at Chicago, 3rd PECO Conference in Washington, D.C., Bocconi University, Columbia University, Vanderbilt University, Warwick University, and King's College London for helpful comments and suggestions. All remaining errors are the authors' responsibility.

Party struggles lend a party strength and vitality; the greatest proof of a party's weakness is its diffuseness and the blurring of clear demarcations; a party becomes stronger by purging itself

From a letter of Lassalle to Marx, June 24, 1852

In 1901, when writing his revolutionary agenda *What is to be done?*, Vladimir Ilyich Ulyanov (alias Lenin) chose one particular sentence as an epigraph. This sentence, reproduced above, calls for the use of purges to shape the membership of communist parties. Often repeated (e.g., Stalin at the 13th Party Congress, the *Pravda* in 1949 as documented by Brzezinski, 1956), it was used to justify the many purges of rank-and-file party members experienced by the Communist Party of the Soviet Union (CPSU). Purges were judged so essential to the good functioning of communist parties that they became a condition of admission to the Communist International: "[t]he communist parties (...) must from time to time undertake purges (re-registration) of the membership of their party organizations in order to cleanse the party systematically of the pettybourgeois elements within it" (thirteenth condition). Far from being limited to the Union of Soviet Socialist Republics (USSR), purges occurred in various forms in Communist China, Nazi Germany, Fascist Italy, and other authoritarian or would-be authoritarian regimes (e.g., Iraq in 1979, Syria in 1980-4, Turkey in 2016-7).

What are these purges Lenin advocates? What is their purpose? Why were they deemed vital by communist leaders? This paper contends that these purges are an instrument of 'top-down accountability,' a response to the autocrat's political problem of motivating rank and files and selecting the most suitable subordinates within the single-party or state bureaucracy. Affecting thousands, if not millions of individuals, these mass purges further take the form of 'many-to-one accountability.'

This paper provides a theoretical framework to understand mass purges, and by extension, many-to-one accountability problems, when the autocrat can only screen her agents using coarse information (e.g., success or failure in meeting a target). Our modeling approach incorporates various motives for the purge—screening opportunists, saboteurs, or opponents—and applies to diverse settings reviewed in the next section. As a caveat, however, it does not cover an autocrats' pure and perfect selection problem when she purges her subordinates based on observable traits such as religion or ethnicity (e.g., the Law for the Restoration of the Professional Civil Service passed in 1933 in Germany or the Sunni military members in Syria in 1966, see Van Dam 2011).

We further use mass purges to highlight one critical characteristic of autocracies, namely the 'ever-present possibility of violence.' Unlike democratic leaders who have limited reward and punishment mechanisms available to motivate subordinates, autocrats face few constitutional constraints. We study whether the absence of de jure checks on her actions converts into de facto absolute power for the autocrat.

Three features characterize mass purges: (i) their breadth—the proportion of subordinates purged—, (ii) their nature—who gets purged—, and (iii) the intensity of violence—the cost of being purged. All these aspects of purges are endogenous in our theory. We show that violence and nature are linked in equilibrium. Discriminate purges (where the set of purged agents is clearly delineated) tend to be relatively mild; semi-indiscriminate purges (where no performance indicator guarantees safety) are violent. Breadth, on the other hand, does not vary monotonically with the intensity of violence or its nature. Further, even though the autocrat faces no de jure check on her choice of violence, de facto she is constrained by her subordinates' strategic behavior. Violence generates an inescapable trade-off between fear (better performance) and love (worse selection).

Our baseline framework comprises an autocrat and a mass of subordinates which can either be congruent (sharing the autocrat's ideological objectives or appealing to her predilections) or noncongruent (opportunists who only care about the benefits reserved to regime insiders or possible opponents). Each agent exerts effort on an individual project and can be successful (e.g., meet the quotas) or unsuccessful. The autocrat only observes the outcome of each subordinate's project. She then chooses a purge breadth. The autocrat also picks the intensity of violence. All purged agents are replaced by new subordinates from an available replacement pool (e.g., the pool of candidates to the party in the USSR and Communist China).

Even though subordinates work on independent tasks, their fate is linked through the purge. An agent makes an inference on his probability of being purged based on his own effort and the performance of all other subordinates. Greater risk of being purged after failure always increases effort; greater risk of being purged after success always decreases effort (since success becomes less valuable).

The autocrat carries out a purge to increase congruence among subordinates. Congruent agents who share the autocrat's predilection always exert more effort and are more likely to succeed. Success is a positive, but noisy signal of congruence. Therefore, the autocrat first targets subordinates who fail in their project, and targets successful agents only after the failure pool has been exhausted. Nonetheless, the screening is never perfect. In any purge, some congruent agents are eliminated, whereas some non-congruent agents survive.

The autocrat's incentive to purge is a function of the informativeness of different performance indicators. Any change in agents' efforts affects the autocrat's evaluation of success and failure. We term this strategic interdependency the 'pool makeup effect,' which we define as positive when the target pool becomes more tainted and the autocrat has increased incentive to purge. We show that the risk of being purged is self-reinforcing. Agents' responses to anticipated increased purge breadth generates a weakly positive pool makeup effect, which leads to greater incentive to purge for the autocrat.

The autocrat also chooses the intensity of violence which affects both agents' effort, what we term the 'fear effect,' and the future composition of rank and files, what we call the 'love effect.' The fear effect is positive: higher intensity of violence always improves present performance by increasing the cost of being purged. This general increase in effort, however, implies that an agent's accomplishment is less informative about his ideological congruence, and the autocrat is never certain what lurks behind the mask of conformity (Dallin and Breslauer, 1970). Heightened violence may thus worsen selection, and the love effect can be negative. The autocrat therefore often faces a trade-off between short-term (subordinates' performance) and long-term (subordinates' congruence) benefits of the purge.

The fear and love effects also have important consequences for other equilibrium quantities. By reducing the proportion of failures, fear makes it less costly to exhaust this pool. In addition, the negative love effect of non-congruent agents entering the success pool renders the latter more tainted. While the (marginal) benefit of purging successful agents is not worth the (marginal) cost at a relatively low intensity of violence, the opposite can hold true at a high intensity. Discriminate purges—only failures are purged—tend to be mild, whereas semi-indiscriminate purges—all failures, but also some successful agents are purged—are violent.

Importantly, the fear effect differentially motivates congruent and non-congruent subordinates. Due to their alignment with the autocrat, congruent agents have more to gain from surviving the purge and are always more sensitive to changes in the risk of being purged. In a partially discriminate purge (not all failures are purged), greater intensity of violence heightens the risk associated with failure as effort increases and the failure pool becomes thinner. Thus, congruent subordinates increase their effort more than non-congruent ones, the failure pool becomes more tainted, and breadth increases with violence. In a semi-indiscriminate purge, greater intensity of violence reduces the benefit of success and thus agents' incentives to exert effort. Congruent subordinates increase their effort less, the success pool becomes more tainted, and breadth increases with violence. However, the success and failure pools are qualitatively different. Discriminate purges do not automatically become semi-indiscriminate when the failure pool is exhausted. For some intermediate intensity of violence, the autocrat purges all failures, nothing less, nothing more. Since the fear effect reduces the proportion of agents who fail their project, the breadth decreases with violence in this range. Overall, purge breadth is non-monotonic in the intensity of violence.

Our theoretical results predict a clear relationship between the nature and violence of purges, whereas no such pattern emerges for breadth and violence. While no large-N study of mass purges exists, historical evidence offers a way to assess our predictions. Our findings are consistent with both the differences and similarities between the Maoist purges of the fifties and Stalinist purges of the thirties. As Teiwes (1993, 25-27) describes, Chinese rectification campaigns were less violent and more discriminate than Stalinist purges when "flouting commands court danger, but even enthusiastic compliance is no guarantee of safety." The proportion of party members purged, however, appears to have been relatively similar in both countries (see Getty, 1987; Teiwes, 1993).

Further, our model also offers an explanation for the significant variation in the features of purges not only across, but also within countries. In many-to-one accountability problems, individual agents' actions generate indirect spillovers modifying the risk of being purged and the informativeness of failure and success (the pool makeup effect). Consequently, small changes in economic or political conditions can have large repercussions for the autocrat's purging decisions, dramatically modifying the intensity of violence, nature, and breadth of purges.

Related Formal Literature

Our paper advances the idea that mass purges are a salient method of accountability in autocracy. In contrast, the literature has put great emphasis on targeted and small elite purges: the shaping of an autocrat's inner circle and contests for power. Bueno de Mesquita and Smith (2015) consider conditions under which a leader might purge the selectorate to increase his payoff and probability of survival. Svolik (2009) examines how autocrats can use elite purges to acquire more power. Acemoglu et al. (2008, 2009) show that an autocrat is unlikely to obtain absolute power through elite purge since members of the autocrat's winning coalition opposes too much purging as it increases the risk of their being removed in the future. Egorov and Sonin (2015) consider how elite purges in the present breed more elite purges in the future as winners of a contest for power anticipate the risk associated with sparing losing contenders. In turn, mass purges can complement power struggles (e.g., screening agents favorable to her opponents), but need not (e.g., cleansing of the single-party from opportunists). The critical aspect of our analysis is that the autocrat uses purges to motivate and select a *mass of subordinates*.

We thus provide a framework for analyzing many-to-one accountability problems. Dozens of studies focus on one-to-many accountability such as electoral accountability where one politician is accountable to many voters (see Ashworth, 2012, for a review). More recently, several papers have examined the specificity of one-to-one accountability. In democratic politics, Dewan and Myatt (2007, 2010) study a prime minister's choice of whether to retain or fire a minister following a scandal. Their analyses share some similarities with ours, including the effect of the quality of the replacement pool on the principal's decision.¹ Other works in this vein focus on the autocrat's (in)ability to hold her top officials accountable. Egorov and Sonin (2011) consider the optimal contract an autocrat can offer to her close circle to avoid betraval. They show that the autocrat can find it optimal to recruit incompetent viziers as it can be too expensive to induce loyalty from competent ones (see also Zakharov, 2016). Gehlbach and Simpser (2015) analyze how an autocrat can use and manipulate elections to increase bureaucrats' effort and the leader's chance of survival. Due to their focus on a single agent, these papers cannot identify the spillovers via change in the risk of being purged or the pool makeup effect which play a key role in our analysis. More closely related to the present study, Jiang et al. (2017) examine how promotions in hierarchies shape subordinates' incentives in a many-to-one setting. They assume that the principal has access to fine-grained information, a setting more suited to the analysis of small rather than large organizations.

Additionally, our paper also innovates by letting the autocrat choose the intensity of violence. A small literature examines the impact of violence on agents' behavior. In the context of industrial organization, Bernhardt and Mongrain (2010) shows how the threat of being fired can lead to overinvestment in firm-specific human capital. Bloch and Rao (2002) and Dal Bó et al. (2006) highlight in various bargaining settings how violence or the threat thereof increases a player's bargaining position and thus improves his payoff. Acemoglu and Wolitzky (2011) model slavery as a principalagent relationship where the slave-owner can reduce the value of her slave's outside option using coercion. Landa and Tyson (2017), in turn, highlight how coercive leadership is necessary to

¹Many models of bureaucracy can be interpreted as one-to-one accountability. See Gailmard and Patty (2012) for a review.

coordinate subordinates who are uncertain about their leader's preferences. All these works focus on various moral hazard problems, ignoring adverse selection, and so cannot examine the effect of violence on the screening of agents as in our paper.

1 Evidence on mass purges

Before proceeding, we summarize historical evidence on mass purges, highlighting their key characteristics. Due to the volume of secondary sources available, we primarily focus on the USSR and China, and discuss purges in other countries at the end of the section.

The Communist Party of the Soviet Union (CPSU) experienced mass purges ("chistka"—a sweeping, a cleansing) in 1919, 1921-23, 1924, 1928, 1929, 1931, 1933-34, 1935, 1936, 1949,1951-53, and 1971 (Brzezinski, 1956; Rigby, 1968; Schapiro, 1977; Getty, 1987). In China, mass purges ("quingchu"—to weed out) were part of the rectification campaigns (which included a reeducation component absent from Soviet purges) which happened under Mao in 1947-48, 1950, 1951-54, 1953, 1957, 1957-58, 1959-60, 1960-61, 1962-63, 1964-65. These lists, you will notice, do not include two prominent mass purges: the Great Terror in the USSR (1936-38) and the Chinese Cultural Revolution (1966-76). Due to their specificities, we return to these two events in Section 7 where we discuss the implications of our results.

The large number of purges in both countries is not accidental. Purges were a system of government (Brzezinski, 1956). Soviet leaders sought to regulate their periodicity (Getty, 1987). Not unlike elections, Mao prescribed the Chinese Communist Party go through rectification campaigns twice every five years (Teiwes, 1993).

The goals of these purges were twofold. First, they provided the necessary momentum to accomplish the grand designs of the totalitarian regimes (Brzezinski, 1956) and sustain a high level of activity (Teiwes, 1993). Second, they were meant to "cleanse the system in anticipation" (Brzezinski, 1956, 19). By removing a proportion of party members, mass purges allowed for the influx of new members (Brzezinski, 1956; Teiwes, 1993) drawn from the pool of candidates to the party (Rigby, 1968).

Mass purges did not target specific individuals; leaders did not have "detailed list[s] of individuals to be purged" (Weinberg, 1993, 23). All rank-and-file party members, *millions* of individuals (Teiwes, 1993), were affected. The most common accusation against purged members was one of opportunism, having joined the party for the social and economic benefits—reserved positions, special shops, etc.—associated with membership (Brzezinski, 1956; Getty, 1987; Teiwes, 1993)). Getty (1987, 38, emphasis in original) asserts that "[i]n the majority of purges, *political* crimes or deviations pertained to a *minority* of those expelled" from the CPSU. Rigby (1968), on the other hand, argues that the purges of the CPSU in the 1930s paved the way for the Great Terror and the show trials of 1936-38. Whatever the justification for the purges, congruence with the autocrat seems to have been difficult to observe. Communist leaders had to use indirect signals such as work performance. Any failure in a member's professional activity "automatically [became] a case of political accountability" (Brzezinski, 1956, 86).

Mass purges in other countries exhibit some similarities. In Nazi Germany, the Sturmabteilung (SA) was purged in 1935-6 following the Röhm putsch (or as Hohne, 1981, puts it, the putsch of Hitler against Röhm), the Schutzstaffel (SS) in 1933-35 (Hohne, 1981), and the party itself (NSDAP) in 1938 (Orlow, 1969). Lack of commitment to the Nazi cause rather than threat to the leadership again seems to have been the main driver of the purges. Even in the SA in the years following the Night of Long Knives, "nearly all of those expelled or punished in the SA's own internal purge were found guilty of moral failings" (Campbell, 1993, 660). In Fascist Italy, the party experienced a mass purge in 1931. Lack of ideological commitment and opportunism again seem to have been the primary cause of removal from the National Fascist Party (PNF) (Morgan, 2012). The Syrian Ba'ath party also removed from its ranks members and (principally) supporters who became "part of the regime, with the aim of profiting from material and other advantages which it supposedly offered" (Van Dam, 2011, 128). In contrast, Saddam Hussain seems to have used the purge of 1979 in Iraq to tighten his hold on power (Coughlin, 2005). Though relatively new members were targeted disproportionately, in none of these historical cases did the autocrat purge her agents based on directly observable traits. The authoritarian leader arguably had to rely on indirect indicators like in the framework we now describe.

2 Set-up

This section and the four that follow present and analyze in detail our theoretical framework as well as various extensions. All our results are summarized in Section 7 where we also discuss their application to historical and recent events. A non-technical reader may wish to proceed directly to Section 7 before returning to the formal part of the paper. We study a two-period $(t \in \{1,2\})$ model with an autocrat (A) and a [0,1] continuum of agents, indexed by the superscript *i*. Each agent is characterized by a type $\tau \in \{c, nc\}$, where $\tau = c$ denotes a congruent type and $\tau = nc$ non-congruent. In the context of communist purges, for example, non-congruent types correspond to opportunistic party members attracted to the party for its associated benefits (Getty, 1987), members who lacked "a wholehearted commitment to the Party's cause" (Teiwes, 1993, 114-115), or members not satisfied with the Party's line (Gregory, 2009). An agent's type is his private information, however it is common knowledge that there is a proportion λ of congruent types within current subordinates.

Each period, agent *i* exerts effort $e^i \in [0,1]$ on an individual project at cost $(e^i)^2/2$. The project can be successful (denoted $\omega = S$) or fail ($\omega = F$). The probability agent *i*'s project is successfully equals (without much loss of generality) e^i . While an agent's effort is not observed by the autocrat, the outcome of his project is (e.g., whether he has fulfilled his quota). This assumption corresponds to historical evidence that officials in charge of the purges had little information about local circumstances and could only judge according to how successful problem cases or certain projects were handled (Rigby, 1968; Teiwes, 1993).

After observing all project outcomes, the autocrat decides to purge a proportion κ_F (κ_S) of agents who failed (succeeded). Denoting the proportion of failures (successes) as α_F (α_S), the purge breadth then equals $\kappa = \alpha_F \times \kappa_F + \alpha_S \times \kappa_S$. Mass purges entail a loss in terms of human capital and organizational knowledge as well as the cost of potentially deporting agents or delay in finding suitable replacements for the purged subordinates. These costs are captured by the cost function $C(\kappa)$ with C(0) = 0 and (for ease of exposition) marginal cost $C'(\kappa) = C_0 + C_1 \times \kappa$, $C_0 \ge 0, C_1 > 0$. When a subordinate is purged, he is replaced by a new agent drawn from a pool of replacement (e.g., candidates for admission to the party in communist regimes). The proportion of congruent types among the replacement pool is r (we study an alternative set-up in which the quality of the replacement pool decreases with the purge breadth κ in Section 6).

Being purged has two distinct consequences for an agent. First, the agent loses the privilege associated with being a regime insider. Second, he suffers a direct loss L which corresponds to the 'intensity of violence' of the purge. The loss L can be relatively low if the agent is only fined or very large if the agent is killed, his or her spouse deported, and their children sent to orphanages as was commonplace in Stalin's USSR (Brzezinski, 1956; Conquest, 2008). The autocrat determines the intensity of violence at the beginning of the game (e.g., investment in the security apparatus) at a cost $\zeta(L)$ with $\zeta(0) = 0$ and marginal cost $\zeta'(L) = \zeta_0 + \zeta_1 L$, $\zeta_0 \ge 0$ and $\zeta_1 > 0$.

In period 1, an agent enjoys a benefit $R \ge 0$, which captures all the special privileges accorded to regime insiders in autocracies. In addition, if he is not purged from the party, $k^i = 0$, an agent obtains a payoff $v(\omega, \tau)$, which depends on the outcome of the project and the agent's type. We assume that $v(F, \tau) = 0$, v(S, c) > 0 and v(S, nc) < v(S, c). To avoid dealing with corner solutions which only complicate the analysis, we further impose $v(S, nc) \ge 0$. When purged, $k^i = 1$, an agent suffers the loss L > 0 (all our results hold if agents enjoy the payoff from their project being successful even after being purged). Agent *i*'s first-period payoff thus assumes the following form:

$$u_1^i(e,\tau) = R + (1-k^i) \times v(\omega,\tau) + k^i \times (-L) - \frac{e^2}{2}.$$
 (1)

In period 2, if i survives the purge, given that there is no subsequent purge, his payoff can be expressed as the sum of the benefit R (we endogenize the second-period payoff from membership in Section 6) and the net gain from a successful project:

$$u_2^i(e,\tau) = R + v(\omega,\tau) - \frac{e^2}{2}.$$
 (2)

To simplify the exposition, we assume throughout that agents do not discount the future.

The autocrat gets a positive payoff—normalized to 1—when a subordinate's project is successful and 0 otherwise. The autocrat thus wants to maximize the proportion of successful projects, which is equal to agents' average effort in each period (in Section 6, we study a set-up in which the autocrat maximizes her probability of survival). In the first period, the autocrat also bears the cost of investing in the intensity of violence and the cost of purging. Letting \overline{e}_t denote the average effort in period $t \in \{1, 2\}$, we can thus express the autocrat's first-period and second-period payoffs as, respectively:

$$u_1^A(\kappa, L) = \overline{e}_1 - c(\kappa) - \zeta(L); \tag{3}$$

$$u_2^A = \overline{e}_2. \tag{4}$$

The autocrat has a discount factor of $\beta \in (0, 1)$, which captures, among other things, the risk (perceived or real) of losing power between the two periods.

To summarize, the timing of the game is:

Period 1:

- 1. The autocrat chooses the intensity of violence $L \ge 0$;
- 2. Agent *i* chooses effort e_1^i after privately observing his type $\tau^i \in \{c, nc\}$;

- 3. Project outcome ($\omega \in \{S, F\}$) is determined by Nature and observed by the autocrat. The autocrat chooses the proportion of agents to purge who failed (κ_F) and succeeded (κ_S);
- 4. Purged agents are replaced, and first-period payoffs are realized.

Period 2:

- 1. (Surviving and new) agent *i* chooses effort e_2^i ;
- 2. Project outcome is determined;
- 3. Game ends and second-period payoffs are realized.

Note that the assumption that the autocrat commits to an intensity of violence at the beginning of the game is not innocuous. Without commitment, at the moment of the purge, the autocrat would either choose no violence (if violence is costly) or the highest feasible intensity (if it is costless). This is because once agents have exerted their effort, violence has no effect on selection (or, clearly, on effort). This assumption, however, has some historical basis. Funds for the Great Terror were earmarked before its launch (Wolton, 2015, ; unfortunately, there is no similar historical evidence for other mass purges). Observe further that the autocrat prefers to commit whenever her preferred intensity of violence is positive.

The equilibrium concept is Perfect Bayesian Equilibrium (PBE), which requires that each agent correctly anticipates the autocrat's purging decision and other agents' efforts when choosing his own effort, and, in turn, the autocrat correctly anticipates the level of effort by each type when determining her investment in violence and purging strategy. For simplicity, we assume that agents are anonymous, so all agents with a successful (failed) project face the same probability of being purged. Finally, to deal with measurability issues, we assume that when the autocrat observes an out-of-equilibrium event, she treats the deviation as a mistake and does not distinguish between the agent who deviated and other subordinates who followed their prescribed strategy.² If, after these restrictions, multiple PBE arise, we select the one which maximizes the incidence of the purge; that is, κ_F and κ_S . Our main insights and comparative statics are robust to change in this last criterion. In what follows, the term 'equilibrium' refers to the PBE which satisfies all our refinements.

Throughout, we use the following notation. $V_2(\tau)$ denotes agent *i*'s expected payoff in period 2 as a function of his type. Simple algebra yields $V_2(\tau) = R + \frac{(v(S,\tau))^2}{2}$. The (ex-ante) average payoffs are denoted by $\overline{v} = \lambda v(S, c) + (1 - \lambda)v(S, nc)$ and $\overline{V_2} = \lambda V_2(c) + (1 - \lambda)V_2(nc)$. For the autocrat, $W_2(\tau)$ denotes her second-period expected payoff induced by an agent of type $\tau \in \{c, nc\}$. It can

²Alternatively, we could define two subsets of agents ξ_0 and ξ_1 who always exert (respectively) effort 0 and 1, implying that there is no out-of-equilibrium event of measure 0.

be verified that $W_2(\tau) = v(S, \tau), \ \tau \in \{c, nc\}$. The gain from replacing a non-congruent type with a congruent type is $\mathcal{D}^{c,nc} := W_2(c) - W_2(nc)$. To limit the number of cases, we assume that the autocrat never purges all subordinates even if there is no congruent agent to begin with (e.g., due to the risk of popular rebellion if work is too disrupted). That is, we impose $\beta r \mathcal{D}^{c,nc} < C_0 + C_1$. Finally, we assume that the highest feasible intensity of violence, denoted \overline{L} , satisfies $\overline{L} := 1 - v(S, c) - V_2(c)$, which guarantees that subordinates' effort choices are interior and simplifies the analysis.

3 Effort and incentive to purge

Due to our equilibrium refinements, there is no equilibrium in which agents exert zero effort.³ When they exert effort, agents endogenously sort into failure and success pools, which constitute, with the proportion of congruent types among current subordinates (λ) and potential replacements (r), the only information available to the autocrat at the time of her purging decision. However, given that congruent types receive a greater intrinsic benefit from a successful project, they always exert more effort than non-congruent types and are more likely to belong to the success pool (i.e., the single-crossing condition holds). Consequently, the autocrat always targets unsuccessful agents first.

Lemma 1. In equilibrium, $\kappa_S > 0$ only if $\kappa_F = 1$.

Proof. All proofs are collected in the Online Appendix available on the authors' websites.⁴ \Box

This lemma implies that only three qualitatively distinct forms of purges can emerge in equilibrium. First, some but not all failures are purged ($\kappa_F \in (0, 1)$ and $\kappa_S = 0$); we label these purges 'partially discriminate.' Second, the entire failure pool is purged with all successful agents surviving ($\kappa_F = 1$ and $\kappa_S = 0$); we refer to these purges as 'fully discriminate.' Finally, all failures and some successful subordinates are purged ($\kappa_F = 1$ and $\kappa_S > 0$); we classify these purges as 'semi-indiscriminate.'

An agent's effort depends critically on the nature of the purge. In a partially or fully discriminate purge, the payoff from belonging to the success pool is large: a successful agent obtains his

³Absent our restrictions, there would exist an equilibrium in which all agents exert zero effort and the autocrat would set $\kappa_S = 1$. This equilibrium would only be sustained by the arguably unreasonable out-of-equilibrium belief that a successful agent is likely to be non-congruent even though congruent types have greater intrinsic motivation to exert effort.

⁴sites.google.com/bpmontagnes/papers and stephanewolton.com/research

flow payoff and is inoculated against the purge. As such, a subordinate exerts high effort when anticipating a discriminate purge. As κ_F increases, so too does the benefit of effort. Conversely, in a semi-indiscriminate purge, the benefit from belonging to the success pool is relatively low. Even if successful, there is a risk an agent's effort is wasted as he may be purged anyway. Furthermore, as κ_S increases, the benefit of success diminishes.

Lemma 2 formally describes an agent's effort as a function of his type. Observe the difference between a (partially or fully) discriminate purge ($\kappa_S = 0$) and a semi-indiscriminate purge ($\kappa_S > 0$). In the former, the subordinate considers the flow payoff from a successful project ($v(S, \tau)$) as well as the expected loss from being purged ($\kappa_F(V_2(\tau) + L)$), whereas in the latter, he takes into account the benefit of success ($v(S, \tau) + V_2(\tau) + L$), weighted by the probability of surviving the purge ($1 - \kappa_S$).

Lemma 2. A type $\tau \in \{c, nc\}$ agent i chooses effort:

$$e_{1}^{i}(\tau) = \begin{cases} v(S,\tau) + \kappa_{F}(V_{2}(\tau) + L) & \text{if } \kappa_{S} = 0\\ (1 - \kappa_{S})(v(S,\tau) + V_{2}(\tau) + L) & \text{if } \kappa_{S} > 0 \end{cases}$$
(5)

Having characterized the agents' effort choices, we now turn to the autocrat's incentive to purge. When deciding whom to purge, the autocrat observes only the outcome of an agent's project, and forms a posterior about the subordinate's congruence based on success (denoted μ^S) or failure (μ^F) . The autocrat's posteriors incorporate her conjectures (correct in equilibrium) of the different levels of effort exerted by congruent and non-congruent types. Due to distinct degrees of intrinsic motivation, a successful project is a positive signal of congruence, so $\mu^F < \lambda < \mu^S$. However, this signal is never perfect: congruent subordinates sometimes fail and non-congruent types sometimes succeed. Consequently, in any purge, some congruent types are purged and some non-congruent subordinates survive.

Consider the autocrat's incentives to purge an agent after observing his project is unsuccessful. If the autocrat retains the agent after failure, her expected payoff is: $\mu^F W_2(c) + (1 - \mu^F) W_2(nc)$. Since the proportion of congruent types in the replacement pool is r, the autocrat's payoff from purging an unsuccessful agent is: $rW_2(c) + (1 - r)W_2(nc)$. The autocrat's expected benefit from purging an agent after failure is thus:

$$\mathcal{W}^F = [r - \mu^F] \mathcal{D}_2^{c,nc}.$$
 (6)

By a similar reasoning, the expected benefit from purging a successful subordinate is:

$$\mathcal{W}^S = [r - \mu^S] \mathcal{D}_2^{c,nc}.$$
(7)

The autocrat's incentive to purge thus depends critically on the informativeness of her posteriors, which are a function of agents' endogenous sorting into the success and failure pools. Thus, there exists a 'pool makeup' effect which captures how changes in agents' efforts affect the autocrat's learning. We say that the pool makeup effect is positive when the target pool—failure pool in a discriminate purge, success pool in a semi-indiscriminate purge—becomes more tainted and the autocrat's incentive to purge increases. To make sense of this effect, we first examine the impact of exogenous increases in purge incidences on the autocrat's relevant posteriors.

Anticipating a discriminate purge ($\kappa_S = 0$), both congruent and non-congruent types increase their effort and exit the failure pool in response to a higher threat of being purged conditional on failure κ_F . Congruent types, however, have more to lose from being purged and so increase their effort relatively more (see Equation 5). Further, congruent types are also less likely to fail to start with. The combination of these two effects implies that relative to their original level, a greater percentage of congruent than non-congruent types exit the failure pool in response to a higher purge incidence. In a partially discriminate purge, the risk of being purged is self-reinforcing with increased purge inference κ_F causing the failure pool to become more tainted and thus generating a positive pool makeup effect.

In a semi-indiscriminate purge ($\kappa_S > 0$), the autocrat's benefit from purging an agent depends at the margin on the informativeness of success (see Equation 7). A higher purge incidence (κ_S) now decreases effort, with congruent types being again more responsive (see Equation 5). However, congruent types are more likely to succeed to start with. In general, it is difficult to determine whether a greater percentage of congruent types exit the success pool. In our set-up, since agents' cost of effort exhibits constant elasticity, the proportion of exits is the same for both types, resulting in a null pool makeup effect.

Remark 1. Consider an exogenous increase in the purge incidence.

(i) In a partially discriminate purge, as κ_F increases, the autocrat's expected benefit of purging an additional agent (\mathcal{W}^F) increases.

(ii) In a semi-indiscriminate purge, as κ_S increases, the autocrat's expected benefit of purging an additional agent (\mathcal{W}^S) remains constant.

While equilibrium quantities ultimately depend on the equilibrium intensity of violence, it is useful to consider how the purging decision is reached. Because all players correctly anticipate each other's strategy, we can simply compare the marginal cost of purging an additional agent with the marginal benefit. As Remark 1 establishes, the marginal benefit only depends on the purge incidence and so can be re-expressed as $\mathcal{W}^F(\kappa_F, \kappa_S)$ and $\mathcal{W}^S(\kappa_F, \kappa_S)$. In turn, the marginal cost is $C_0 + C_1 \times \alpha_F \times \kappa_F$ in a discriminate purge (as only failures get purged) and $C_0 + C_1 \times (\alpha_F \times 1 + \alpha_S \times \kappa_S)$ in a semi-indiscriminate purge (as all failures and some successful subordinates are purged). Let $\hat{\alpha}_F(L)$ be the proportion of failures in a fully discriminate purge ($\kappa_F = 1$, $\kappa_S = 0$) for intensity of violence L and note that $\hat{\alpha}_F(L) = 1 - \overline{v} - \overline{V}_2 - L$. The next Lemma characterizes the purge incidences κ_F and κ_S as well as the nature of the purge as a function of the model primitives when the intensity of violence is held fixed.

Lemma 3. For each L, there exist unique equilibrium purge incidences $\kappa_F^*(L)$, $\kappa_S^*(L)$. Further, whenever (i) $C_0 + C_1 \times \hat{\alpha}_F(L) > \mathcal{W}^F(1,0)$, the purge is partially discriminate; (ii) $W^S(1,0) \leq C_0 + C_1 \times \hat{\alpha}_F(L) \leq \mathcal{W}^F(1,0)$, the purge is fully discriminate; (iii) $C_0 + C_1 \times \hat{\alpha}_F(L) < \mathcal{W}^S(1,0)$, the purge is semi-indiscriminate.

While condition (iii) appears stringent, as Corollary 1 establishes, there exists an open set of parameters that satisfy it as long as the replacement pool is sufficiently congruent relative to the current set of agents.

Corollary 1. If $r > \lambda$, there exists a non-measure zero set of parameter values such that the equilibrium purge is semi-indiscriminate.

Lemma 3 and our earlier analysis imply that for each intensity of violence we can reconstruct the unique purge breath from its constituting parts. Extending notation in a natural way, the equilibrium purge breadth is $\kappa^*(L) = \alpha_F^*(L) \times \kappa_F^*(L) + \alpha_S^*(L) \times \kappa_S^*(L)$. It now remains to determine the autocrat's optimal choice of violence. To do so, we proceed in two steps. We first examine how exogenous changes in L affect the autocrat's and agents' choices. We then solve the autocrat's maximization problem.

4 The love-fear trade-off

We first consider the effect of violence on subordinates' effort, which we label the 'fear effect.' From Equation 5, the intensity of violence has a direct effect which always leads to increased first-period performance. That is, the fear effect is always positive. This direct effect of L on agents' efforts, in turn, induces a series of second-order effects which need to be carefully unpacked.

Let us first consider the indirect effect of violence on effort through purge incidence. To do so, suppose that the purge breadth κ is held fixed. In a partially discriminate purge, greater L then increases the risk of being purged conditional on failure (as $\kappa = \alpha_F \times \kappa_F$ with α_F decreasing due to the direct effect). Since higher purge incidence κ_F increases effort, the indirect effect through purge incidence is positive. Turning to semi-indiscriminate purges, we similarly find that greater intensity of violence increases the risk of being purged conditional on success (as $\kappa = \alpha_F \times 1 + \alpha_S \times \kappa_S =$ $1 - (1 - \kappa_S)\alpha_S$ with α_S increasing). However, greater purge incidence κ_S now tends to depress effort (see Equation 5) and the indirect effect through purge incidence is then negative.

While the direct effect of violence is similar for all types, this is not the case for the indirect effect through purge incidence. As we have seen in the previous section, congruent types are more responsive to change in the threat of being purged. Fear differentially motivates congruent and non-congruent agents and thus triggers a pool makeup effect.

In a partially discriminate purge, the positive indirect effect triggers a positive pool makeup effect as congruent agents increase their effort more than non-congruent subordinates (the failure pool thus becomes more tainted). The autocrat then purges more failures, further increasing the purge incidence. Direct and indirect effects all go in the same direction. Equilibrium effort by all subordinates increases at a (relatively) high rate with the intensity of violence.

This has a further consequence. As violence increases, the failure pool is quickly exhausted (the failure pool becomes thinner and the purge incidence increases). Consequently, there exists a threshold of violence—denoted $L^{full} \geq 0$ —such that for all $L > L^{full}$, all failures are purged. In this case, the autocrat does not immediately start purging from the success pool because the pools are qualitatively distinct (recall $\mathcal{W}^F(\kappa_F, \kappa_S) > \mathcal{W}^S(\kappa_F, \kappa_S)$). Instead, for intermediate intensity of violence, the purge is fully discriminate ($\kappa_F^*(L) = 1$ and $\kappa_S^*(L) = 0$).

In a fully discriminate purge, the purge incidence is constant. The direct effect of violence encourages higher performance with congruent and non-congruent agents exiting the failure pool at the same rate. This increases the autocrat's incentive to purge from the success pool since (i) the marginal cost $(C_0 + C_1 \alpha_F^*(L) \times 1)$ is reduced and (ii) the marginal benefit increases (relative to their original level, a greater percentage of non-congruent types enters the success pool). Consequently, there exists a threshold of violence—denoted $L^{ind} \leq \overline{L}$ —such that for all $L \geq L^{ind}$, some successful agents are purged; the purge becomes semi-indiscriminate.

In a semi-indiscriminate purge, the indirect effect through purge incidence is negative, implying that congruent subordinates increase their effort less than non-congruent agents. The success pool thus becomes more tainted, the pool makeup effect is positive, and the autocrat purges more successful agents further increasing the purge incidence. In a semi-indiscriminate purge, the direct effect increases effort, but the indirect effects (through purge incidence and the pool makeup effect) both dampen it. Consequently, average first-period performance increases at a relatively low rate with the intensity of violence.⁵

Before formalizing the analysis above, it is useful to introduce the following condition:

$$C_0 + C_1 \hat{\alpha}_F(\overline{L}) < \beta r \mathcal{D}^{c,nc}.$$
(8)

The right-hand side represents the marginal benefit of purging an additional failure when the purge is fully discriminate (i.e., $\kappa_F = 1$, $\kappa_S = 0$) and the intensity of violence is maximal (i.e., $L = \overline{L}$). In this case, congruent types never fail ($e^i(c) = 1$) and the failure pool is only composed of noncongruent agents. The left-hand side, in turn, is the marginal cost of removing an additional subordinate when all failures and no successful subordinates are purged at the maximal intensity of violence. When Equation 8 does not hold, the marginal cost of purging all failures is thus too high compared to the marginal benefit and the purge is always partially discriminate.

Using Equation 8, the proposition that follows establishes that the nature of the purge depends on the intensity of violence.

Proposition 1. 1. Suppose Equation 8 does not hold. Then for all intensities of violence, the purge is partially discriminate: $\kappa_F^*(L) \in [0, 1)$.

2. Suppose Equation 8 holds. Then there exist unique $L^{full} < \overline{L}$ and $L^{ind} \in (L^{full}, \overline{L}]$ such that: (i) For $L < L^{full}$, the purge is partially discriminate $(\kappa_F^*(L) \in [0, 1))$; (ii) For $L \in [L^{full}, L^{ind}]$, the purge is fully discriminate $(\kappa_F^*(L) = 1 \text{ and } \kappa_S^*(L) = 0)$; (iii) For $L > L^{ind}$, the purge is semi-indiscriminate $(\kappa_S^*(L) > 0)$.

Proposition 1 highlights a pattern that is unique to our theory as the intensity of violence fully determines the nature of the purge. Mild purges tend to be partially discriminate, violent purges semi-indiscriminate. For intermediate intensity, purges are fully discriminate.

To limit the number of cases under consideration, we assume in what follows that Equation 8 holds so the purge is fully discriminate for sufficiently high intensity of violence and potentially

⁵The attentive reader may wonder whether the negative second-order effects we describe can dominate the direct effect of violence. The answer is no, average effort is increasing in L even in a semi-indiscriminate purge. This is best shown by contradiction. If the first-period performance were to decrease with L, fixing the purge breadth, the purge incidence κ_S would decrease, and the indirect effect on effort would become positive. As a result, all the forces described in the text would go in the opposite direction which would imply greater effort, a contradiction.



Figure 1: Fear effect Parameter values: $\lambda = 1/3$, r = 2/3, R = 0, v(S, c) = 1/4, v(S, nc) = 0, $\beta = 0.9$, $C_0 = 0$, and $C_1 = 0.17$.

semi-indiscriminate for very large L. Observe that all the comparative statics we establish for partially discriminate purges when Equation 8 is satisfied also hold when it is not. To avoid dealing with corner solutions, we further suppose that the purge breadth is always strictly positive (formally, the marginal cost satisfies $C_0 < \beta(r - \lambda \frac{1-v(c)}{1-\overline{v}})\mathcal{D}^{c,nc}$).

While greater violence always generates better first-period performance (i.e., the fear effect is always positive), the responsiveness of effort to L depends on the nature of the purge and thus (using Proposition 1) the original intensity of violence. In a partially discriminate purge, as violence increases, direct and indirect effects complement each other until the nature of the purge changes to fully discriminate. In a semi-indiscriminate purge, the indirect effects (through purge incidence and the pool makeup effect) mitigate the direct effect. Consequently, the fear effect tends to be strongest in a partially discriminate purge (if L is sufficiently large), lowest in a semi-indiscriminate purge, and intermediate in a fully discriminate purge (when there are no indirect effects). This relationship between effort and violence is summarized in Proposition 2 and illustrated in Figure 1.

Proposition 2. The total derivative of average effort with respect to violence $\frac{d\bar{e}(L)}{dL}$ is always strictly positive. Further, there exists a unique $L^{fear} \leq L^{full}$ such that the derivative satisfies: (i) $\frac{d\bar{e}(L)}{dL} > 1$ for all $L \in (L^{fear}, L^{full})$; (ii) $\frac{d\bar{e}(L)}{dL} = 1$ for all $L \in [L^{full}, L^{ind})$; (iii) $\frac{d\bar{e}(L)}{dL} < 1$ for all $L \geq L^{ind}$. While the nature of the purge changes from partially to fully discriminate and then to semiindiscriminate as the intensity of violence increases, the purge breadth does not vary monotonically with the intensity of violence. The equilibrium purge breadth is determined by equating the marginal cost of a purge to the marginal benefit: $C_0 + C_1 \times \alpha_F^*(L) \times \kappa_F^*(L) = \mathcal{W}^F(\kappa_F^*(L), 0)$ in a partially discriminate purge and $C_0 + C_1 \times (\alpha_F^*(L) \times 1 + \alpha_S^*(L) \times \kappa_S^*(L)) = \mathcal{W}^S(1, \kappa_S^*(L))$ in a semi-indiscriminate purge. As we have already noted, fear generates a positive pool makeup effect in partially discriminate purges and semi-indiscriminate purges. That is, the marginal benefit of purging $(\mathcal{W}^F(\kappa_F^*(L), 0)$ in a partially discriminate purge and $\mathcal{W}^S(1, \kappa_S^*(L))$ in a semi-indiscriminate purge) increases with the intensity of violence. Therefore, the purge breadth $(\alpha_F^*(L) \times \kappa_F^*(L)$ in a partially discriminate purge and $\alpha_F^*(L) \times 1 + \alpha_S^*(L) \times \kappa_S^*(L)$ in a semi-indiscriminate purge) must be positively correlated with the intensity of violence. However, we also need to account for fully discriminate purges (i.e., $L \in [L^{full}, L^{ind}]$) when the failure pool shrinks as violence increases and the purge breadth decreases. Proposition 3 summarizes these findings and Figure 2 illustrates them.

Proposition 3. The relationship between purge breadth, $\kappa^*(L)$, and the intensity of violence, L, exhibits the following properties:

(i) For $L < L^{full}$, $\kappa^*(L)$ is strictly increasing with L;

(ii) For $L \in [L^{full}, L^{ind}]$, $\kappa^*(L)$ is strictly decreasing with L;

(iii) For $L > L^{ind}$, $\kappa^*(L)$ is strictly increasing with L.



Figure 2: Equilibrium purge breadth and intensity of violence Parameter values: $\lambda = 1/3$, r = 2/3, R = 0, v(S, c) = 1/4, v(S, nc) = 0, $\beta = 0.9$, $C_0 = 0$, and $C_1 = 0.17$.

Having examined the effect of violence on effort and on the nature and breadth of the purge, we now consider its impact on selection. Observe that since a new agent is always better on average (from the autocrat's perspective) than a purged agent, the purge always improves the second-period congruence (defined by the proportion of congruent types among subordinates). An *increase* in the intensity of violence, however, changes how much the autocrat benefits from the purge in terms of selection. We refer to this comparative static as the love effect, which is positive (resp. negative) if greater violence improves (resp. worsens) selection. The next proposition establishes conditions under which the love effect is negative for the survivors of the purges and *for all subordinates*. Figure 3 illustrates this result.

Proposition 4.

(i) The proportion of congruent types among surviving subordinates of the purge strictly increases with L if and only if $L < L^{full}$, and strictly decreases otherwise.

(ii) The proportion of congruent types among subordinates in the second period weakly increases with L for all L if and only if $\lambda \ge r$.

(iii) If $r \in (\lambda, 2\lambda]$, the proportion of congruent types among subordinates in the second period strictly increases with L for $L < L^{full}$ and strictly decreases otherwise.



Figure 3: Love effect

The solid line corresponds to the proportion of congruent types among all subordinates in period 2, the dashed line to the proportion of congruent types among survivors of the purge. Parameter values: $\lambda = 1/3$, r = 2/3, R = 0, v(S,c) = 1/4, v(S,nc) = 0, $\beta = 0.9$, $C_0 = 0$, and $C_1 = 0.17$.

The first part of the proposition highlights that the autocrat's ability to screen agents decreases with L unless the purge is partially discriminate. For $L < L^{full}$, as the intensity of violence increases, more agents exit the failure pool and more failures are purged. Among surviving subordinates, a greater proportion of agents thus belongs to the success pool, which is of higher quality than the failure pool. The love effect is then positive. When the purge is fully discriminate or semi-indiscriminate, surviving agents all belong to the success pool. Screening then worsens because the success pool becomes more tainted as more non-congruent types enter the pool *relative to* the stocks of both types. The love effect is then negative.

Even though purged agents are replaced by more congruent types (in expectation), the love effect can still be negative when it comes to the (second-period) composition of subordinates. Maybe surprisingly, the love effect is always positive for high enough intensity of violence $(L \ge L^{full})$ only if the replacement pool is worse than the existing pool of subordinates (Proposition 4(ii)). This occurs because in a fully discriminate purge, the purge breadth decreases (Proposition 3(ii)) and existing agents constitute a greater proportion of the second-period subordinates. The negative love effect among second-period agents is not driven by decreasing purge breadth. Unity worsens with violence in a semi-indiscriminate purge (despite increasing $\kappa^*(L)$) whenever the quality of the replacement pool is not too high: $r \le 2\lambda$. Under this sufficient condition, the lower congruence of the surviving subordinates dominates the gain from replacement.

This section highlights the multidimensional impact of violence, through fear and love. The fear of violence always engenders higher effort in the first period (Proposition 2). The relationship between violence and selection is more nuanced and depends on both the congruence of the replacement pool and the intensity of violence (Proposition 4). When the replacement pool is sufficiently congruent $(r > \lambda)$ and violence is sufficiently large $(L > L^{full})$, the autocrat faces a stark trade-off between fear and love.

5 Intensity of violence

In this section, we examine how the autocrat optimally selects violence in light of the (positive) fear and (potentially negative) love effects. Throughout, we assume that parameter values are such that the autocrat's marginal benefit of increased violence is continuous in L. It needs not be the case. Due to the positive direct and indirect effects of violence, the benefit of purging is strictly convex in L, and the equilibrium purge incidence may exhibit discontinuities as a function of L. A technical discussion of this issue is relegated to Online Appendix C. Here, it suffices to note that imposing continuity does not change this section's results, but greatly simplifies the exposition.

Before proceeding to the choice of violence, it is helpful to describe the autocrat's marginal benefit from violence when the pool of replacement is high quality $(r > \lambda)$ as illustrated in Figure 4. Even though the benefit of violence is continuous, the marginal benefit is not. Both at $L = L^{full}$ and $L = L^{ind}$, it exhibits downward shifts. In addition, the marginal benefit reaches a maximum in a partially discriminate purge and is lowest in a semi-indiscriminate one. These are consequences of the fear and love effects. From Proposition 2, we know that the fear effect decreases discontinuously at the $L = L^{full}$ and $L = L^{ind}$ thresholds. From Proposition 4, the love effect turns negative (at least for survivors) when the purge moves to fully discriminate or semi-indiscriminate.



Figure 4: Marginal benefit of violence Parameter values: $\lambda = 1/3$, r = 2/3, R = 0, v(S, c) = 1/4, v(S, nc) = 0, $\beta = 0.9$, $C_0 = 0$, and $C_1 = 0.17$.

Observe further that the marginal benefit of violence is strictly increasing for relatively low L $(L \leq L^{full})$. In this range, all indirect effects reinforce the direct effect of increased violence on effort: the fear effect is maximized, the love effect is positive, and the purge threat is self-reinforcing. Consequently, even though all utility functions are concave, because of the indirect effects resulting from holding a mass of subordinates accountable, the autocrat's marginal benefit of violence is strictly increasing.

Having described the properties of the marginal benefit, we can turn to the autocrat's choice of violence. Proposition 5 characterizes the optimal intensity of violence when the cost of investing in L is relatively large and relatively low.⁶

⁶At $L = L^{full}$, by definition $\kappa_F^*(L) = 1$. Effort thus satisfies $e_1^i(\tau) = v(\tau) + V_2(\tau) + L$ (see Lemma 2). The posterior $\mu^F = \lambda \frac{1-e_1^i(c)}{1-\overline{e_1}}$ depends only on model parameters and L. For $L > L^{full}$, the posterior μ^S does not depend on $\kappa_S^*(L)$ (see the discussion prior to Lemma 3). Hence, the quantities used in the text of Proposition 5 only depend on the intensity of violence and model primitives.

Proposition 5. There exists (almost always) a unique equilibrium intensity of violence L^* . Further, (i) If at $L = L^{full}$, $\zeta_0 + \zeta_1 L \ge 1 + \beta \mathcal{D}^{i,o}(\lambda - \mu^F(1,0))$, then $L^* \le L^{full}$; (ii) If at $L = L^{ind}$, $\zeta_0 + \zeta_1 L < 1 - \frac{\partial \kappa_S^*(L)}{\partial L}(\overline{v} + \overline{V_2} + L) + \frac{\partial \mu^S(1,0)}{\partial L}\beta \mathcal{D}^{c,nc}(\overline{v} + \overline{V_2} + L)$, then $L^* > L^{ind}$.

Because of the high marginal benefit of violence described above, a purge is partially discriminate only if the marginal cost of investing in L is relatively large as described in point (i) (observe the high fear effect, equal to 1, and the positive love effect, equal to $\beta \mathcal{D}^{i,o}(\lambda - \mu^F(1,0)))$. In turn, because of the low fear effect $(1 - \frac{\partial \kappa_S^*(L)}{\partial L}(\overline{v} + \overline{V_2} + L))$ with $\frac{\partial \kappa_S^*(L)}{\partial L} > 0$) and negative love effect $(\frac{\partial \mu^S(1,0)}{\partial L}\beta \mathcal{D}^{c,nc}(\overline{v} + \overline{V_2} + L))$ with $\frac{\partial \mu^S(1,0)}{\partial L} < 0$, the marginal benefit of violence is small for high L(i.e., $L \ge L^{ind}$) and the purge is semi-indiscriminate only if the marginal cost of investing in violence is relatively low. The next corollary lists three conditions under which a semi-indiscriminate purge occurs (we have omitted dependence on other parameter values for notational simplicity).

Corollary 2. A purge is semi-indiscriminate if the following conditions are satisfied:

- 1. The proportion of congruent types in the replacement pool r is strictly higher than some $\underline{r} \geq \lambda$;
- 2. The cost parameters C_0 and C_1 are respectively strictly below some $\overline{C_0}(r) > 0$ and $\overline{C_1}(r, C_0) > 0$;
- 3. The cost parameters ζ_0 and ζ_1 are respectively strictly below some $\overline{\zeta_0}(r) > 0$ and $\overline{\zeta_1}(r, \zeta_0) > 0$.

The three conditions are relatively intuitive. First, the replacement pool must be of sufficiently good quality, and in particular, better on average than existing subordinates (condition 1, also necessary). Second, the cost of purging (C_0, C_1) must be sufficiently low to compensate for the relatively low marginal benefit of purging a successful agent (condition 2). Finally, as explained above, the cost of investing in the security apparatus must also be relatively small (condition 3).

Observe that Proposition 5 does not characterize the intensity of violence when points (i) and (ii) do not hold. This is due to the convexity of the benefit of violence, which has two consequences. First, points (i) and (ii) are not satisfied, the purge is not necessarily fully discriminate since the marginal benefit may intersect the linear marginal cost more than once in the range $[0, L^{full}]$, as illustrated in Figure 5. The autocrat may then have to choose between two possible maxima: the lowest intersection L^1 in Figure 5 and L^{full} (the highest intersection L^2 is a local minimum). Second, the presence of multiple maxima implies that small changes in the underlying fundamentals can be associated with a large increase in the intensity of violence. This result is summarized in the next proposition.

Proposition 6. There exists a non-measure zero set of parameter values \mathcal{P}^d such that if $(\lambda, r, v(S, c), v(S, nc), C_0, \zeta_0) \in \mathcal{P}^d$, there exists C_1^d and ζ_1^d satisfying $\lim_{C_1 \uparrow C_1^d} L^* < \lim_{C_1 \downarrow C_1^d} L^*$ and $\lim_{\zeta_1 \uparrow \zeta_1^d} L^* < \lim_{\zeta_1 \downarrow \zeta_1^d} L^*$.



Figure 5: Marginal benefit of violence in a partially discriminate purge The solid blue line is the marginal benefit, the dashed black line the marginal cost. Parameter values: $\lambda = 1/3$, r = 2/3, R = 0, v(S, c) = 1/4, v(S, nc) = 0, $\beta = 0.9$, $C_0 = 0$, $C_1 = 0.17$, $\zeta_0 = 0.5$, and $\zeta_1 = 4.8$.

6 Extensions

In this section, we briefly discuss some extensions of our baseline model. A formal analysis is relegated to Online Appendix D. We consider four different variations. First, we suppose that the autocrat can supplement the second-period benefit for regime insiders. We then assume that the pool of replacement is decreasing in the purge breadth. The third extension suppose that the autocrat seeks to maximize her survival probability. Lastly, we discuss how our framework can be used to study repression.

Endogenous reward

Assume that at marginal cost $\xi'(R_2) = \xi_0 + \xi_1 R_2$, surviving and new agents receive a payoff $R + R_2$ in period 2. This approach facilitates comparison with the baseline model as the latter is equivalent to fixing R_2 to 0. Throughout, to simplify the analysis, we assume that $\xi_0 = \zeta_0$ and that the highest feasible endogenous reward $\overline{R_2}$ and intensity of violence \overline{L} jointly satisfy $\overline{R_2} + \overline{L} = 1 - v(S, c) - V_2(S, c)$ (to exclude corner solutions).

Using a similar reasoning to that in Lemma 2, we obtain that an agent's effort as a function of the purge inferences κ_F and κ_S is:

$$e_1^i(\tau) = \begin{cases} v(S,\tau) + \kappa_F(V_2(\tau) + L + R_2) & \text{if } \kappa_S = 0\\ (1 - \kappa_S)(v(S,\tau) + V_2(\tau) + L + R_2) & \text{if } \kappa_S > 0 \end{cases}$$
(9)

As Equation 9 establishes, the intensity of violence and endogenous rewards are perfect substitutes. It is then useful to think in terms of effort inducement. In this extension, it takes the form $T := L + R_2$, the appropriate mix of violence and reward, whereas it is simply L (with R_2 fixed to 0) in the baseline model. As T and L have the same impact on agents, all the baseline comparative statics with respect to L now hold with respect to T in the extension. Further, the cost of inducing any level of effort is strictly lower with endogenous reward than without (denote $\mathcal{T}(T)$ the cost of T, for any $Z_o > 0$, $\mathcal{T}(Z_o) < \zeta(Z_o)$ and $\mathcal{T}'(Z_o) < \zeta'(Z_o)$). As a result, the equilibrium effort inducement in the extension \hat{T} is strictly higher than the equilibrium intensity of violence in the baseline model L^* .

Combining these results, the next proposition describes how introducing endogenous rewards changes equilibrium quantities.

Proposition 7. Suppose the autocrat can also propose additional reward R_2 in period 2. Compared to the baseline model, in equilibrium:

- (i) The purge incidence is always weakly higher;
- (ii) The purge breadth can be higher or lower;
- *(iii)* The intensity of violence can be higher or lower.

As we have established in Section 4, the purge incidence is increasing (strictly when the purge is not fully discriminate) with effort inducement. Given $\hat{T} > L^*$, the purge inference is weakly higher with endogenous reward than without. In turn, we have found that the purge breadth is non-monotonic in effort inducement so an increase from L^* to \hat{T} can increase or decrease the proportion of subordinates being purged.

Point (iii) of Proposition 7 is slightly more surprising since it states that even though the autocrat can use carrots (R_2) in addition to sticks (L), the intensity of violence may actually be *strictly* higher than in the baseline model. This is a consequence of the indirect effects present in our many-to-one accountability framework which generates a convex benefit of effort inducement starting from a relatively low level (see Figure 4). Since the marginal cost of effort inducement is strictly lower in the extension, the difference between \hat{T} and L^* is very large whenever $L^* < L^{full}$; so much so that equilibrium violence can be strictly higher even if rewards are available.

Decreasing replacement pool

Throughout, we have assumed the autocrat suffers a cost when purging agents. Alternatively, the quality of the replacement pool may be decreasing with the purge breadth κ . In this case, the marginal replacement has a probability $r(\kappa)$ of being congruent with $r'(\kappa)$ strictly negative. In Online Appendix D, we show that when the probability $r(\kappa)$ is linear $(r(\kappa) = \bar{r} - r_1 \kappa)$, the autocrat's problem is isomorphic with the baseline model and all results carry through to this environment.

Observe that a decreasing replacement pool may increase the likelihood that the purge is semiindiscriminate. Since a semi-indiscriminate purge is not necessarily broad, the pool of replacement may be of high (average and marginal) quality when the autocrat starts purging from the success pool.

Mass purges and autocrat's survival

Our results are also robust to a change in the autocrat's objective. Rather than caring about first- and second-period performances, the autocrat may seek to maximize her survival probability, which we assume is a function of the first-period performance (e.g., economic performance) and the proportion of non-congruent agents—to be consistent with our analysis of repression below (our results are robust to maximizing the proportion of congruent subordinates). More formally, the autocrat seeks to maximize

$$P(survives) = \gamma \overline{e}_1 + (1 - \gamma)\beta(1 - \mathcal{P}),$$

with \mathcal{P} the proportion of congruent agents, γ the weight on first-period performance, and $\beta(\cdot)$ a strictly decreasing and weakly concave function (in Online Appendix D.3 we study a more general objective function allowing for some complementarity between performance and the proportion of non-congruent subordinates).

In turn, for ease of comparison, we suppose that a type $\tau \in \{c, nc\}$ subordinate obtains a payoff $v(S, \tau)$ from a successful project in the first period. He also gets $V_2(\tau)$ from surviving the purge, (e.g., congruent agents get a higher payoff when the autocrat survives which implies $V_2(c) > V_2(nc)$ like in the baseline model).

Agents then behave like in the baseline model. Consequently for a replacement pool of high enough quality $(r \ge \lambda)$, a sufficient condition to counteract the non-linear effect of $\beta(\cdot)$, we again obtain in this setting that (i) partially discriminate purges tend to be mild, semi-indiscriminate purges violent and (ii) the purge breadth is non-monotonic in violence. When it comes to the effect of violence on performance and selection, however, the autocrat needs to consider an additional effect through the changes in the marginal value of purging an additional subordinate ($\beta(\cdot)$ is not necessarily linear, in contrast to the baseline model). As long as $\beta(\cdot)$ is not "too concave" (see Propositions D.3 and D.4 for details), this additional effect is second order and the fear and love effects remain present in this modified environment.

This extension establishes that all our findings are robust to changes in the autocrat's objectives, suggesting that the strategic interactions between one principal and a mass of subordinates is the key force behind the equilibrium features of purges and the trade-offs we identify.

Repression

A few recent formal papers have explored the use of repression in autocracies. Whether they are interested in the relationship between the autocrat and her repressive agents (Tyson, 2017; Dragu and Przeworski, 2017) or the signaling (Shadmehr and Boleslavsky, 2017) or preventive (Rozenas, 2017) role of repression, these works consider an autocrat facing a single citizen. In turn, the framework described in the previous subsection can be modified to study repression when the autocrat faces a mass of individuals (see also Gregory et al. 2011 for a framework with exogenous information and thus none of the effects we identify). As such, we can explicitly model one of the key characteristics of repression, the elimination (permanent if repressed citizens are killed or temporary if they are imprisoned) of the repressed population. This distinguishes repression from purges where targets are subordinates which may be replaced by more suitable agents.

To study repression, we assume that the autocrat's survival probability depends on the mass (rather than proportion) of non-congruent citizens denoted \mathcal{N} .⁷ That is, the autocrat now maximizes the following function:

$$P(survives) = \gamma \overline{e}_1 + (1 - \gamma)\beta(\mathcal{N}),$$

with $\beta(\cdot)$ strictly increasing and weakly concave. The inference $\kappa_F(\kappa_S)$ now represents the probability a failed (successful) citizen is repressed. As above, we assume that an individual gets $v(S, \tau)$, $\tau \in \{c, nc\}$ from a first-period successful project and $V_2(\tau)$ from surviving the repression.

⁷If the autocrat's survival probability depends on the mass of congruent individuals, she would never repress as repression, by decreasing the number of citizens, reduces the mass of congruent types. If her survival probability depends on the proportion of congruent types, repression is never semi-indiscriminate since in this case, the proportion of congruent types among survivors is: $\bar{e}_1 \mu^S / (\bar{e}_1 \mu^S + \bar{e}_1 (1 - \mu^S)) = \mu^S$, which does not depend on κ_S .

Despite the lack of replacement, repression resembles purges in equilibrium. Mild repressions tend to be discriminate, violent repressions—or terror using Gregory's 2009 terminology—are generally semi-indiscriminate. The breadth of repression is non-monotonic in violence. Violent repression generates more effort, but may improve non-congruent citizens' survival chances (in relative terms).

7 Discussion

In this last section, we discuss the implications of our findings for the understanding of mass purges, the impact of violence in autocracies, and top-down accountability more generally.

Are purges rational?

Our approach presupposes that mass purges are an instrument of top-down accountability in autocracy. Doing so, our theory yields some distinctive patterns. First, the nature of the purge is linked to the intensity of violence: partially discriminate purges tend to be relatively mild, semi-indiscriminate purges brutal. Second, whenever the autocrat's information is coarse (but not necessarily binary as in our set-up), the purge breadth is non-monotonic in the intensity of violence: violent purges do not necessarily translate into a greater proportion of subordinates being purged.

These findings can help distinguish our theory from others. For example, if mass purges are simply random, we should not expect the nature of the purge and its intensity of violence to be related. In turn, if violent purges are simply purges that get out of hand, the correlation between purge breadth and the intensity of violence is likely to always be positive. Using the patterns described above, our theory can thus be falsified.

Historical evidence offers some prima facie evidence in line with our results (a more comprehensive analysis is left for future research). In his comparison of Chinese and Soviet purges, Teiwes (1993) remarks that Chinese rectification campaigns were characterized by low intensities of violence and high levels of predictability, whereas in the USSR in the thirties intensities of violence were high and the targets of the purges less delimited as "flouting commands court danger, but even enthusiastic compliance is no guarantee of safety" (*ibid.*, 25). In the language of our paper, Maoist purges resemble discriminate purges and Stalinist purges can be thought of as examples of semi-indiscriminate purges. This relationship is not circumscribed to China and USSR. In Fascist Italy, Giuriati—in charge of the purge of the PNF—"clearly intended the purge to be firm and selective" with temporary suspension of membership the most common sanction (Morgan, 2012, 330). In contrast, the purge of the Iraqi Ba'ath party in 1979 was violent, with many receiving long sentences, and "directed at anyone *suspected* of opposing Saddam Hussein" (Coughlin, 2005, 163, emphasis added).

One can wonder whether the relatively stringent conditions for semi-indiscriminate purges (see Corollary 2) were met in the USSR in the 30s and Iraq in 1979. While we do not have definitive evidence, it should be noted that both countries experienced a significant number of purges in the preceding years, potentially gradually building up the necessary infrastructure of violence. Additionally, the replacement pool in 1930s USSR was more aligned with Stalin than existing party members. The main beneficiaries of the purges of the thirties were the hundreds of thousands of students who graduated from the Stalinist state schools between 1928 and 1938 (Brzezinski, 1956). These new cadres were more loyal to both the Soviet regime and Stalin (Fitzpatrick, 1979; Wolton, 2015).

Historians have also collected some basic statistics on the proportion of subordinates purged in the USSR (see Table 1) and Maoist China (see Table 2). Despite the differences in intensity of violence and nature, Soviet and Chinese purges had similar breadth, in line with our predictions. During the 1930s Stalinist purges, the proportion of purged members varied from 5% in 1930, 1931, and 1937 to 22% in 1933-34. In turn, the expulsion rate in Chinese rectification campaigns fluctuated between 9% in 1957-58 and 23% in 1947-48.⁸

Our framework even offers some rationale for why the purge breadth differs widely from one purge to the next. As we have seen, small changes in parameters can be associated with large swings in intensity of violence and, consequently, in breadth and nature. While the violence and breadth of the purge may be carefully planned by an autocrat with superior knowledge of the fundamentals, it may be difficult to predict for external observers and even appear random.

We can also reassess two critical historical events—the Great Terror (1936-38) and the Cultural Revolution (1962-76)—in light of our theoretical framework. In their broad aspects, both events exhibit interesting similarities. Both were meant to increase the leader's control over his party and his hold on power and so are best understood in light of our extensions. Both were violent and semi-indiscriminate when it comes to the purges of the communist party or the army as well as the

⁸Given the differences in party size and population, proportion of party members purged provides a better comparison and more closely fits our model. The number of members expelled from the party was always much larger in China.

repression of the general population (Conquest, 2008).⁹ A low cost of violence, our theory contends, can explain these similarities. Stalin drastically simplified procedures (troikas, confessions in place of evidence) to facilitate the Great Terror (Gregory, 2009) and Mao delegated the implementation of the Cultural Revolution to the Red Guards in 1966-68 and to the army in 1968-71 (Dikötter, 2016). Importantly, our theoretical explanation holds for both purge and repression, despite their qualitative differences (only purged agents are replaced).

While purges have become less common following the deaths of Stalin and Mao, they have not completely disappeared. A recent example of a mass purge is the ongoing Erdogan purge in Turkey. While the first victims of the purge were clearly connected with the Gulanist movement (New York Times Magazine, 2017), the purge has changed in recent months. The regime now uses proxies such as criticizing the regime or signing petitions to select its targets in the bureaucracy, schools, and media (Turkey Purge Website, 2017). Given its relative mildness (purged people are expelled or face arrest), our theory predicts that the Erdogan purge is unlikely to become (semi-)indiscriminate, though its breadth may still be large and affect thousands of individuals.

The cases of the Great Terror, Cultural Revolution, or Erdogan purge, however, highlight one limitation of our theory. While useful to understand the strategic interactions between a (would-be) autocrat and her surbordinates or the general population, our framework has little to say about elite purges—the almost exclusive focus of all previous research on the subject—which have occurred in all three events. We hope that our set-up can serve to explore the connections between elite and mass purges.

The ever-present possibility of violence

Unlike representatives in democracies, autocrats face few constitutional restraints on their actions (although they may face international pressures). As such, autocrats have a wider range of tools at their disposal, especially violence.

Our framework shows that even if autocrats face no de jure checks (though investing in violence infrastructure is costly), subordinates' strategic behavior de facto constrains autocrats' actions. As violence increases, so does the fear of being purged; agents work more whether or not they are aligned with the autocrat. But this general increase in effort has a clear downside: an agent's

⁹Some groups, it should be noted, suffered more from the Great Terror due to observable characteristics, especially their nationality (Conquest, 2008). We can incorporate this feature in our framework by allowing for population-specific priors (different λ 's).

accomplishment is less informative about his ideological congruence with the autocrat. As Stalin himself remarked, "saboteurs disguise themselves by over-fulfilling the plan" (cited in Dallin and Breslauer, 1970, 57). We show that even after accounting for the autocrat's response through her choice of purge breadth, the gain in performance can induce a loss in selection, which we have termed the love-fear trade-off.

The Great Terror—and its high intensity of violence—may have been helpful to deal with a lack of performance in all sectors of the Soviet economy (Gregory, 2009), but it was ill-adapted to screen agents and eliminate a so-called fifth column within the USSR (a justification advanced by Molotov, Kaganovich, former U.S. ambassador Davies, or Sayers and Kahn [1946] to defend Stalin's actions).

While absent from the formal literature, the observation that fear and loyalty may not go together is not entirely new. Machiavelli (2005, Chapter 17) remarked that it is difficult for a Prince to be both loved and feared. And so did Stalin, according to an anecdote circulating among Moscow party members in 1931 (reported by Dallin and Breslauer, 1970, 42 footnote 37). "Yagoda was alleged to have asked Stalin: 'Which would you prefer Comrade Stalin: that party members should be loyal to you from conviction or from fear?' Stalin is alleged to have replied: 'From fear.' Whereupon Yagoda asked, 'Why?' To which Stalin replied: 'Because convictions can change: fear remains'."

Many-to-one accountability

Our framework consists of many subordinates accountable to a single principal, the autocrat. This contrasts with all previous political agency models in which one agent is accountable to one (e.g., prime minister) or many (e.g., voters) principals. In this subsection, we highlight the particularities of many-to-one accountability.

In our set-up, all agents work on independent projects. However, they are not evaluated in isolation. The autocrat makes her purging decision based on relative performance; on the informativeness of success and failure.

The interdependencies between agents we uncover have important consequences. They generate indirect effects which must be taken into account. An increase in the intensity of violence does not just affect effort directly, but also indirectly via change in the threat of being purged (the purge incidence) and the autocrat's incentive to purge (the pool makeup effect). These indirect effects, in turn, generate convexities in a world that is inherently concave (see the autocrat's and agents' utility functions). As a result, small changes in fundamentals can have large consequences for equilibrium values such as the purge breadth, the intensity of violence, or even the nature of the purge. Further, allowing for the autocrat to choose reward in addition to punishment can lead to an increase rather than a decrease in violence.

All these findings are specific to many-to-one accountability (in Online Appendix E, we highlight some of the main differences with a single-agent setting). Many-to-one accountability cannot be approximated by a one-to-one setting. In a host of hierarchical settings, such as large firms or the army, the principal shares the same problem as the autocrat, with different tools at her disposal (mass layoffs, up or out promotion instead of purges). We thus believe the approach developed in this paper has a large range of applications for the study of public and private organizations.

References

- Acemoglu, Daron and Alexander Wolitzky, "The economics of labor coercion," *Econometrica*, 2011, 79 (2), 555–600.
- -, Georgy Egorov, and Konstantin Sonin, "Coalition formation in non-democracies," The Review of Economic Studies, 2008, 75 (4), 987–1009.
- _ , _ , and _ , "Do Juntas Lead to Personal Rule?," The American Economic Review: Papers & Proceedings, 2009, 99, 298–303.
- Ashworth, Scott, "Electoral accountability: recent theoretical and empirical work," Annual Review of Political Science, 2012, 15, 183–201.
- Bernhardt, Dan and Steeve Mongrain, "The Layoff Rat Race," The Scandinavian Journal of Economics, 2010, 112 (1), 185–210.
- Bloch, Francis and Vijayendra Rao, "Terror as a bargaining instrument: A case study of dowry violence in rural India," *The American Economic Review*, 2002, 92 (4), 1029–1043.
- Bó, Ernesto Dal, Pedro Dal Bó, and Rafael Di Tella, "Plata o Plomo?: Bribe and Punishment in a Theory of Political Influence," American Political Science Review, 2006, 100 (01), 41–53.
- Brzezinski, Zbigniew K, The permanent purge, Harvard University Press Cambridge, MA, 1956.
- Campbell, Bruce B, "The SA after the Röhm purge," Journal of Contemporary History, 1993, 28 (4), 659–674.
- Conquest, Robert, The great terror: A reassessment, Oxford University Press on Demand, 2008.
- Coughlin, Con, Saddam: His Rise and Fall, Harper Collins, 2005.
- Dallin, Alexander and George W Breslauer, Political terror in communist systems, Stanford University Press, 1970.
- **Dam, Nikolaos Van**, The struggle for power in Syria: Politics and society under Asad and the Ba'th Party, IB Tauris, 2011.

- de Mesquita, Bruce Bueno and Alastair Smith, "Political Succession: A Model of Coups, Revolution, Purges and Everyday Politics," *Journal of Conflict Resolution*, 2015, 61 (4), 707 – 743.
- Dewan, Torun and David P Myatt, "Scandal, protection, and recovery in the cabinet," American Political Science Review, 2007, 101 (01), 63–77.
- and _ , "The declining talent pool of government," American Journal of Political Science, 2010, 54 (2), 267–286.
- Dikötter, Frank, The Cultural Revolution: A People's History, 19621976, Bloomsbury Publishing, 2016.
- Dragu, Tiberiu and Adam Przeworski, "Preventive Repression: Two Types of Moral Hazard," https://ssrn.com/abstract=3011020 2017.
- Egorov, Georgy and Konstantin Sonin, "Dictators And Their Viziers: Endogenizing The Loyalty–Competence Trade-Off," *Journal of the European Economic Association*, 2011, 9 (5), 903–930.
- and _ , "The Killing Game: Reputation and knowledge in politics of succession," Research in Economics, 2015, 69 (3), 398–411.
- Fitzpatrick, Sheila, "Stalin and the Making of a New Elite, 1928-1939," *Slavic Review*, 1979, *38* (3), 377–402.
- Gailmard, Sean and John W Patty, "Formal models of bureaucracy," Annual Review of Political Science, 2012, 15, 353–377.
- Gehlbach, Scott and Alberto Simpser, "Electoral manipulation as bureaucratic control," American Journal of Political Science, 2015, 59 (1), 212–224.
- Getty, John Arch, Origins of the great purges: the Soviet Communist Party reconsidered, 1933-1938, Vol. 43, Cambridge University Press, 1987.
- **Gregory, Paul R**, Terror by Quota: State Security from Lenin to Stalin: (an Archival Study), Yale University Press, 2009.

- -, Philipp JH Schröder, and Konstantin Sonin, "Rational dictators and the killing of innocents: Data from Stalins archives," *Journal of Comparative Economics*, 2011, 39 (1), 34–42.
- Hohne, Heinz, The Order of the Death's Head: The Story of Hitlers' SS, Pan Books, 1981.
- Jiang, Junyan, B. Pablo Montagnes, and Stephane Wolton, "Political Tournaments: Accountability and Selection," 2017.
- Landa, Dimitri and Scott A Tyson, "Coercive Leadership," American Journal of Political Science, 2017, 61 (3), 559–574.
- Machiavelli, Niccolò, The prince, OUP Oxford, 2005.
- Morgan, Philip, "The Trash Who are Obstacles in Our Way: the Italian Fascist Party at the Point of Totalitarian Lift Off, 1930–31," *The English Historical Review*, 2012, 127 (525), 303–344.
- New York Times Magazine, "Inside Turkey's purge," https://www.nytimes.com/2017/04/ 13/magazine/inside-turkeys-purge.html April 2017.
- **Orlow, Dietrich**, *The history of the Nazi party: 1933-1945*, Vol. 2, [Pittsburgh]: University of Pittsburgh Press, 1969.
- Rigby, Thomas Henry, "Communist Party membership in the USSR," Princeton: Princeton University Press, 1968, 178, 369–491.
- Rozenas, Arturas, "The Logic of Collective Repression," http://cpd.berkeley.edu/ wp-content/uploads/2017/03/CollectiveRepression1.2.pdf 2017.
- Sayers, Michael and Albert Eugene Kahn, *The great conspiracy against Russia*, Boni & Gaer, 1946.
- Schapiro, Leonard Bertram, The government and politics of the Soviet Union, Taylor & Francis, 1977.
- Shadmehr, Mehdi and Raphael Boleslavsky, "Institutions, Repression, and the Spread of Protest," https://ssrn.com/abstract=2731749 2017.
- Svolik, Milan W, "Power sharing and leadership dynamics in authoritarian regimes," American Journal of Political Science, 2009, 53 (2), 477–494.

- Teiwes, Frederick C, Politics and Purges in China: Rectification and the Decline of Party Norms, 1950-1965, ME Sharpe, 1993.
- Turkey Purge Website, https://turkeypurge.com/ 2017.
- Tyson, Scott A., "The Agency Problem Underlying Repression," *Journal of Politics*, 2017, *forth-coming.*
- Weinberg, Robert, "Purge and Politics in the Periphery: Birobidzhan in 1937," *Slavic review*, 1993, 52 (1), 13–27.
- Wolton, Thierry, Histoire mondiale du communisme, tome 1: Les bourreaux, Grasset, 2015.
- Zakharov, Alexei V, "The Loyalty-Competence Trade-Off in Dictatorships and Outside Options for Subordinates," The Journal of Politics, 2016, 78 (2), 457–466.

Historical data

Year	Proportion	Source
1919	10-15	Getty (1987, Table 2.1)-Rigby (1968, 76)
1921-23	25	Getty (1987, Table 2.1)-Rigby (1968, 97)
1924	3	Getty (1987, Table 2.1)
1925	4	Getty (1987, Table 2.1)
1926	3	Getty (1987, Table 2.1)
1927	6	Rigby (1968, 127)
1928	13	Getty (1987, Table 2.1)
1929	11	Getty (1987, Table 2.1)
1930	5	Getty (1987, Table 2.1)
1931	5	Getty (1987, Table 2.1)
1933-34	17-22	Getty (1987, 55)-Rigby (1968, 204)
1935	9-13	Getty (1987, Table 7.1)-Rigby (1968, 209) ⁺
1936	10	Rigby $(1968, 209)^+$
1937	5	Getty (1987, Table 7.1)
1951-53	5	Rigby $(1968, 281)^+$

Table 1: Proportion of party members purged in USSR

All proportions are approximation. + denotes authors' calculation using Rigby (1968, 52)

Year	Proportion	Source
1947-48	23	Teiwes $(1993, 75)^+$
1951-54	10	Teiwes (1993, 110)
1957-58	9	Teiwes (1993, 268)
1959-60	20	Teiwes (1993, 339)
1964-65	10	Teiwes (1993, 425)

Table 2: Proportion of party members purged in China

All proportions are approximation. + denotes authors' calculation