Mass Purges: Top-down Accountability in Autocracy

B. Pablo Montagnes

Assistant Professor

Department of Political Science, Emory University

Email: pablo.montagnes@emory.edu

Stephane Wolton

Associate Professor

Department of Government, London School of Economics

Email: S.Wolton@lse.ac.uk

Abstract

This paper proposes a novel theoretical framework to study the features of 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 show that the set of purged agents is well delineated in mild purges, whereas no performance indicator is a guarantee of safety in violent purges. The proportion of purged agents is non-monotonic in the intensity of violence. For the autocrat, increasing the intensity of violence always raises performance, but it improves the selection of subordinates only if violence is low to begin with. Hence, even absent de jure checks, the autocrat is de facto constrained by her subordinates' strategic behavior. We use historical (including the Soviet purges and the Cultural Revolution) and recent (the Erdogan purge) events to illustrate our key theoretical findings.

Word count: approx. 11,815 words

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., Josef Stalin at the 13th Party Congress), 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 the thirteenth 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 petty-bourgeois elements within it." 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 1980-84, Turkey in 2016-17).

What are these purges that Lenin advocates? What is their purpose? What are their common features? This paper contends that these purges are an instrument of 'top-down accountability,' a response to the autocrat's political problem of motivating her agents and selecting the most suitable subordinates within the single party or state bureaucracy. Affecting thousands, if not millions of individuals, these mass purges, furthermore, are a case of many accountable to one.

This paper provides a theoretical framework to study 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 autocrat's 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).

We focus on three characteristics of purges and study how they correlate with each other. First, we consider the intensity of violence, defined as the direct cost of being purged. Second, we look at the link between agents' performance and their risk of being purged. Third, we analyze the proportion of subordinates purged, referred to as the 'purge breadth.' We only find a clear

relationship between violence and the risk of being purged. For low intensity of violence, the purge solely targets bad performers; for high intensity, under some conditions, no performance indicator is a guarantee of safety. Breadth, in contrast, does not vary monotonically with the intensity of violence.

We also consider the impact of violence on two key indicators for the autocrat: average performance of subordinates and the selection of suitable agents. A higher intensity of violence always encourages subordinates to work harder. Greater violence, however, improves selection only when violence is low to begin with. In many-to-one accountability settings, performance and selection move together in some circumstances, but move in opposite directions in others.

Our baseline framework consists of a two-period game with an autocrat and a mass of subordinates. Each agent is either congruent (sharing the autocrat's ideological objectives or appealing to her predilections) or non-congruent (opportunists who only care about the benefits reserved to regime insiders or possible opponents) and his degree of congruence (type) is his private information. Each subordinate exerts effort to increase the probability his individual and independent project is successful (e.g., to meet the assigned quotas). The autocrat seeks to maximize the performance of her agents. After observing each agent's project outcome (success or failure), the autocrat chooses, at a cost, a proportion of failures and successes to be purged or, equivalently, a purge breadth. For the autocrat, purging is beneficial as purged agents are replaced by new subordinates from an available replacement pool with the same proportion of congruent types as existing subordinates. For the agents, being purged is costly. They lose the material benefits reserved to regime insiders and face a direct cost due to the violence of the purge (e.g., fine or deportation). To provide a possible rationale for the intensity of violence in purges, we assume that the autocrat can credibly commit to a certain intensity of violence at the beginning of the game.

As congruent subordinates benefit more from successful projects, they exert greater effort than their non-congruent counterparts. As a result, on average, subordinates who succeed are more congruent and agents who fail are less congruent than their possible replacements. The autocrat, who seeks to maximize the proportion of congruent agents in the second period, thus, only purges from the failure pool (i.e., the set of subordinates who fail). In our baseline model, the set of purged agents is clearly delimited and we refer to this form of purges as 'discriminate' purges. Nonetheless, since project outcomes are only partially correlated with types, in all purges, screening is imperfect: Some congruent agents are eliminated and some non-congruent subordinates survive.

Things becomes more involved when we consider the consequences of a change in the intensity of violence. Greater violence motivates all agents, congruent and non-congruent, to exert more effort. We show that, consequently, the autocrat's assessment of failures worsens. When some failures survive (which can occur because purging is costly), the proportion of failed agents who are purged increases with violence. This further motivates subordinates which, in turn, further decreases the autocrat's evaluation of failed agents, triggering the same response by the autocrat. We refer to this reinforcing effect as the 'pool makeup effect.' As a consequence of these direct and indirect effects, a higher intensity of violence yields a wider purge breadth (despite the increased effort). Greater violence also raises the proportion of congruent types amongst second-period subordinates as more failures are eliminated and replaced by better agents.

Above a certain intensity threshold, all failures are purged. The pool makeup effect no longer plays a role. Greater effort induced by increased violence decreases the size of the failure pool and, thus, the purge breadth. As fewer agents failed and, as a result, are purged, selection then no longer improves with violence: Under the assumptions of the baseline model, increased intensity of violence has no effect on the proportion of second-period congruent subordinates.

Our analysis reveals a rich set of positive results describing the relationship between the intensity of (potentially exogenous) violence and equilibrium behavior. For external observers, greater violence always increases the proportion of failures purged, but has an ambiguous effect on the purge breadth. For the autocrat, more violence always raises effort, but it improves selection only when its intensity is low to begin with. These findings can also be thought of as inputs in the autocrat's choice of an optimal intensity of violence. They then imply that the autocrat's marginal benefit of investing in violence is especially high for low intensity of violence. The marginal benefit, in turn, is relatively low for high intensity suggesting that the autocrat replace all failures only in specific circumstances.

We also study the features of mass purges when the replacement pool is of higher quality than existing subordinates. If the difference between the two pools is large enough, the autocrat may have incentives to purge agents from the success pool. Still, the purge is not random as all failures are purged, whereas only some successful agents are replaced. We label this form of purge as 'semi-indiscriminate.' Semi-indiscriminate purges, we show, only occur when the intensity of violence is high. With high violence comes great effort and two important consequences. First, the failure pool is thin, resulting in a low marginal cost of purging successful agents. Second, the success pool is tainted, resulting in a relatively high marginal benefit of purging successful agents. As the intensity

of violence increases in a semi-indiscriminate purge, the success pool further deteriorates yielding greater purge breadth and worse selection. Hence, semi-indiscriminate purges are characterized by high violence and high effort, but relatively poor selection as the autocrat is never certain what lurks behind the mask of conformity (Dallin and Breslauer, 1970).

Historical evidence provides a useful illustration of some of the patterns we uncover. Our findings are broadly consistent with both the differences and similarities between the Stalinist purges of the thirties and the Maoist purges of the fifties. 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."

Related formal literature

Our paper studies the main features of mass purges. By contrast, the literature has emphasized elite purges: the shaping of an autocrat's inner circle and the resulting contests for power (e.g., Acemoglu et al., 2008; Svolik, 2009; Bueno de Mesquita and Smith, 2015). Mass purges can complement power struggles (e.g., screening agents favorable to her opponents), but need not (e.g., cleansing the single-party of opportunists). A critical aspect of our analysis is that mass purges are an instrument of accountability used to motivate and select subordinates.

As such, our paper is in close connection with a large literature on agency problems in economics (see Laffont and Martimort, 2009 for a textbook treatment) and political science (see Ashworth, 2012 for a review). Most of this literature consider the relationship between one principal and one (possibly representative) agent. Some papers, however, analyze settings in which a principal faces multiple agents, as we analyze in our framework. Beginning with Holmström (1982), several scholars have examined the problem of moral hazard in teams. Others have looked at agents' effort on an individual project when promotion is the only tool available to the principal (e.g., Lazear and Rosen, 1981; Green and Stokey, 1983). Yet, another strand of the literature studies the ex-ante principal's choice of one agent among many in democracy (e.g., Dewan and Myatt, 2010; Dewan and Hortala-Vallve, 2011; Gailmard and Patty, 2012)) or in autocracy (e.g., Egorov and Sonin, 2011; Zakharov, 2016). Few of these works focus on the relationship between performance and selection when the principal faces many agents as is the case in our paper (one exception is Jiang et al., 2017). Further, our paper innovates on the literature by looking at the principal's choice

of replacing a *mass of subordinates* and, as a result, studies different strategic interactions than previous works.

Our paper also joins a small literature interested in the impact of violence. In Esteban et al. (2015), a leader can kill a mass of opponents to increase her chances of survival, but, by doing so, lowers her rents in the future. In bargaining settings, Bloch and Rao (2002) and Dal Bó et al. (2006) highlight how violence, or the threat thereof, strengthens a player's position and thus improves his payoff. Acemoglu and Wolitzky (2011) model slavery as a principal-agent relationship in which the slave-owner can reduce the value of her slave's outside option and, thus, his wage using coercion. Landa and Tyson (2017) highlight how coercive leadership is necessary to coordinate subordinates who are uncertain of their leader's preferences. All of these works focus on various moral hazard problems, ignoring adverse selection, and so cannot examine the effect of violence on screening and selection as our paper does.

Evidence on mass purges

Before proceeding, we summarize historical evidence on mass purges, highlighting some of 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 rectification campaigns, which happened under Mao Zedong in 1947-48, 1950, 1951-54, 1953, 1957, 1957-58, 1959-60, 1960-61, 1962-63, and 1964-65. Due to their specificities, these lists do not include the Great Terror in the USSR (1936-38) and the Chinese Cultural Revolution (1966-76). We return to these two events in the 'Discussion' section when 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) and leaders in the USSR and China sought to regulate their periodicity (Getty, 1987; 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. All rank-and-file party members, representing *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. Accusations of opportunism, however, may simply have been a way to easily get rid of recalcitrant members and to hide substantial opposition within the party. Indeed, Rigby (1968) 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 occurred in other communist regimes across the world (e.g., Albania in 1948-52, Bulgaria in 1948-50, East Germany in 1950-53 and 1956-60, Hungary in 1950, 51, and 55, North Korea repeatedly before 1957; see Janda, 1980) and even in communist parties in democracies (e.g., the exiled Spanish communist party in France and the French communist party; see Wolton, 2017). But mass purges were not just a communist phenomenon. In Nazi Germany, the Sturmabteilung (SA) was purged in 1935-6 following the "Röhm putsch," the Schutzstaffel (SS) was purged in 1933-35 (Hohne, 1981), and the party itself (NSDAP) was purged in 1938 (Orlow, 1969). Lack of commitment to the Nazi cause 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, with opportunism again the primary cause of removal (Morgan, 2012). In the 1980s, 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). By contrast, Saddam Hussein seems to have used the purge of 1979 in Iraq to tighten his hold on power (Coughlin, 2005). In all of these historical cases, directly observable traits did not seem to play a major role in the autocrat's purging decision (though new members were targeted disproportionately). Instead,

the authoritarian leader arguably had to rely on indirect indicators, as in the framework we now describe.

Set-up

This section and the five that follow present and analyze our theoretical framework as well as various extensions. In the 'Discussion' section, we summarize our main findings and use historical and recent events to illustrate them. A non-technical reader may wish to proceed directly to this section 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. 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 $e_t^i \in [0, 1]$ on an individual and independent project at cost $\frac{(e_t^i)^2}{2}$. The project can succeed (denoted $\omega_t = S$) or fail ($\omega_t = F$). The probability agent i's project is successful equals e_t^i. The autocrat observes the first-period outcome of the agent's project, but not his effort. We denote α_F and α_S as the proportion of agents who fail and succeed, respectively.

After observing all first-period project outcomes, the autocrat decides to purge a proportion κ_F of agents who failed and κ_S of those who succeeded, yielding a purge breadth of $\kappa = \alpha_F \times \kappa_F + \alpha_S \times \kappa_S$. A purge is costly. This cost is captured by the function $C(\kappa)$ with $C(0) = 0$ and (for ease of exposition) *marginal* cost $C'(\kappa) = C_0 + C_1 \times \kappa$, $C_0 \geq 0$, $C_1 > 0$. When a subordinate is purged, he is replaced by a new agent drawn from the replacement pool. The proportion of congruent types among the replacement pool is r with $r = \lambda$ (we relax this assumption when we discuss another form of purges and in our extensions below).

A purged agent faces (i) the loss of privileges associated with being a regime insider and (ii) a direct loss L which corresponds to the 'intensity of violence' of the purge. The autocrat determines the intensity of violence at the beginning of the game at a cost $\zeta(L)$ with $\zeta(0) = 0$ and *marginal* cost $\zeta'(L) = \zeta_0 + \zeta_1 L$, $\zeta_0 \geq 0$ and $\zeta_1 > 0$.

In period 1, an agent enjoys a benefit $R \geq 0$, which captures all 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 his project and his 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 that

only complicate the analysis, we further impose $v(S, nc) \geq 0$. When purged, $k^i = 1$, an agent suffers the loss $L \geq 0$. Agent i's first-period payoff thus assumes the following form:

$$u_1^i(e_1^i, \tau) = R + (1 - k^i) \times v(\omega_1, \tau) + k^i \times (-L) - \frac{(e_1^i)^2}{2}. \tag{1}$$

In period 2, there is no subsequent purge and a (surviving or new) subordinate i's payoff can be expressed as the sum of the benefit R and the net gain from a successful project:

$$u_2^i(e_2^i, \tau) = R + v(\omega_2, \tau) - \frac{(e_2^i)^2}{2}. \tag{2}$$

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 she receives 0 otherwise. The autocrat, thus, seeks to maximize the proportion of successful projects, which is equal to agents' average effort in each period. In the first period, the autocrat also bears the costs of investing in the intensity of violence and of purging. Letting $\bar{e}_t$ 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:

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

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

The autocrat discounts the future with factor $\beta \in (0, 1)$, 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 $L \geq 0$;

2. Agent i chooses his effort e_1^i after privately observing his type $\tau^i \in \{c, nc\}$;

3. Each project outcome ($\omega_1 \in \{S, F\}$) is determined by Nature and observed by the autocrat. The autocrat chooses to purge a proportion κ_F of agents who failed and κ_S of agents who succeeded;

4. Purged agents are replaced, and the first-period payoffs are realized.

Period 2:

1. Surviving or new agent i chooses effort e_2^i;

2. Each project outcome ($\omega_2 \in \{S, F\}$) is determined;

3. The game ends and the second-period payoffs are realized.

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 her purging strategy. For simplicity, we assume that agents are anonymous, all agents who fail face an identical probability of being purged, and all agents who succeed face the same risk of being removed. To avoid uninteresting PBE in which all agents exert no effort, we further impose that when the autocrat observes success as an out-of-equilibrium event, she treats the deviation as a mistake and does not distinguish between the agent who exerted effort and other subordinates who followed their prescribed strategy.[1] If, after these restrictions, multiple PBE arise, we select the one which maximizes the purge incidences; 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 of our refinements.

Before proceeding to the analysis, a few remarks on the set-up are in order. First, we remain agnostic about what constitutes non-congruent types. It could correspond to opportunistic agents attracted to the regime for the associated benefits of being an insider (Getty, 1987), subordinates who lacked "a wholehearted commitment to the Party's cause" (Teiwes, 1993, 114-115) or to the leader, or agents not satisfied with the regime's line (Gregory, 2009). In our theory, it only matters that the autocrat prefers some types of subordinates over others.

We also suppose that the autocrat is relatively well informed. As it is common in the agency literature, the autocrat does not observe her agents' effort. She, however, learns the project outcome of all agents (in Appendix F.1, we show that our results hold when the autocrat only learns about a subset of outcomes). Historically, communist leaders in China or the USSR were sending special commissions or work teams from higher up to evaluate local performance (Rigby, 1968; Teiwes, 1993). Hence, project outcomes can be understood as fulfilling quotas, dealing with problem cases, or ensuring the provision of local services.

Finally, both carrying out the mass purge and investing in the intensity of violence are costly. 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. When it comes to the intensity of violence, fines—which can correspond to a low L from being

---

[1]Other refinements could be used to eliminate this 'no-effort' PBE. For example, we could assume that there exist two subsets of agents ξ_0 and ξ_1 who always exert (respectively) effort 0 and 1. Alternatively, we could assume that the minimum level of effort is $\epsilon > 0$ so success is never an out-of-equilibrium event.

purged—requires less infrastructure and personnel than killing subordinates, deporting their spouse, and sending their children to orphanages, a very high L quite common in Stalin's USSR (Brzezinski, 1956; Conquest, 2008).

Note that the autocrat commits to an intensity of violence at the beginning of the game. This assumption is not innocuous. Absent commitment, at the time of the purge, the unique PBE features no violence if violence is costly (violence has no effect on selection and effort then) or any intensity can be part of a PBE if violence is costless. With commitment, we offer one possible rationale for the variation in the intensity of violence of purges across countries and over time. From the onset, let us stress that all results from the baseline model described here remain unchanged if the autocrat can commit to both the intensity of violence (L) and the purge incidences (κ_F and κ_S) at the beginning of the game (see Appendix F.2).

Early commitment to an intensity of violence can be thought of as earmarking funds for the purge (like in the case of the Great Terror, Wolton, 2015). This assumption can also be interpreted as investment in the infrastructure of violence such as the Solovski prison camp (opened in the USSR in 1923) or the Qincheng Prison (opened in China in 1958). The cost function $C(\kappa)$ then corresponds to the cost of filling these coercive structures with purged subordinates (in this case, C_0 and C_1 may be a function of L; our model can easily accommodate this possibility and all of our results would hold as long as the two parameters do not increase too fast with the intensity of violence).

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 $V_2(\tau) = R + \frac{(v(S,\tau))^2}{2}$. For the autocrat, $W_2(\tau)$ denotes her second-period expected payoff induced by an agent of type $\tau \in \{c, nc\}$, with $W_2(\tau) = v(S, \tau)$. To limit the number of cases, we impose $\beta r(W_2(c) - W_2(nc)) < C_0 + C_1$ so the autocrat never purges all agents. Finally, to simplify the analysis, we assume that the highest feasible intensity of violence, which we denote $\overline{L}$, satisfies $\overline{L} := 1 - v(S, c) - V_2(c)$ so that no subordinate exerts effort 1.

Effort and incentive to purge

In this section, we determine the agents' equilibrium effort and the purge incidences κ_F, κ_S for a given intensity of violence L. Our first result, quite intuitively, states that in any equilibrium, the autocrat only purges from the failure pool (i.e., the set of agents who fail in their project).

Lemma 1. *In equilibrium, $\kappa_F \in [0,1]$ and $\kappa_S = 0$.*

Proof. All proofs are collected in the Supplemental Appendix. □

Due to our equilibrium refinements, there is no equilibrium in which agents exert zero effort.[2] When they exert effort, agents endogenously sort into failure and success pools, which constitute, with the proportion (λ) of congruent types among current subordinates and potential replacements, the only information available to the autocrat at the time of her purging decision. Given that congruent types receive a greater intrinsic benefit from succeeding, they always exert more effort than their non-congruent counterparts and are more likely to belong to the success pool (i.e., the single-crossing condition holds). Consequently, success becomes a positive signal of congruence, and failure a negative one. The autocrat never purges successful agents then since she (correctly) believes the success pool is strictly better than the replacement pool.

When deciding his level of effort, a subordinate takes into account his present and future payoff from succeeding (since he survives the purge) and the risk of facing a loss L and of forgoing future payoffs when he fails. This risk is proportional to the (anticipated) proportion of failures being purged, κ_F. Greater κ_F, therefore, translates into higher effort.

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

$$e_1^i(\tau) = v(S, \tau) + \kappa_F(V_2(\tau) + L). \tag{5}$$

Having characterized the agents' effort choices, we now turn to the autocrat's incentive to purge. After observing all project outcomes, the autocrat forms a posterior μ_F that a subordinate who fails ($\omega_1 = F$) is congruent, with $\mu_F < \lambda$ as discussed above. The autocrat then chooses the purge incidence, κ_F, to equate the marginal cost of purging an additional agent—$C_0 + C_1\alpha_F \times \kappa_F$—with its associated marginal benefit. The marginal benefit is a function of two elements: (i) the increased probability that the replacement is congruent, $\lambda - \mu_F$, and (ii) the expected payoff gain from having a congruent agent rather than a non-congruent subordinate in period 2, $\beta(W_2(c) - W_2(nc))$. The autocrat's expected marginal benefit from purging an additional failure is:

$$\mathcal{W}^F = [r - \mu^F]\big(W_2(c) - W_2(nc)\big). \tag{6}$$

---

[2] Absent our restrictions, there would exist a 'no-effort' PBE in which the autocrat would set $\kappa_S = 1$. This PBE would be sustained by the unintuitive out-of-equilibrium belief that an agent who succeeds is likely to be non-congruent even though $v(S, c) > v(S, nc)$.

In our baseline model, there exists a clear demarcation between purged and non-purged agents. Success is always a guarantee of safety and failures risk being replaced. The purge is thus 'discriminate.' Project outcomes, however, are only a noisy signal of types (by Equation 5, a congruent agent exerts effort less than 1, while a non-congruent subordinate exerts positive effort). Hence, in any purge, some congruent agents are purged and some non-congruent subordinates survive.

As purges are costly, not all failures may be removed ($\kappa_F \in (0,1)$). This case exhibits an interesting property. To describe it, suppose that κ_F increases exogenously. Both congruent and non-congruent types then raise their effort in response to a higher threat of being purged conditional on failure. Congruent agents, 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 more congruent agents (in relative and absolute terms) exit the failure pool. The autocrat's evaluation of subordinates who fail then deteriorates (μ_F decreases) and her marginal benefit of purging increases (see Equation 6).

The analysis above highlights a feedback mechanism whereby an anticipated increase in the purge incidence yields higher effort by all agents and, in turn, a greater incentive to purge for the autocrat. We label this reinforcing effect the 'pool makeup effect.' As the pool makeup effect plays a crucial role below, we introduce the following definition as a point of reference.

Definition 1. *The pool makeup effect corresponds to the negative effect of the purge incidence* $\kappa_F \in [0,1)$ *(as anticipated by the agents) on the autocrat's posterior upon failure:* $\frac{\partial \mu_F}{\partial \kappa_F} < 0$.

Having examined the agents' efforts and the autocrat's incentive to purge, we turn in the next section to the consequences of (for now) exogenous changes in the intensity of violence.

The consequences of violence

Greater intensity of violence increases subordinates' efforts (see Equation 5). This direct effect of L on agents' action, in turn, induces a series of indirect effects that needs to be carefully unpacked.

Let us first consider the indirect effect of violence on effort through the purge incidence, assuming not all failures are purged ($\kappa_F \in (0,1)$). To do so, fix the purge breadth $\kappa = \alpha_F \times \kappa_F$ (the proportion of agents who fail times the proportion of failures purged). As the intensity of violence increases, all subordinates exert more effort, and the pool of failures shrinks ($\alpha_F \downarrow$). As the purge breadth is held constant by design, the purge incidence must necessarily increase ($\kappa_F \uparrow$), which, in turn,

induces greater effort. Thus, the indirect effect through purge incidence is positive. Higher purge incidence then generates a pool makeup effect (Definition 1). The autocrat has a greater incentive to purge more failures, further increasing the purge incidence, which again induces even greater effort.

As efforts increase at a relatively high rate with the intensity of violence due to the direct and indirect effects, the failure pool is quickly exhausted (fewer subordinates fail and more failures are purged). Consequently, there exists a threshold of violence—denoted $L^{full} \geq 0$—above which all failures are purged. Using the superscript '$*$' to denote equilibrium values, we thus obtain:

Proposition 1. *There exists $L^{full} > 0$, unique whenever $L^{full} < \overline{L}$, such that:*
(i) Some agents who fail survive ($\kappa_F^*(L) \in (0,1)$) if and only if $L < L^{full}$;
(ii) All agents who fail are purged ($\kappa_F^*(L) = 1$) if and only if $L \in [L^{full}, \overline{L}]$ (possibly an empty interval).

To simplify the exposition, we assume in what follows that $L^{full} < \overline{L}$ so all failures may be purged if violence is high enough (we provide a precise condition for this inequality to hold in the Online Appendix).

As noted above, as the intensity of violence increases, the autocrat's assessment of the failure pool deteriorates (μ^F declines). When some failures survive, the equilibrium purge breadth is determined by equating the marginal cost and the marginal benefit of purging. Greater marginal benefit then yields greater purge breadth. When the autocrat purges all failures, the breadth equals the size of the failure pool, which shrinks with the intensity of violence. The relationship between the breadth $\kappa^*(L)$ and the intensity of violence L is, thus, non-monotonic.

Proposition 2. *The equilibrium purge breadth, $\kappa^*(L)$, satisfies:*
(i) For $L < L^{full}$, $\kappa^*(L)$ is strictly increasing with L;
(ii) For $L \in [L^{full}, \overline{L}]$, $\kappa^*(L)$ is strictly decreasing with L.

We now turn to the impact of violence on effort. While greater violence always generates better first-period performance, the responsiveness of effort to L depends on how many failures are purged and, thus, the original intensity of violence (see Proposition 1). When some failures survive, as violence increases, direct and indirect effects complement each other, inducing a relatively large response in subordinates' efforts. In turn, when all failures are purged, there is no indirect effect. Effort then increases linearly with violence (Equation 5 with $\kappa_F = 1$).

Proposition 3. *The total derivative of average effort with respect to violence $\frac{d\bar{e}(L)}{dL}$ is always strictly positive. Further, there exist $L^{eff} \geq 0$, unique if $L^{eff} < L^{full}$, such that the derivative satisfies:*

(i) $\frac{d\bar{e}(L)}{dL} > 1$ for all $L \in (L^{eff}, L^{full})$;

(ii) $\frac{d\bar{e}(L)}{dL} = 1$ for all $L \in [L^{full}, \overline{L}]$.

We finally consider the relationship between violence and selection. As a first observation, note that a new subordinate is always more likely to be congruent than a purged agent. Thus, the purge always raises the proportion of congruent types among second-period subordinates. Here, we study the impact of *increasing* the intensity of violence on screening and selection.

Consider first the consequence of greater violence on screening (i.e., the proportion of congruent types among surviving subordinates). When some failures survive ($\kappa_F^*(L) \in (0, 1)$, equivalently $L < L^{full}$), a greater intensity of violence is associated with fewer agents failing and more failures being purged. Among surviving subordinates, a greater proportion of agents thus belongs to the success pool, which is of higher quality than the failure pool, resulting in better screening. When surviving agents all belong to the success pool ($\kappa_F^*(L) = 1$, or equivalently $L \geq L^{full}$), screening worsens because the success pool becomes more tainted as more non-congruent types enter the pool relative to the stocks of both types.

Let us now study the relationship between violence and selection (i.e., the proportion of congruent types among second-period agents, both old and new). When some failures survive ($L < L^{full}$), selection strictly improves with the intensity of violence. Higher intensity yields better screening of surviving agents and more failures are replaced by subordinates of higher quality. When all failures are purged (i.e., $L \geq L^{full}$), new subordinates are still better than purged agents, but there are fewer of them and screening deteriorates. In general, the overall consequences of violence on selection would be undetermined. Under our assumptions, the positive and negative effects fully cancel out and violence has no effect on selection.

Proposition 4.

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

(ii) The proportion of congruent types among second-period agents strictly increases with the intensity of violence if and only if $L < L^{full}$ and is constant otherwise.

Greater violence always engenders higher effort in the first period (Proposition 3). The relationship between violence and selection is more nuanced, and violence may impede screening

(Proposition 4). With these comparative statics, we can now turn to the autocrat's optimal choice of violence.

Intensity of violence

Recall that effort is very responsive to violence at low intensity (i.e., $L < L^{full}$) due to the direct effect (on payoff) and indirect effects (especially, the pool makeup effect) of violence. In addition, greater intensity also improves selection. As a result, the marginal benefit from increased violence is large *and increasing in L* for $L < L^{full}$. The marginal cost can then intersect the marginal benefit more than once, as Figure 1 illustrates. If so, the autocrat chooses between two possible maxima: the lowest intersection L^1 in Figure 1 and L^{full} (the highest intersection L^2 is a local minimum). The presence of multiple maxima implies that small changes in the underlying fundamentals can be associated with a large swing in equilibrium values, including the intensity of violence or the purge breadth.
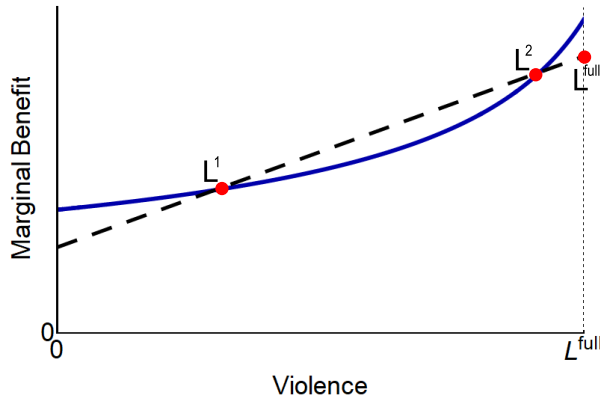

Figure 1: Marginal benefit of violence in a partially discriminate purge

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 $\zeta_1 = 4.8$.

Remark 1. *Small changes in the cost of purging (C_0, C_1) or in the cost of investing in violence (ζ_0, ζ_1) can be associated with large shifts in the observed intensity of violence and purge breadth.*

At $L = L^{full}$, the marginal benefit drops. There is no pool makeup effect so effort increases linearly with violence (Proposition 3), and the selection effect is null (Proposition 4). The benefit is then concave in the range $[L^{full}, \overline{L}]$. Consequently, the intensity of violence is such that all failures

17

are purged ($L^* \geq L^{full}$) if and only if the cost parameters determining the investment in violence are not too high.

Remark 2. *All failures are purged ($L^* \geq L^{full}$) only if the cost parameters of investing in violence ζ_0 and ζ_1 are below some thresholds $\zeta_0^{disc} > 0$ and $\zeta_1^{disc}(\zeta_0) > 0$.*

Another form of purges

So far, we have assumed that the proportion of congruent types is the same among existing subordinates and their possible replacements. In this section, we relax this assumption and suppose that the pool of replacement is better than the pool of first-period agents: $r > \lambda$.

When the replacement pool is of sufficiently high quality (r sufficiently large relative to λ), the autocrat may decide to purge agents from the success pool (i.e., $\kappa_S > 0$). The purge then appears somewhat indiscriminate: Now, even success does not inoculate a subordinate against the purge. However, as before, success is a signal of congruence, and failure is a signal of non-congruence. Hence, the autocrat always replaces all failures before purging agents belonging to the success pool. Formally, in any equilibrium, the purge incidence conditional on success satisfies $\kappa_S > 0$ only if $\kappa_F = 1$. This form of purges is not exactly random, and we label it as 'semi-indiscriminate.'

Turning to the agents' first-period effort, anticipating a semi-indiscriminate purge, each subordinate now anticipates that success is no longer sufficient to survive the purge. A type-τ agent then exerts effort ($\tau \in \{c, nc\}$):

$$e_1^i(\tau) = (1 - \kappa_S)(v(S, \tau) + V_2(\tau) + L) \tag{7}$$

Quite intuitively, effort is decreasing with the purge incidence κ_S since a greater risk of being purged conditional on success reduces the value of effort.

The autocrat's marginal benefit of purging a successful agent is, in turn:

$$\mathcal{W}^S = [r - \mu^S](W_2(c) - W_2(nc)) \tag{8}$$

Using Equation 7 and Equation 8, we proceed as in the section on the consequences of violence and describe the main observable features of semi-indiscriminate purges. A first characteristic of semi-indiscriminate purges is that they tend to be violent. For all intensity of violence, the marginal benefit of purging a successful agent is strictly lower than the marginal benefit of purging a failure (using $\mu_S > \lambda > \mu_F$ and comparing Equation 6 and Equation 8). When the intensity of violence

18

is such that all failures are purged (i.e., at $L = L^{full}$), the autocrat does not immediately start purging successful agents. Rather, there is a range of intensity of violence such that the purge remains discriminate.

As the intensity of violence increases, two combined forces make a semi-indiscriminate purge more attractive to the autocrat. First, greater violence yields the same increase in effort from congruent and non-congruent subordinates (see Equation 5 with $\kappa_F = 1$). The success pool then becomes more tainted as relatively more non-congruent agents enter the pool. Consequently, the autocrat's posterior, μ_S, decreases with L, and the marginal benefit of purging a successful agent increases. Second, greater effort by all subordinates reduces the size of the failure pool and the purge breadth, and the marginal cost of purging an additional agent decreases. Thus, there exists a threshold $L^{ind} > L^{full}$ such that the purge is semi-indiscriminate if and only if the intensity of violence is above L^{ind}.

A second characteristic of semi-indiscriminate purges is that they are not necessarily large. Indeed, they occur only if the failure pool is sufficiently depleted. However, once a semi-indiscriminate purge has begun ($L \geq L^{ind}$), the breadth again is positively correlated with the intensity of violence. As violence increases, the success pool is more tainted ($\mu_S \downarrow$), the marginal benefit of purging is greater, and, as a consequence, the purge breadth is wider.[3]

Proposition 5. *There exists $L^{ind} > L^{full}$, unique whenever $L^{ind} < \overline{L}$, such that:*
(i) The purge is semi-indiscriminate ($\kappa_F^*(L) = 1$ and $\kappa_S^*(L) > 0$) if and only if $L > L^{ind}$ and discriminate ($\kappa_S^*(L) = 0$) otherwise;
(ii) The purge breadth $\kappa^*(L)$ is strictly increasing with L for all $L \geq L^{ind}$.

Figure 2 illustrates the non-monotonic relationship between the intensity of violence and the purge breadth combining our earlier analysis (Proposition 2) with the result described in Proposition 5.(ii). The purge breadth is increasing with L when the intensity of violence is relatively low ($L < L^{full}$) or relatively high ($L \geq L^{ind}$). In turn, the correlation is negative for intermediary intensity of violence when the purge is discriminate and all agents are purged ($L \in [L^{full}, L^{ind}]$).

We now turn to effort and selection. In a semi-indiscriminate purge, the direct effect of violence on payoffs implies that greater intensity always increases effort. This positive impact, however, is

---

[3]Higher intensity of violence also increases the purge incidence $\kappa_S^*(L)$. Unlike in a discriminate purge, this has no impact on the autocrat's posterior, which does not depend on κ_S (omitting effort anticipation, $\mu_S = \dfrac{\lambda\left(v(S,c)+V_2(c)+L\right)}{\lambda\left(v(S,c)+V_2(c)+L\right)+(1-\lambda)\left(v(S,nc)+V_2(nc)+L\right)}$). There is no pool makeup effect in a semi-indiscriminate purge.
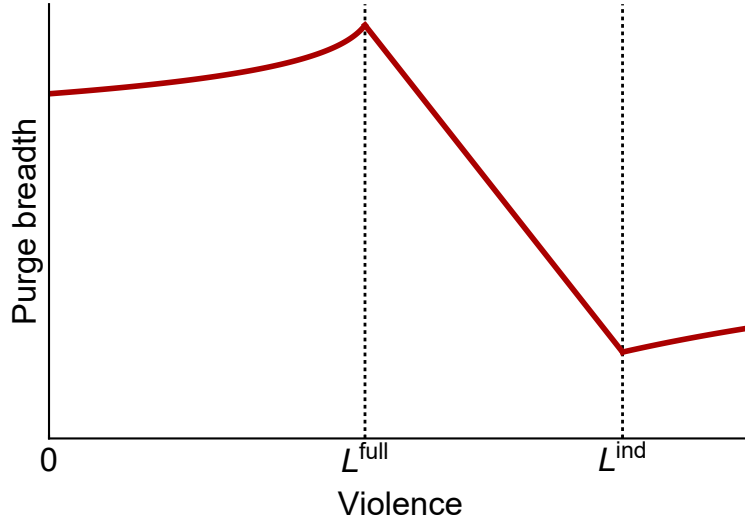
Figure 2: Equilibrium purge breadth and intensity of violence

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 $C_1 = 0.17$.

dampened by the increased purge incidence κ_S. Indeed, as the purge breadth increases, so does the risk that success is not enough to inoculate an agent against the purge. As a result, a subordinate's effort is generally less responsive to violence in a semi-indiscriminate purge than in a discriminate purge.

In a semi-indiscriminate purge, selection worsens as the intensity of violence increases. Greater violence yields a more tainted success pool from which survivors are drawn and, thus, leads to less efficient screening. Further, the purge breadth tends to increase slowly with violence (due to the low marginal benefit of purging). Combining these two effects, the overall congruence of second-period subordinates decreases with L as long as the quality of the replacement pool is not too high ($r \leq 2\lambda$, a sufficient condition).

The higher quality of the replacement pool relative to existing subordinates also implies that greater violence decreases the proportion of congruent second-period subordinates when the purge is discriminate and all failures are purged ($L \in [L^{full}, L^{ind}]$). Then, the purge breadth decreases with violence so that existing agents, who are worse on average than their potential replacements, constitute a greater proportion of the second-period subordinates. Overall, when the quality of the replacement pool is high ($r > \lambda$), an increase in violence improves selection only in a discriminate purge as more agents who fail are purged.

Proposition 6. *The total derivative of average effort with respect to violence $\frac{d\bar{e}(L)}{dL}$ satisfies $0 < \frac{d\bar{e}(L)}{dL} < 1$ for all $L \in [L^{ind}, \overline{L}]$.*

If $r \in (\lambda, 2\lambda]$, the proportion of congruent types among second-period agents decreases with L for all $L \geq L^{full}$.

Figures 3 and 4 illustrate the consequences of violence on effort and selection, combining the results of this section and those of our baseline model (Section 'The consequences of violence'). Figure 3 shows that effort increases at a high rate in a discriminate purge when some failures survive (for L close enough to L^{full}), at a relatively low rate in a semi-indiscriminate purge ($L > L^{ind}$), and at an intermediate speed when all failures are purged. Figure 4 highlights that screening (dashed line) and selection (solid line) improve with the intensity of violence only if violence is low and some failures survive.
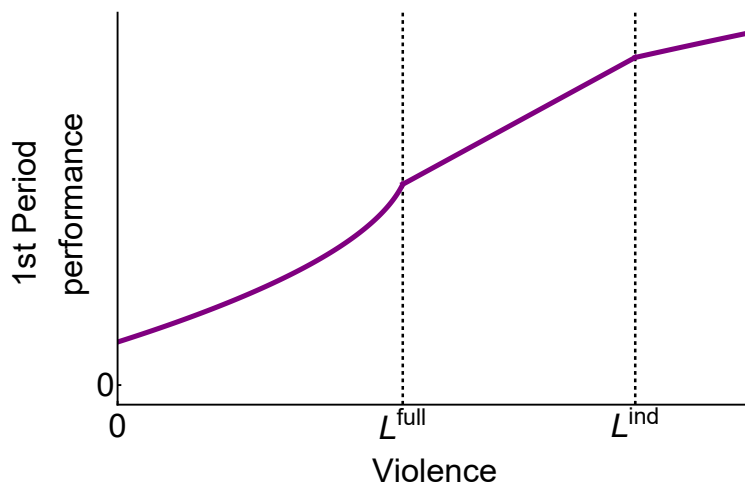

Figure 3: Violence and effort

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 $C_1 = 0.17$.

Overall, when the replacement pool is of high quality relative to existing subordinates, our analysis suggests that, at a high intensity of violence, the autocrat faces a trade-off between effort and selection when choosing L. Further, violence has a relatively low effect on first-period performance. So can semi-indiscriminate purges actually occur? The next remark states three conditions for a purge to be semi-indiscriminate.

Remark 3. *A purge is semi-indiscriminate ($L^* > L^{ind}$) 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 $\overline{C_1}(r, C_0) > 0$; and

3. The cost parameters ζ_0 and ζ_1 are, respectively, strictly below some $\overline{\zeta_0}(r) > 0$ and $\overline{\zeta_1}(r, \zeta_0) > 0$.
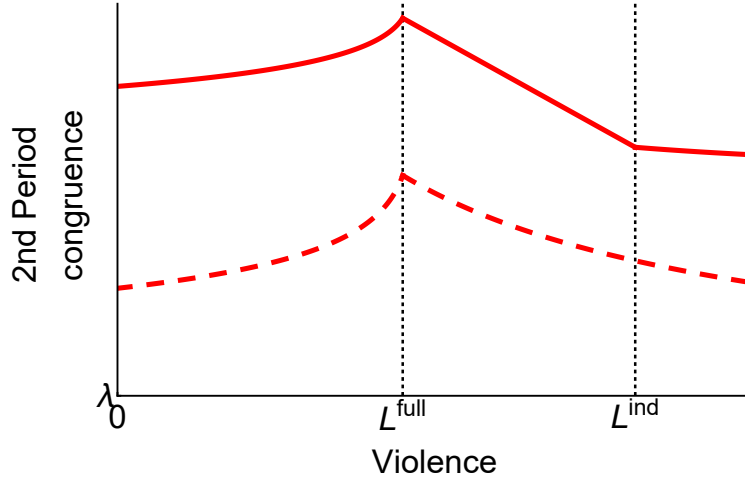
Figure 4: Violence and selection

The solid line corresponds to the proportion of congruent types among second-period subordinates, the dashed line to the proportion of congruent types among surviving agents. 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 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). 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, the cost of investing in violence must also be relatively small so that the autocrat is willing to choose a high L despite the low marginal benefit of doing so (condition 3).

Extensions

In this section, we briefly discuss three extensions of our baseline model, with the full formal analysis available in Online Appendix E.

Decreasing replacement pool

Throughout our analysis, we have assumed the autocrat suffers a cost from purging agents. Alternatively, the quality of the replacement pool may decrease with the purge breadth κ. The marginal replacement then has a probability $r(\kappa)$ of being congruent, with $r'(\kappa)$ strictly negative. In the special case when $r(\kappa)$ is linear (e.g., $r(\kappa) = \bar{r} - r_1\kappa$), the average congruence of new subordinates is $\bar{r} - r_1\frac{\kappa}{2}$. The autocrat then faces a quadratic loss in the quality of the replacements as the purge breadth increases, and the problem is isomorphic to the baseline model. For example, in

22

a discriminate purge, the value of replacing a proportion κ of existing agents is proportional to $(\bar{r} - \mu_F)\kappa - r_1\frac{\kappa^2}{2}$ and all our results hold.

A declining replacement pool arguably facilitates the occurrence of semi-indiscriminate purges. While it is improbable that all new agents are better than existing subordinates, it is likely that some of them are (formally, $r(1) < \lambda < r(0)$). Recall that in a discriminate purge when all failures are purged, the breadth decreases with violence. For relatively high intensity of violence, the replacement pool is barely touched and the marginal new subordinate is, thus, of high quality. This would then induce the autocrat to start purging from the success pool.

Carrots and sticks

So far, we have assumed that the autocrat can only choose the intensity of violence, the sticks. In practice, she can also vary the material benefits available to insiders, the carrots. In some situations, such as democratic regimes, positive incentives may be the only tools available to the principal. This subsection studies the role of carrots in addition to sticks.

Formally, we return to the case when the replacement pool does not depend on the purge breadth. We further assume that the replacement pool may be of higher quality than existing subordinates (i.e., $r \geq \lambda$). To facilitate comparison with the original set-up (baseline model with $r \geq \lambda$), we suppose that the autocrat can only supplement the second-period material benefit at marginal cost $\xi'(R_2) = \xi_0 + \xi_1 R_2$. The original model then corresponds to the case when R_2 is constrained to zero (equivalently, $\xi_1 \to \infty$). Throughout, to simplify the analysis, we further impose that $\xi_0 = \zeta_0$ and that the highest feasible endogenous reward $\overline{R_2}$ and intensity of violence $\overline{L}$ jointly satisfy $\overline{R_2} + \overline{L} = 1 - v(S,c) - V_2(S,c)$.

In this extension, an agent's effort as a function of the purge incidences κ_F and κ_S is:

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

Essentially, the additional reward R_2 is a perfect substitute for the intensity of violence L. As a result, when the autocrat has the ability to use both carrots and sticks, she can provide the same amount of total inducement at a strictly lower cost. Quite logically, the autocrat uses this lower cost to increase the overall incentives provided to her agents. This generates a few differences in equilibrium outcomes summarized in the next proposition.

Proposition 7. *Suppose the autocrat can also offer an additional reward R_2 in period 2. Compared to the original 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.

Point (i) parallels the analysis of the impact of greater intensity of violence in the original model. As total inducement (punishment and reward combined) increases, the target pool becomes more tainted, yielding a strictly higher marginal benefit of purging and weakly higher purge incidence (strictly if some failures survive or if some successful agents are purged). Point (ii) follows from the non-monotonic relationship between violence and breadth. Depending on the form of the purge, the greater total incentive can be associated with larger purge breadth (e.g., if some failures survive with and without carrots) or lower purge breadth (e.g., if all failures and no successes are purged with and without carrots).

Point (iii) of Proposition 7 is slightly more surprising. 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 original model when no rewards are available. This is a consequence of the indirect effects present in our many-to-one accountability framework. Recall that absent carrots, the benefit of investing in violence is convex when some failures survive (i.e., $L < L^{full}$, as shown in Figure 1). Given that reward and punishment are perfect substitutes, this property carries over to the case where the autocrat can also use carrots. With the lower cost of providing incentives, the equilibrium total inducement is then much larger than the equilibrium intensity of violence in our original model. Indeed, the difference may be such that, even though violence only constitutes a portion of the total incentives, the intensity of violence can be strictly higher with rewards than without.

Mass purges and the autocrat's survival

Our results are also robust to a change in the autocrat's objective. Rather than caring about the 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 second-period non-congruent agents. Formally, the autocrat seeks to maximize

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

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

The rest of the model is similar to the original model. In particular, 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, with $V_2(c) > V_2(nc)$ (e.g., congruent agents receive a higher payoff when the autocrat survives).

Agents then behave as in the original model and the observable features of purges remain the same. For relatively low intensity of violence the purge is discriminate and some failures survive, whereas the purge is semi-indiscriminate for high intensity if r is sufficiently large. The purge breadth still varies non-monotonically with the intensity of violence.

As before, effort increases with the intensity of violence (at a high rate when some failures survive and at a low rate when the purge is semi-indiscriminate), and selection is non-monotonic with L (assuming $r > \lambda$) whenever $\beta(\cdot)$ is not "too concave" (we provide exact conditions in Propositions E.3 and E.4). This last restriction guarantees that any second-order effect due to the shape of the survival function is dominated by the first-order direct and indirect effects of violence that we have identified in the analysis of our original model.

This extension thus suggests that the strategic interactions between one principal and a mass of subordinates, rather than the particular goal of the autocrat (performance or survival), is the key force behind the equilibrium features of purges and the trade-offs we identify.

Discussion

In this last section, we discuss the implications of our findings on the study of mass purges, violence in autocracies, and many-to-one accountability problems more generally.

Rational purges?

Our approach presupposes that after deciding to launch a mass purge, the autocrat strategically chooses the proportion of failures and successes to be replaced. Doing so, our theory yields some distinctive patterns. First, the form of the purge is linked to the intensity of violence: Discriminate purges with some failures surviving tend to be relatively mild, while semi-indiscriminate purges

with some successes being purged tend to be brutal. Second, whenever the autocrat's information is coarse (but not necessarily binary as in our original model; see Appendix F.1), the purge breadth is non-monotonic in the intensity of violence: Violent purges do not necessarily translate into a greater proportion of subordinates being replaced.

These findings can help distinguish our approach from others. For example, if mass purges are simply random, we should not expect the form 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 the purge breadth and the intensity of violence is likely to be always positive. Using the comparative statics described above, our theory can thus be falsified.

To illustrate some of the patterns uncovered in our paper, we turn to historical evidence. In his comparison of Chinese and Soviet purges, Teiwes (1993) remarks that Chinese rectification campaigns were characterized by a low intensity of violence and a high level of predictability, whereas in the USSR in the thirties, purges were violent and their targets were 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 between violence and targets of the purge extends beyond China and the USSR. In Fascist Italy, Giovanni Giuriati—in charge of the purge of the party—"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).

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 their intensity of violence and in their form, Soviet and Chinese purges had on average a similar breadth, in line with our results. During the 1930s Stalinist purges, the proportion of purged members varied from 5% in 1930, 1931, and 1937 to 22% in 1933-34. The expulsion rate in Chinese rectification campaigns fluctuated between 9% in 1957-58 and 23% in 1947-48.[4]

Pushing the model a bit further, our framework can provide some insights for the ongoing mass purge in Turkey (the so-called 'Erdogan purge'). While the first victims of the purge were clearly

---

[4]Given the differences in party size and population, the proportion of purged party members more closely fits our model. The number of members expelled from the party was always much larger in China.

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 face expulsion or arrest), our theory predicts that the Erdogan purge is likely to have predictable targets, but it will also affect thousands of individuals.

Observe that the discussion here only addresses the rationality of the autocrat's purging choices conditional on a mass purge occurring. It does not deal with the broader question of the rationality of mass purges vis-a-vis other instruments of repression. Since our paper focuses on the features of mass purges, for reason of space, we can only offer some conjectures on the reasons for their use. Our analysis reveals that mass purges significantly increase effort but moderately improve the congruence of subordinates, especially if violent. Hence, we should expect mass purges to occur in contexts when the autocrat seeks to increase the work rate. For example, Levytsky (1972) argues that the objective of terror under Stalin was to raise productivity of bureaucrats and party members, meant to blindly obey and execute orders, and, more generally, "to squeeze the last drops of effort out of the 'free' workers" (p.316).

Mass purges further seem to have occurred during, arguably, the most personalist phase, to borrowGeddes's (2003) terminology, of the communist regimes in the USSR and China.[5] We see two possible complementary reasons for this. According to Geddes (2003), personalist leaders control appointments, potentially raising the congruence of new agents, and the security apparatus, potentially reducing the cost of carrying out the purge. Purges may then have almost disappeared in China and the USSR following the deaths of Stalin and Mao because of the subsequent return to a form of collective leadership to avoid a repeat of past excesses (Levytsky, 1972; Teiwes, 2017).

Obviously, much more needs to be learned about why autocrats decide to start a mass purge. However, our framework can be seen as a possible starting point for a more general theory of coercive instruments in autocracy.

The ever-present possibility of violence

Autocrats face few constitutional constraints on their actions (notwithstanding international pressures). As such, autocrats have a wide range of tools at their disposal, especially violence. Our framework proposes one possible rationale for the observed violence in mass purges. The autocrat

---

[5]Note that Geddes (2003) characterizes these two countries as single-party autocracies, not personalist autocracies (pages 227-232).

commits, at a cost, to a certain intensity of violence in order to obtain the right level of effort from agents and to select the best possible subordinates.

We show that even though an autocrat faces no de jure checks (they can credibly optimize the intensity of violence), subordinates' strategic behavior de facto constrains the autocrat's decision. As violence increases, so does the fear of being purged; agents work more whether they are aligned with the autocrat. But this general increase in effort has a clear downside: An agent's accomplishment is less informative about his ideological congruence with the autocrat whenever all failures are purged. 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 engendered by higher violence is not always paralleled with an improvement in term of selection (greater violence always improves selection only if the proportion of congruent types in the pool of replacement is low—$r < \lambda$—; that is, when selection is impeded to begin with).

This trade-off between effort and selection is especially strong when purges turn violent and semi-indiscriminate. It does not seem to have remained unnoticed by 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.'"

Yet, despite this trade-off, Stalin chose to engage in violent semi-indiscriminate purges, at least according to Teiwes (1993). One can wonder whether the relatively stringent conditions for semi-indiscriminate purges described in Remark 3 were met in the USSR in the 1930s. While we do not have definitive evidence, it should be noted that the USSR experienced both a significant number of purges (see Table 1) and investment in security services (Levytsky, 1972) in the 1920s, potentially gradually building up the necessary infrastructure of violence. Additionally, 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 the Soviet regime and to Stalin (Fitzpatrick, 1979; Wolton, 2015).

Two other important historical events can be cited as examples of violent and semi-indiscriminate purges: the Great Terror (1936-38) and the Cultural Revolution (1962-76). In both cases, the leader of the regime aimed to increase his hold on power, making these two events best understood as part

of our extension on the autocrat's survival. In both cases, the autocrat targeted subordinates from the communist party, the army, and the bureaucracy (Conquest, 2008). In both cases, the general population also fell victim to a violent repression (in Appendix F.3, we show how our framework can be adapted to study repression; there are, however, different approaches, e.g. Tyson, 2017).[6] The particular characteristics of the Great Terror and the Cultural Revolution, our theory contends, can be explained by a low cost of violence. Stalin drastically simplified procedures (troikas, confessions sufficient for conviction) 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). Notice that our framework also allows us to reassess some claims about the objective of the Great Terror. Due to its high intensity of violence, the Great Terror may have been helpful to deal with a lack of performance in all sectors of the Soviet economy (as argued by 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 or Kaganovich to defend Stalin's actions).

The cases of the Great Terror, Cultural Revolution, or the Erdogan purge, however, highlight one limitation of our paper. Violence often targets the elite of the regime as well as the mass of rank-and-file. Our paper has little to say about elite purges, which are the object of a large literature that, in turn, ignores mass purges. Future research would do well to explore the connections between elite and mass purges, possibly building on the theoretical framework developed in this paper.

Many-to-one accountability

A key feature of our theory is that many subordinates are accountable to a single principal, the autocrat. This contrasts with most previous political agency models in which one agent is accountable to one principal (e.g., a prime minister) or many principals (e.g., voters). In this subsection, we discuss some particularities of many-to-one accountability.

In our set-up, all agents work on an independent project. However, they are not evaluated in isolation. The autocrat makes her purging decision based on relative performance—in other words, on the informativeness of success and failure. This interdependency between agents generates important indirect effects. An increase in the intensity of violence does not just affect effort

---

[6]Some parts of the USSR also experienced harsher purges and repression during 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).

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). Due to these indirect effects, small changes in economic or political conditions can have large consequences for equilibrium values such as the intensity of violence, the purge breadth, or even the form of the purge (Remark 1). Thus, our model provides one possible rationale for the significant variation in breadths from one purge to the next in the USSR and China (as documented in Tables 1 and 2). While the violence and breadth of the purge may be carefully planned by an autocrat with superior knowledge of the situation she faces, these features may be difficult to predict for external observers and, in fact, may even appear random.

Another particularity of many-to-one accountability regards the trade-off between effort and selection we discuss at length above. We are not the first to formalize it. The literature on accountability in democracies has highlighted a similar trade-off when it comes to providing access to better information to the electorate (e.g. Ashworth et al., 2017) or offering better pay to office-holders (e.g. Besley, 2004). There is, however, a critical difference. When it comes to purging a mass of subordinates, this trade-off arises only above a certain intensity of violence. Below this threshold, effort *and* selection are positively correlated with the intensity of violence.

Many-to-one accountability thus exhibits non-monotonic relationships that appear absent from other accountability settings. In general, many-to-one accountability cannot be adequately approximated by a one-to-one setting (in Online Appendix F.4, we highlight how the results of our set-up differ from those of a single-agent setting). Many-to-one accountability is not limited to autocracy. In democratic settings, constitutional restraints limit the use of violence (at most, membership can be terminated), but a leader can use rewards, a perfect substitute for violence as one extension shows, to motivate members. Outside of politics, in large firms or in the army, the principal shares the same problem as the autocrat, with different tools at her disposal (mass layoffs or up-or-out promotions instead of purges). We, thus, believe the approach developed in this paper has a wide range of applications in 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.

Ashworth, Scott, "Electoral accountability: recent theoretical and empirical work," *Annual Review of Political Science*, 2012, *15*, 183–201.

__ , **Ethan Bueno de Mesquita, and Amanda Friedenberg**, "Accountability and information in elections," *American Economic Journal: Microeconomics*, 2017, *9* (2), 95–138.

Besley, Timothy, "Paying politicians: theory and evidence," *Journal of the European Economic Association*, 2004, *2* (2-3), 193–215.

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.

Brzezinski, Zbigniew K, *The permanent purge*, Harvard University Press Cambridge, MA, 1956.

Bueno de Mesquita, Bruce and Alastair Smith, "Political Succession: A Model of Coups, Revolution,Purges and Everyday Politics," *Journal of Conflict Resolution*, 2015, *61* (4), 707 – 743.

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.

Dal Bó, Ernesto, 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.

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.

Dewan, Torun and David P Myatt, "The declining talent pool of government," *American Journal of Political Science*, 2010, *54* (2), 267–286.

_ **and Rafael Hortala-Vallve**, "The three As of government formation: Appointment, allocation, and assignment," *American Journal of Political Science*, 2011, *55* (3), 610–627.

Dikötter, Frank, *The Cultural Revolution: A People's History, 19621976*, Bloomsbury Publishing, 2016.

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.

Esteban, Joan, Massimo Morelli, and Dominic Rohner, "Strategic mass killings," *Journal of Political Economy*, 2015, *123* (5), 1087–1132.

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.

Geddes, Barbara, *Paradigms and sand castles: Theory building and research design in comparative politics*, University of Michigan Press, 2003.

Getty, John Arch, *Origins of the great purges: the Soviet Communist Party reconsidered, 1933-1938*, Vol. 43, Cambridge University Press, 1987.

Green, Jerry R and Nancy L Stokey, "A comparison of tournaments and contracts," *Journal of Political Economy*, 1983, *91* (3), 349–364.

Gregory, Paul R, *Terror by Quota: State Security from Lenin to Stalin:(an Archival Study)*, Yale University Press, 2009.

Hohne, Heinz, *The Order of the Death's Head: The Story of Hitlers' SS*, Pan Books, 1981.

Holmström, Bengt, "Moral hazard in teams," *The Bell Journal of Economics*, 1982, pp. 324–340.

Janda, Kenneth, *Political parties: A cross-national survey*, New York: Free Press; London: Collier Macmillan, 1980.

Jiang, Junyan, B. Pablo Montagnes, and Stephane Wolton, "Political Tournaments: Accountability and Selection," 2017.

Laffont, Jean-Jacques and David Martimort, *The theory of incentives: the principal-agent model*, Princeton university press, 2009.

Landa, Dimitri and Scott A Tyson, "Coercive Leadership," *American Journal of Political Science*, 2017, *61* (3), 559–574.

Lazear, Edward P and Sherwin Rosen, "Rank-order tournaments as optimum labor contracts," *Journal of political Economy*, 1981, *89* (5), 841–864.

Levytsky, Boris, *The uses of terror: the Soviet secret police 1917-1970*, Coward, McCann & Geoghegan, 1972.

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.

Schapiro, Leonard Bertram, *The government and politics of the Soviet Union*, Taylor & Francis, 1977.

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.

____ , *Leadership, legitimacy, and conflict in China: From a charismatic Mao to the politics of succession*, Macmillan, 2017.

Turkey Purge Website, `https://turkeypurge.com/` 2017.

Tyson, Scott A., "The Agency Problem Underlying Repression," *Journal of Politics*, 2017, *forthcoming.*

Wolton, Thierry, *Une histoire mondiale du communisme, tome 1: Les bourreaux*, Grasset, 2015.

____ , *Une histoire mondiale du communisme, tome 3: Les complices*, Grasset, 2017.

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

Table 1: Proportion of party members purged in USSR

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)[+]

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

Table 2: Proportion of party members purged in China

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)

All proportions are approximation. [+] denotes authors' calculation