

# Who to Target? Strategic Elite Elimination in the Contest for Power

Allison Cuttner  
Department of Political Science  
University of Toronto  
[allison.cuttner@utoronto.ca](mailto:allison.cuttner@utoronto.ca)

September 16, 2022

## Abstract

When a dictator moves to consolidate his own power vis-a-vis the elite allies in his inner circle, who should he target for a purge? By introducing power heterogeneity into the elites of the ruling coalition, I formalize a dictator's attempt to seize power from regime insiders with a dynamic and novel adaptation of a classical flexible contest framework. In general, strong dictators will target strong elites who yield the most rewards in a successful purge. This drive is mitigated by uncertainty as a relatively strong dictator will maintain his status quo power instead rather than initiate a risky conflict. The conclusions I derive show that elite coalitions can lead to more purge attempts instead of deterring them and that both powerful and relatively weak dictators will initiate conflict with their inner circle of elites, an important implication for empirical studies of purges and the personalization of power in dictatorships.

Keywords: Authoritarian Politics, Purges, Consolidation, Personalization, Coalitions, Conflict

During his reign at the top of the Soviet regime, Stalin conducted not only mass purges of rank and file party members, bureaucrats, and military officials (Getty, 1987) but removals and eliminations of elite Politburo, Central Committee, and high-ranking CPSU members. The expulsions, arrests, and executions of prominent Bolsheviks like Kamenev and Zinoviev, heads of government like Rykov and Syrtsov, and previous Stalin allies like Bukharin and Rudzutak, are frequent examples of Stalin's growing personalist power through the 1920s-30s (Khlevniuk, 2008). These powerful men, however, were not the only elite Soviet officials that Stalin systematically removed during his reign. Candidate members of the Politburo (Postyshev, Eikhe, and Chubar), Central Committee members (Yenukidze and Pyatakov), and regional party and NKVD (secret police) bosses (Kosior and Uspensky) were similarly purged from the party and their positions, eventually being executed. Clearly Stalin did not only attack his most powerful and influential colleagues, nor did he only target weaker, less-established elites, but a variety of officials from different ranks and institutions, different ideological backgrounds, and different geographic areas over his period of personalist consolidation.

Such elites that form the ruling coalition at the top of an authoritarian regime are distinct from one another in a variety of ways. The formal positions of power they hold as party, cabinet, junta, or politburo members entail different levels of policy influence and access to resources (Arriola, DeVaro and Meng, 2017; Meng, 2019). Influence and patronage could be tied to an elite's ethnic, clan, or religious identity (Coughlin, 2005; Hornsby, 2013; Van Dam, 1996), networks of connections with other elites through education or previous positions (Lu and Lorentzen, 2016), or military command (Barros, 2002). One way that an autocratic leader can solidify his position as an unconstrained or personalist dictator is through purges—forcible removals—of these various elites that make up his ruling coalition (Geddes, Wright and Frantz, 2018; Goldring, 2020; Svoboda, 2009). While the removal of elites from the upper echelons of power could occur for a variety of reasons such as ideological disagreement, elite malfeasance, or a demonstration of personnel power (Lu and Lorentzen, 2016; Montagnes

and Wolton, 2019; Woldense, 2018), the focus of this study is consolidation purges in which the dictator removes or eliminates members of his inner circle in order to take more power for himself (Svolik, 2009; Geddes, Wright and Frantz, 2018). When faced with a heterogeneous ruling coalition of elites, which types of elites will be purged in the dictator's movement toward consolidation? In other words, who will a power-hungry dictator target?

The makeup of the dictators ruling coalition, in particular who remains to govern after a purge, has downstream implications for both the inner circle and the regime at large. The members of the ruling coalition that continue in power after a consolidation attempt may be more or less ambitious and threatening to the dictator (Kosterina, 2017) or have an unstable balance of power with the leader (Acemoglu, Egorov and Sonin, 2008; Svolik, 2009). For bureaucrats, lower-level security officials, and citizens, the identity of the members of the inner circle have consequences for personnel (Carter and Hassan, 2021; Hassan, 2017; Woldense, 2018) as well as the distribution of public goods and services (Kramon and Posner, 2016). Furthermore, with ever increasing numbers of dictators consolidating power at the expense of their elite partner (Geddes, Wright and Frantz, 2014), understanding the process by which this unfolds generates important insights into regimes that are frequently closed, secretive, and the least likely to democratize (Barros, 2016; Geddes, Wright and Frantz, 2014).

A ruling coalition with which to share power and govern effectively is necessary to help a regime come to power (Bueno De Mesquita et al., 2003; Haber, 2006), but is the greatest threat to a dictator's tenure (Svolik, 2012). Violating an agreement to share power with his elites by attempting to remove one of them in a consolidation attempt destabilizes a regime and will be met with resistance from his coalition (Svolik, 2009). Thus when choosing the path of personalization and consolidation through elite eliminations, the dictator must consider the tradeoff between his desire for power and the instability that purges introduce (Luo and Rozenas, 2019). I argue that the various potential targets of his purge create a similar tradeoff: a more powerful, influential member of the ruling coalition would yield the

dictator a high reward when removed, but attempting to purge them is risky. Violating the power-sharing agreement could cause elites to retaliate in a coup, the success of which is subject to uncertainty (Meng, 2019; Svolik, 2009; Singh, 2014).

I model a dictator facing two elite actors with differing amounts of power in a dynamic setting. In the first round, the dictator must decide whether to initiate a purge and which elite, who vary by the amounts of power they possess, to target. If an elite avoids being targeted, he, in turn, must decide whether to form a defensive coalition against the dictator or join the dictator against his elite colleague. Conflicts between the dictator and his allies are subject to contest-style uncertainty: while having a power advantage over your opponent makes you much more likely to defeat him, your success is not guaranteed. The most powerful winner of the conflict in the first round will usurp the power of the defeated side, and anyone left standing will move to a second round in which they must make similar conflict decisions. This builds on Acemoglu, Egorov and Sonin (2008) non-democratic coalition model in which the relative power of coalition members determines who is removed from the coalition by incorporating that relative power and stochastic noise into an uncertain conflict over that removal. How the dictator resolves the risk-reward tradeoff of targeting a member of his ruling coalition for a purge yields different expectations about who he will target (or if he moves to consolidate at all) depending on both his relative power as well as the uncertainty of the ensuing conflict. The results show how the dictator's preferred target changes with his own power advantage, the difference between the possible elite targets, and the uncertainty of the conflict. When the dictator is weak relative to the elites and the difference between the elites is low, the dictator targets the weaker elite. As the difference between the elites increases, the weaker elite is a less lucrative target and the dictator targets the stronger elite. If the dictator is weak relative to the uncertainty of the potential conflict, he will not initiate conflict and instead maintain his status quo power instead of initiating a risky conflict.

In addition to the dictator's targeting decision, this model of dictatorial consolidation generates predictions about elite coalition behavior in response to a purge attempt. I build

on similar models of dictator-elite conflict in which the elites overthrow the dictator to punish him for violating the power sharing agreement (Boix and Svobik, 2013; Meng, 2019; Svobik, 2009) but do not assume that the elites, who I model as individual actors, will always coalesce against the dictator. Instead, the elites makes a similar risk-reward calculation in weighing the likelihood that they can successfully overthrow the dictator with the benefits that ousting him will yield. I show the conditions under which the formation of an elite coalition leads to more consolidation conflict than there would otherwise be, not deterring the dictator from violating the power-sharing status quo, because defeating multiple elites at once offers the dictator even greater power than eliminating one. The non-monotonic relationship between the dictator’s power relative to his elites and his moves to consolidate power through elite purges has implications for the burgeoning empirical literature on personalism, consolidation, and purges in authoritarian regimes (Geddes, Wright and Frantz, 2018; Goldring, 2020; Keller and Wang, N.d.; Sudduth, 2017).

## **Power Consolidation and Conflict**

Rulers come to power with the support of allies, often formally grouped into a party, council, royal family, or military junta (Gandhi, 2008; Geddes, Wright and Frantz, 2018; Luo and Rozenas, 2019). While some agreement to share power among the members of the ruling coalition may have been agreed upon (Bueno De Mesquita et al., 2003) or even formalized through institutions (Boix and Svobik, 2013; Myerson, 2008), “institutions do not eliminate the ruler’s primal instinct to accumulate power” (Luo and Rozenas, 2019, p. 1). Consider an authoritarian regime that has already formed: there is a leader, or dictator, and some set of elite regime members with various positions, holdings, and factional ties. While the dictator individually has more power than any single elite (which is why he is the dictator), this advantage may be very small. One of the methods by which the dictator can accumulate more power for himself is by breaking the status quo power-sharing agreement and eliminate

the members of his own coalition, removing, expelling, imprisoning, exiling, or executing them and thereby taking their power for himself (Geddes, Wright and Frantz, 2018; Svolik, 2009).

## **The Rewards of Elite Purges**

Removing high-ranking members of the ruling coalition could offer the dictator a variety of benefits. While some purges could be conducted to remove corrupt or truly treacherous officials (Lu and Lorentzen, 2016) or as a public display of arbitrary power over personnel (Woldense, 2018), many elite eliminations are used to consolidate power directly from the elites to the dictator (Gandhi and Sumner, 2020; Geddes, Wright and Frantz, 2018; Svolik, 2009). An elite in the dictator's inner circle often holds formal positions in the government, military, and/or party and, with the executive powers over personnel and appointment that are *de jure* or *de facto* held by the leader, the dictator has control over what happens to those offices. Removing the elite can open up the position for a protégé, family member, or loyalist, increasing the dictator's influence and inspiring loyalty among the promoted. For example, in the Soviet Union, by eliminating Beria and other close associates of Stalin after the latter's death, Nikita Khrushchev was able to promote his protégés Kirichenko, Brezhnev, Zhukov, and Furtseva to positions of power; they were then able to support Khrushchev against a coup attempt by other members of the elite in June of 1957 (Taubman, 2003).

Alternatively, the dictator can leave prominent positions open, maintaining the post-purge balance of power in the ruling coalition with *de facto* control over the domains of the open offices. Félix Houphouët-Boigny, founding president of Côte d'Ivoire, maintained vacancies in his cabinet including such prominent offices as the minister of defense and the vice presidency (Meng, 2019). A dictator could even keep such titles and offices for himself, taking full control over the material resources, policy control, and prestige that the position incurs. In Mali, Moussa Traoré took on the role of general secretary of the Democratic Union of the Malian People as well as minister of defense and security in addition to his position

as president. Saddam Hussein accumulated positions as the chairman of the Revolutionary Command Council, regional secretary of the Ba'ath Party, general secretary of the National Command of the Arab Socialist Ba'ath Party, and eventually Prime Minister all in addition to being president of Iraq. Collecting governmental offices, either by installing lackeys or directly controlling them himself, puts greater control over policy, personnel, and resources in the hands of the dictator and has been used as an empirical measure of the consolidation of power (Gandhi and Sumner, 2020).

How much the dictator benefits from an elite elimination depends on **who** is being eliminated: their position in the government, party, or military, the resources that once flowed to them, etc. A higher level position such as a head of party or government or minister of a major department like defense, finance, or interior, will yield the dictator much more control when he installs his loyal protégé (or himself) into that office. Targeting such a powerful member of the ruling coalition for removal, in violation of an agreement to share the power of the regime, is not without risk of potentially grave consequences.

## **The Risks of Elite Purges**

The uncertainty inherent in a conflict between dictator and elites where institutions might exist but don't constrain the way they do in democracies (Gandhi, 2008; Meng, 2019; Svolik, 2012) and violence underlies every contest (Svolik, 2012) has been widely recognized, but mostly from the perspective of elites. Elites may be uncertain about the dictator's relative strength (Meng, 2019), hidden actions that the dictator could be taking (Boix and Svolik, 2013), what signals other elites receive (Luo and Rozenas, 2019), or simply whether a coup could succeed (Little, 2017; Singh, 2014). The dictator, however, can be equally uncertain about whether his attempt at a purge will succeed. Even with full information about the actions and intentions of the members of the inner circle, an attempt at an elite removal could fail.



In Malaysia, for example, the dismissal and imprisonment of Deputy Prime Minister Anwar Ibrahim sparked the *Reformasi* movement of mass demonstrations against the sitting government. After several regime party politicians lost seats to opposition candidates in 1999, the Prime Minister responsible for initially ousting Anwar, Mahathir Mohamad, resigned in 2003.<sup>1</sup> In the Soviet Union, the formal removal of elites from their official positions required a vote from the Politburo, the members of which were not always beholden to the leader. Early in Stalin's tenure,

members of the Politburo conducted themselves independently... forming diverse and unexpected (given subsequent events) tactical coalitions... The votes were evenly divided on whether or not Trotsky and Zinoviev should be immediately expelled from the Central Committee... On 20 June 1927, a bare majority voted to expel Trotsky and Zinoviev, but only after Stalin demanded that his vote be counted in absentia and Kalinin joined those in favor of immediate expulsion (Khlevniuk, 2008, p. 3)

Even a dictator like Stalin, who is often considered a pinnacle of consolidated power and highly effective at purging his elite comrades (as well as mass party and military members), was “forced...to act cautiously...and keep an eye on the mood of his comrades-in-arms” (Khlevniuk, 2008, p. 5).

Much in the way that coups are more likely to succeed if the target of the coup is relatively weak and the coupers are numerous and powerful (Little, 2017; Singh, 2014), I argue that the dictator is more likely to succeed in his attempt to purge a member of the ruling coalition when he (and any other elites who join him in a coalition) is more powerful than his opponent. Pavel Postyshev, a relatively young candidate member of the Politburo, was easier to remove than the chairman of the Council of People's Commissars (and thus the head of government), Aleksei Rykov, whose replacement “dragged on for sometime... which leaves room for speculation that Stalin was wavering, weighing the advantages and

---

<sup>1</sup>Mahathir resumed power under a new party label in 2018.

disadvantages before making up his mind” (Khlevniuk, 2008, p. 34). Strong targets can be strong opponents, as Burkina Faso’s president Jean-Baptiste Ouedraogo found when he attempted to remove Thomas Sankara from his post as prime minister and was immediately overthrown in a coup. In addition to their relative power, the conflict between a dictator and his purge target is subject to the “fundamental and irreducible uncertainty” of conflict that results from stochastic processes (Bas and Schub, 2017). The critical substantive feature is that some features of conflict between the dictator and powerful elites, whether from a reactionary mass protest, elite coup, or lack of control over personnel who carry out arrests or violence, makes the outcome of the conflict uncertain to both the dictator and his elite rivals. Thus while eliminating a member of the ruling coalition with great *de jure* and *de facto* power can offer the dictator great rewards if he succeeds in his purge, the resistance of this powerful target to removal is an important consideration. The greater the target, the greater risk the dictator will fail to remove him and instead face ouster himself.

## **The Path to Personalist Power**

When he seeks more power for himself at the expense of his elite allies, who will the dictator target for a purge? How he resolves the tradeoff between the greater power that he could achieve from a stronger, more influential target and the greater uncertainty of defeating such a target determines the dictator’s targeting decision. When the potential elite targets are very similar to one another, the dictator will target a weaker opponent because the greater likelihood of successfully purging him is sufficient to outweigh the lower power he can take. In general, a stronger dictator will attack a more powerful opponent, whether that is the elite with higher power or a coalition of elites who coalesce against the dictator and can all be removed. This relationship between the power advantage of the dictator and the power of his target, however, is conditioned by the uncertainty of the resulting conflict between them. In a highly uncertain environment where the dictator is *ex ante* unsure whether his purge attempt will give him the power of the eliminated elite or result in his own ouster, even an

extremely powerful dictator would rather avoid a risky conflict and maintain the status quo arrangement with his elites, resisting the temptation to attempt any consolidation at the elites' expense.

How will the elites react to their leader's attempt at removing them? While I assume that the purge target will immediately launch a coup against the power-hungry dictator, I do not assume other elites will necessarily join the effort. In an elite-dictator conflict, other members of the ruling coalition have the option to join their peer against the dictator, join the dictator in ousting the elite, or stay on the sidelines of the conflict. By allowing non-targeted elites to join the conflict only when it is in their best interest to do so and not because of any previous commitment to their colleagues, I am able to further explore how the risk-reward tradeoff of conflict in the upper echelons of power affects coalition behavior from the elite's perspective. Not only will the elites not always coalesce against the dictator's consolidation attempt as some previous models assume, elite coalitions will lead to more conflict as defeating a coalition offers greater rewards despite the higher risk the dictator faces. Further, the elites' ability to form coalitions has heterogeneous effects on the elites themselves: the lower-powered elite is worse off as the higher-powered elite joins him in a coalition and alters the dictator's targeting decision. It is the anticipation of the elite's coalition behavior and the risk-reward tradeoff that conflict at the top of the regime engenders that determines when the dictator consolidates power vis-a-vis elites and who he targets in the purge.

## **Modeling Elite-Dictator Conflict**

Dictators desire to simultaneously amass more power for themselves and avoid the instability that elite-dictator conflict may bring to the regime. Elite elimination is an effective way to achieve personalist power, but there are potentially drastic consequences: a counter-coup or mass elite defection could lead to the dictator's ouster. My theory of elite-dictator conflict

has two (related) stages: targeting and coalition formation. Considering the potential for his ouster or the possibility of making the regime more vulnerable to outside opposition, the dictator must first consider whether initiating any intra-regime conflict is worth the consequences. If he deems conflict beneficial, in expectation, for his personal power, the dictator must then decide who to target for elimination.

These modeling assumptions yield a few substantive scope conditions on the applicability of this model to the inner workings of autocratic regimes. First, the dictator, as the conflict “agenda-setter,” must be able to target an elite for a purge. This could include introducing articles of expulsion to a legislature, ordering police to make an arrest, or prompting a paramilitary to attempt assassination. This does not require that the dictator must be such a powerful strongman that he can unilaterally remove an elite from their position in the inner circle and usurp their power. Indeed, by incorporating uncertainty over the outcome of the conflict, even a strongly advantaged dictator in my model will not be certain that his purge attempt will succeed. Second, in order to focus on the dictator’s targeting of his allies, preemptive coups are outside the scope of the model. The elites only respond in a coup when the dictator violates the power-sharing agreement by initiating a purge (Meng, 2019; Svobik, 2009). Relatedly, there is no first-mover advantage: when the dictator and elites are in conflict, only their relative power and the conflict uncertainty determine the outcome, not an element of surprise which would make the dictator’s initiation relative to a possible elite-initiated coup more relevant.<sup>2</sup> Lastly, I assume that there is no credible sharing of the spoils of conflict. In practice, this means that all of the benefits of removing a member of the inner circle—whether it is an elite or the dictator himself who loses—flows to the most powerful member of the coalition that removed him.<sup>3</sup>

I formalize the heterogeneous levels of power of the dictator and two elite regime members, the uncertainty of the interested parties over the conflict outcomes, and the targeting

---

<sup>2</sup>Without a first-mover advantage, in the parameter space where the dictator does not initiate conflict but the elites may want to coup, in equilibrium the dictator would simply be indifferent between initiating conflict and allowing the elites to initiate it.

<sup>3</sup>This assumption is relaxed in Cuttner (2021) in a focus on sharing power with rivals.

mechanism below. The variation in targeting and elite coalition formation depends not on coordination-promoting institutions or non-credible commitments, but the self-interest of heterogeneously powerful members of the ruling coalition.

## Model Setup

Three players, a dictator (D), an elite with a high initial endowment of power (H), and an elite with a lower initial endowment of power (L), together form a regime. Each player's type,  $\tau_i > 0$ , is his power endowment. I use the following notation to measure the power disparity among the players: H's initial endowment of power is fixed at  $H$  ( $\tau_H \equiv H > 0$ ); the dictator's endowment is  $d > 0$  larger than this ( $\tau_D \equiv H + d$ ) and L's endowment is  $w \in (0, H)$  less ( $\tau_L \equiv H - w$ ). The dictator has the most power initially (which is why he is the dictator) and the high elite has more initial power than the low elite. Thus  $d$  (and  $d + w$ ) can be interpreted as the dictator's initial power advantage over the other two elites while  $w$  denotes the relative difference between elites H and L.

Play takes place over the course of two rounds. First, the dictator chooses which elite to target for elimination or does not initiate conflict, ending the round. If he chose a target, the non-targeted elite can choose to join a coalition with the dictator, the targeted elite, or remain out of the conflict. The dictator and target(s) then participate in a contest where the probability that each side wins is the difference between their relative power plus mean-zero noise; e.g. participant (or coalition)  $i$  wins the conflict if  $\tau_i \geq \tau_j + \epsilon_t$  where  $\epsilon_t \sim U[-a, a]$  independent of the round.<sup>4</sup> Thus the probability that participant  $i$  wins is  $F_\epsilon(\tau_i - \tau_j)$ . While the amount of uncertainty the potential conflict parties are subject to can vary, I assume all conflicts are uncertain; in particular, I assume that  $a$  is sufficiently large such that no individual or coalition can win a conflict with certainty.<sup>5</sup>

---

<sup>4</sup>This distribution is symmetric around 0 and the likelihood of a draw of noise is equal between the bounds. The parameter of interest regarding the uncertainty of the conflict is  $a$ : as  $a$  increases, the spread of the noise distribution increases in both directions (maintaining symmetry around 0). The more advantaged party is less certain of their victory as  $a$  increases.

<sup>5</sup>Technically,  $a > H + d + w$ .

The power of the loser(s) of the conflict is transferred to the most powerful winner. Note that the winner cannot split the winnings with a coalition member because any *ex ante* commitment to split gains upon winning would be violated when the time to share power came. While both exogenous and endogenous sharing rules are commonplace in contests that are nested or allow for alliances (Konrad, 2009), such agreements may have credibility problems in authoritarian regimes (Gehlbach, Sonin and Svulik, 2016; Svulik, 2012). Thus I focus on the baseline case of complete lack of enforceability. This assumption is relaxed in Cuttner (2021), where the dictator and elite are able to share the spoils of coalescing against another elite.<sup>6</sup>

The first round ends and whichever remaining player has the most power now becomes the dictator and can choose to initiate conflict with the other remaining member of the regime (if there is a second member). If there is only one player remaining in round two, the game ends. If conflict is initiated, it occurs as previously described, with a new, independent draw of  $\epsilon$ . The power of the loser(s) is transferred to the winner and the game ends.

## Sequence of Play

$t = 1$

- The dictator, D chooses to target H, to target L, or not to target anyone. If no target is selected, the round ends.

- The non-targeted elite chooses whether to participate in the conflict on the side of the dictator, the target, or not participate.

- The dictator (or coalition) and target(s) participate in a contest where participant(s)  $i$  win if  $\tau_i \geq \tau_j + \epsilon_1$  where  $\epsilon_t \sim U[-a, a]$

- The loser(s)'s power is transferred to the strongest winner; whichever remaining player now has the most power is now the dictator. If only one player remains, the game ends.

---

<sup>6</sup>The general relationships between the dictator's power advantage, conflict uncertainty, and conflict initiation remain when sharing is an option. When conflict does occur, however, the dictator will only utilize the sequential targeting strategy instead of allowing an elite coalition to form and targeting them both.

$t = 2$

- The dictator chooses to target one of the remaining players, or not to target anyone.

If no target is selected, the game ends.

- If there is a non-targeted elite, he chooses whether to participate in the conflict on the side of the dictator, the target, or not participate.

- The dictator (or coalition) and target(s) participate in a contest where participant(s)  $i$  win if  $\tau_i \geq \tau_j + \epsilon_2$  where  $\epsilon_t \sim U[-a, a]$

- The loser's power is transferred to the winner and the game ends.

## Payoffs

All members of the ruling group derive utility from their endowments of power at the end of the game. These *ex post* power endowments are a function of the dictator's target choice and the coalition choices of both elites in both rounds. The actions of each player are, in turn, a function of their initial power endowments and the conflict uncertainty.

$$u_i(H, d, w, a) = \tau_{i,t=2}$$

## Targeting without Coalitions

First consider a benchmark in which coalitions are not possible: the dictator simply makes his targeting decision (H, L, or no conflict) and conflict ensues accordingly, subject to the uncertainty described above. The contest function that determines each actor's probability of winning the conflict creates a tradeoff for this dictator: he is less likely to be successful in a conflict against the higher-powered elite, H, but beating him would yield higher rewards. While he is never certain of victory, a successful purge is much more likely against his lower-powered opponent, L, though the rewards are less. The parameter  $w$  defines the difference between H and L: as  $w$  approaches 0, H and L are very similar both in the rewards from

ousting them and the difficulty in defeating them; as  $w$  approaches its upper bound, L is very weak and more likely to be defeated, but the rewards from the conflict are also less enticing. While all possible conflicts are subject to uncertainty<sup>7</sup>, how much “noise” there is in the dictator’s chance of winning depends on  $a$ , the dispersion of the distribution. When  $a$  is low, the dictator’s power advantage strongly affects his chances of winning; as  $a$  increases, the conflict is more uncertain even when there is a large difference in power. As  $a$  approaches infinity (in the limit), the likelihood that the dictator wins a conflict goes to  $\frac{1}{2}$ , even if he is extremely advantaged over his opponent.

How these concerns affect the dictator’s targeting decision is easily illustrated with a single round of conflict. Consider a one-shot conflict<sup>8</sup> in which the dictator makes a single targeting decision: he can fight H, L, or not initiate any conflict. The characterization of his optimal targeting strategy is summarized in Lemma 1 and visualized in Figure 1. When the uncertainty of the conflict ( $a$ ) is relatively low, the dictator will target L, the low powered elite, when L is not too weak. The benefit for defeating L is sufficiently high (with low  $w$ ) that the dictator will take the higher conflict win probability and attack his weakest opponent. When L is particularly weak, however, as can be seen in the right region of Figure 1(a), the benefit from defeating L is insufficient. Instead, the dictator will take on the greater risk of fighting his high powered opponent for the greater reward.

**Lemma 1.** *When the elites cannot form coalitions and there is a single round of conflict, the dictator will target L for  $a \in (H + d + w, 2H - w)$ , target H for  $a \in (2H - w, 2H + d)$ , and not initiate conflict for  $a > 2H + d$*

---

<sup>7</sup>Note that I assume the spread of the noise distribution ( $a$ ) is sufficiently large to avoid deterministic conflict outcomes (assuming  $\epsilon$  is distributed uniform). Using the uniform supports the tractability, but a variety of probability distributions would maintain the tradeoff between conflict uncertainty and the gains of a high-power purge. If a continuous noise distribution is highly disperse (has a high variance), the high uncertainty makes the dictator less conflict-prone if he is advantaged and more conflict-prone if he is disadvantaged. A noise distribution that is tight around zero will reduce the risk of conflict and make the dictator’s expected utility of a high-powered target for a purge greater.

<sup>8</sup>Which is equivalent to a second round of conflict with all three players remaining in the full game



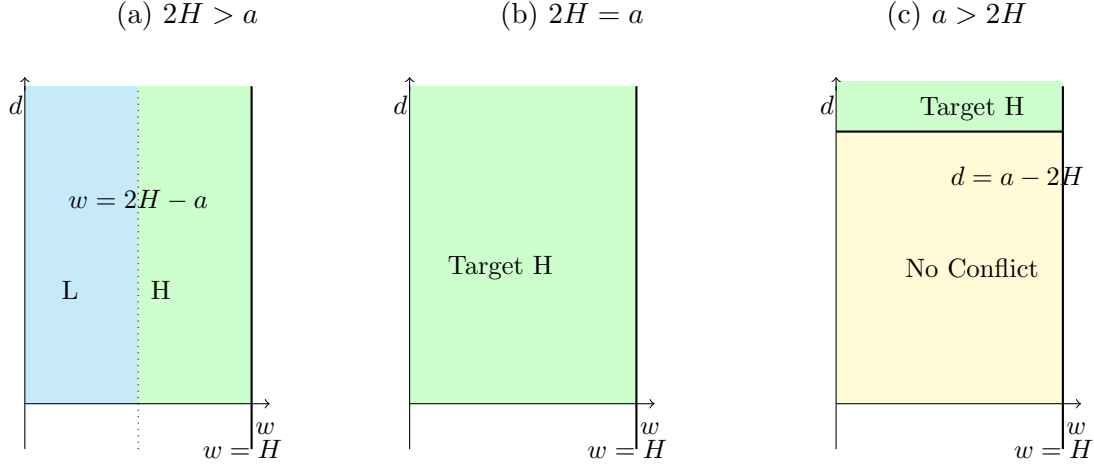


Figure 1: No Coalitions, Single Round of Conflict

As the possible conflicts become more uncertain, moving across the panels of Figure 1, the dictator will only target H to achieve the higher expected benefit of taking his power. When uncertainty is sufficiently high, even the prospect of taking H's power is not enough to tempt the dictator into a risky conflict. He will maintain his status quo power and not initiate any conflict, especially if his relative advantage,  $d$ , is low. As conflict uncertainty continues to grow (higher  $a$ ), the point at which the dictator is advantaged enough relative to the elites that he will risk conflict gets higher.

## Multiple Rounds of Conflict

While the above described his targeting decision in a single conflict, the dictator can use two rounds of conflict to oust the elites. In a multi-round conflict, a second round in which only two players remain is possible if there was conflict in the first round. Whoever won the first round conflict now has the most power and is the dictator; his targeting strategy when there is one opponent remaining is described in Lemma 2.

**Lemma 2.** *The round 2 dictator will target the remaining elite for conflict if the dictator's advantage is sufficiently high relative to the conflict uncertainty ( $d > a + w - 3H$ ), otherwise he will not initiate conflict.*

The full characterization of the dictator's two round targeting strategy takes both rounds of conflict into account: he will only receive the benefits of the full conflict if he survives both rounds, but taking one elite's power in the first round of conflict increases his power advantage in the second round. This is especially true if he fights H first: he takes H's power and uses this advantage against L, the weak opponent, defeating him with more certainty.<sup>9</sup> The dictator's optimal targeting strategy is described in Lemma 3 and Figure 2.

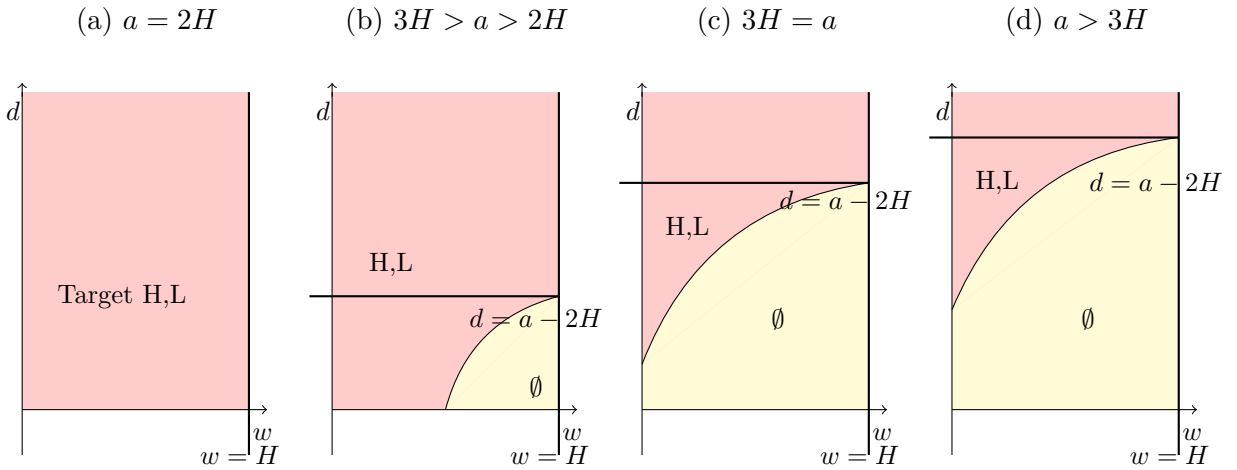
**Lemma 3.** *When the elites cannot form coalitions and there are two possible rounds of conflict, the dictator will target H in the first round and L in the second round for*

$$a \in (H + d + w, (\frac{1}{(2(3d+H+w))})((d + 3H - w)(2d + H + w) - \sqrt{(d + 3H - w)(16d^3 + (3H - w)(H + w)^2 + 16d^2(2H + w) + d(H + w)(17H + w))}))$$

and not initiate conflict for

$$a > (\frac{1}{(2(3d+H+w))})((d + 3H - w)(2d + H + w) - \sqrt{(d + 3H - w)(16d^3 + (3H - w)(H + w)^2 + 16d^2(2H + w) + d(H + w)(17H + w))})$$

Figure 2: Two Rounds of Conflict, No Coalitions



<sup>9</sup>Though the outcome is never certain (winning with probability 1) due to restrictions on  $a$ .

For low levels of uncertainty, the dictator will target H first and then, if he succeeds in his first conflict, target L in the second round. For all  $d > a - 3H + w$ , the expected utility from targeting two elites sequentially (H then L or L then H) is preferred to only targeting one (H or L, respectively). When the dictator is sufficiently advantaged relative to the uncertainty of the conflict, he prefers to target his stronger opponent first, then use those additional gains to take on L with an even stronger advantage. A weak dictator in an uncertain conflict environment would prefer the weak opponent first, targeting L first then using those lesser gains to stand against H. However, in the parameter space where L then H is preferred to H then L, the dictator is sufficiently weak and the conflicts sufficiently uncertain that no conflict is a dominant strategy. For more uncertain conflict environments, the level of power advantage that the dictator has and the level of overall expected benefit from beating both elites (i.e.  $w$  is not too high such that L's power is not too low) must be higher to induce him into a risky conflict.

The general trend of conflict initiation is the same whether the dictator is facing one or two rounds of potential consolidation. When uncertainty is low, the dictator will still initiate conflict; if two rounds of conflict are possible, however, he will always target H first whereas in the single round he would target L if L was not too weak to offer sufficient benefits. When conflict is more uncertain, having second round of conflict makes the dictator more conflict-prone: because the possible rewards from defeating both elites are higher than the single conflict rewards, the dictator is willing to take on the risk of conflict for more of the parameter space when beating both elites is possible. Note that this conflict behavior is limited by L's power: as L becomes weaker, the additional benefit of fighting the elites becomes too low to induce conflict and the dictator prefers to maintain his status quo power. In general, greater possible rewards from conflict will lead the dictator to take on more uncertain conflicts in order to achieve those rewards.

## Coalitions and Conflict with Three Players

Now allow the members of the regime to form coalitions with one another after the dictator has made his targeting choice. Whichever elite is not targeted can join the dictator, join the targeted elite, or stay out of the conflict. We can use backwards induction, starting with  $t = 2$  for Subgame Perfect Nash Equilibria. Depending on what happened in the previous round, there could be two or three players remaining. The only way that there could be three players remaining in the second round is if the dictator did not initiate conflict in the first round. Thus all three players have their initial endowments of power.<sup>10</sup>

### Coalition Formation

Given their expectations about the potential outcomes of the conflict and the losses or rewards that those outcomes entail, whichever elite is not targeted by the dictator for conflict must decide whether to join the conflict and, if so, on which side. First, should a non-targeted elite join the dictator in a coalition? As credible commitment to sharing spoils is ruled out by assumption, neither elite will ever join the dictator in a coalition as there is nothing to gain.<sup>11</sup> For similar reasoning, the weakest member of the regime, L, will not join a coalition with the stronger elite H. Because any possible gains from the conflict would accrue to H, L is at most indifferent between joining a coalition and remaining out of the conflict to keep his status quo power. As long as the conflict is uncertain, the elites are strong enough that the potential gains of beating the dictator (which would go to H) outweigh the risk of defeat. This logic is summarized in Lemma 4.

---

<sup>10</sup>Note that round 2 with three players remaining, if considered on its own, constitutes a one-shot version of the target selection, coalition formation, and conflict among the three regime actors as all individuals have their initial power endowments and no shadow of a future conflict.

<sup>11</sup>Recall that the dictator, by definition, is stronger than either of the elites individually (regardless of his relative power to an elite coalition). Thus the spoils of any successful conflict would accrue to the dictator—he cannot credibly commit to sharing any of the power gained from the target with a coalition partner. The non-targeted elite will not gain anything from joining the conflict on the side of the dictator: he is, at most, indifferent between joining the conflict and staying out. If the conflict outcome is uncertain, joining the conflict would give the elite, in expectation, a non-zero probability of losing everything, with no possible gains to offset the risk. Thus neither elite would ever join a coalition with the dictator.

**Lemma 4.A.** *L will never join a coalition with either the dictator or elite H.*

**Lemma 4.B.** *H will never join a coalition with the dictator, but will always join L in a coalition if L is targeted.*

Recall that this is the second round:<sup>12</sup> there is no shadow of future conflict driving H to fight the dictator, and he has no concern for L. The driver of coalition formation here is solely H's desire for power. He will join a coalition with L for the opportunity to defeat the dictator and take D's power upon winning, despite the risk this entails.

## Targeting Behavior

When choosing to initiate conflict and, if so, which elite to target, the dictator must take into account the potential formation of a coalition between H and L. In particular, for the parameter space in which H is willing to join L in a coalition (which is the case for all uncertain conflict), targeting L alone is no longer an on-path option for the dictator. The dictator's targeting decisions that take his relative power, the conflict uncertainty, and the expected behavior of the elites into account is summarized in Proposition 1 and Figure 3.

**Proposition 1.A.** *If the dictator is strongly advantaged ( $d > H - w$ ) and there is a single round of conflict, he will target H for  $a \in (H + d + w, 3H - w)$ , target L and fight the elite coalition for  $a \in (3H - w, 3H + d - w)$ , and not initiate conflict for  $a > 3H + d - w$ .*

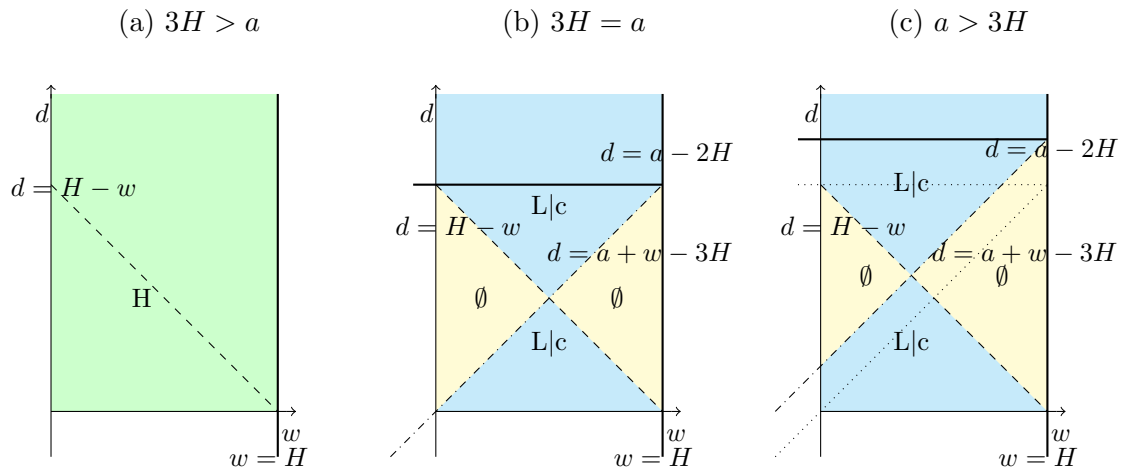
**Proposition 1.B.** *If the dictator is weakly advantaged ( $d < H - w$ ) and there is a single round of conflict, he will target H for  $a \in (H + d + w, 3H - w)$ , not initiate conflict for  $a \in (3H - w, 3H + d - w)$ , and target L and fight the elite coalition for  $a > 3H + d - w$ .*

---

<sup>12</sup>Or a one-shot conflict.

When the uncertainty of the conflict is relatively low and the difference between the two elites,  $w$ , is relatively low, the dictator will only target H. As conflict uncertainty increases, the dictator's behavior depends on his advantage relative to the elites. If the dictator is relatively strong ( $d > H - w$ ), he will target L in anticipation that H will join the elite coalition so that he can fight them both and take their power. Once L's power is too low to induce him into conflict, a relatively strong dictator will not initiate conflict. A weak dictator ( $d < H - w$ ), on the other hand, benefits from greater conflict uncertainty when fighting both elites as it washes out their advantage over him with greater noise. Moving from Figure 3(b) to 3(c) shows how increasing uncertainty adjusts the parameter space. Note that the impact of conflict uncertainty on conflict behavior is nuanced here. It is not the case that a dictator facing greater uncertainty will always “exhibit prudence” and avoid conflict, as is the case in interstate war onset (Bas and Schub, 2016). The relationship between the dictator's advantage over the elites and his advantage relative to the uncertainty of the conflict, when both taken into account, suggest that both extremely weak **and** extremely strong dictators will initiate conflict within their inner circle.

Figure 3: Coalitions, Single Conflict



Note how the possibility of coalitions makes the dictator more likely to initiate conflict relative his targeting strategy in a single conflict with no coalitions. Without coalitions, a high uncertainty environment would deter all but the dictators with very high relative from

initiating conflict (Figure 1 (c)). When coalitions are possible and the dictator can fight both elites at once, he will target an elite (L) and initiate conflict for much more of the parameter space.

## Two Players Remaining

If there are two players remaining, that implies that the dictator did initiate conflict in the first round and the third regime member either (a) did not join a coalition on either side or (b) joined a coalition and the coalition won, leaving the winning coalition members as the two remaining members of the regime, the stronger of whom is now the dictator.<sup>13</sup> Regardless of the identity of the dictator in the second round<sup>14</sup>, if the dictator's advantage (which may have been transferred to the new dictator) is sufficiently large, he will initiate conflict in the second round against his remaining opponent, regardless of if it is H or L who remains. If the dictator's advantage relative to uncertainty and the power of his opponents is sufficiently low, the conflict outcome is uncertain enough to deter the dictator from conflict and he will not initiate conflict, instead maintaining the power he gained from round 1.

**Lemma 5.** *The round 2 dictator will target the remaining elite for conflict if the dictator's advantage is sufficiently high relative to the conflict uncertainty ( $d > a + w - 3H$ ), otherwise he will not initiate conflict.*

Note that the dictator is now never interested in starting a risky, uncertain conflict as the weak dictator was in the previous subgame. Even if he began the game as a relatively weak

---

<sup>13</sup>Possible Conflicts: (1) D and H fought in round 1, now the winner (note that it does not matter whether there was a coalition or not. If L and H fought D together and won, H got all of D's power and L's power did not change. If D and L fought H together and won, D got all of H's power and L's power did not change. So regardless of what exactly the conflict was in round 1, the round 2 power distribution is  $2H + d$  for the current dictator versus  $H - w$  for the potential opponent.) is the dictator and can target L (or no conflict); (2) D and L fought in round 1, the winner (note that round 1 coalitions do matter for this option as the only way L can acquire D's power is through a one-on-one conflict. If H was involved, he would take D's power as the stronger coalition member and we would be in case 1.) is the dictator and can target H (or no conflict).

<sup>14</sup>Note that the identity of the dictator in the second round does not matter because there is no loss or decay of power upon transfer: it simply flows from the loser to the winner.

dictator ( $d < H - w$ ) or he began the game as an elite, by reaching the subgame with two remaining players as the dictator, he necessarily won the conflict in the first round and is now relatively advantaged in terms of power. Thus he will act as a strong dictator, preferring to maintain his status quo power than risk losing it in an uncertain conflict, similar to the advantaged dictator in the three-player subgame.

## Dynamic Coalitions and Conflict

The round 1 targeting behavior must take into account both the expected round 2 behavior as well as the possible coalition formations in round 1. Note that, as above, H joining the dictator in a coalition against L is a weakly dominated strategy. This is intuitive as H has nothing to gain from siding with the dictator: if he and the dictator win together, all of the gains go to the dictator. If they lose, H loses everything. H's round 1 coalition behavior is summarized in Lemma 6(a). As in the one-shot version, H will join a coalition with L. The benefits of H joining L in a coalition and possibly defeating the dictator are no longer just the dictator's power, but also the potential for taking L's power in the second round conflict. If H and L beat the dictator as a coalition in the first round, all of D's power is transferred to H and he will become the new dictator, targeting his former ally L when there is a sufficient dictatorial advantage. As before, L never gains from joining the conflict on either the side of H or the side of the dictator as neither party can commit to sharing the gains of conflict with him. As the weakest member of any coalition, the power of the loser will always go to the other coalition member, not to L. L's strategy is summarized in Lemma 6(b).

**Lemma 6.A.** *H will always join L in an elite coalition.*

**Lemma 6.B.** *L will never join a coalition*



## Targeting Behavior

The dictator's full equilibrium targeting decision depends on both the expected coalition behavior of the elites as well as the round 2 conflict and coalition decisions that he anticipates. Note that a choice of no target and thus no conflict in the first round does not end the game and lead to absolutely no conflict, it simply moves all regime members with their initial power endowments to the second round, the equilibrium strategies of which are described above in the round 2 subgame. In terms of expected utilities, the dictator is indifferent between targeting one player (H) in the first round and not initiating conflict in the second and not initiating conflict in the first round and targeting H in the second.

The dictator's round 1 targeting strategy is summarized in Proposition 2 and visualized in Figure 4. The dictator must choose between targeting L and fighting the coalition, targeting H in the first round and L in the second round, only targeting H (in the first or second round), and no conflict. Note that L alone is not available as a target option because H will join a targeted L in an elite coalition.

**Proposition 2.A.** *If the dictator is strongly advantaged ( $d > H - w$ ) and there are two possible rounds of conflict, he will target H in the first round and L in the second for  $a \in (H + d + w, \frac{1}{2}(3H - w + \sqrt{4d^2 + (-3H + w)^2 + 4d(H + w)})$ )*

*will target L and fight the coalition for*

*$a \in (\frac{1}{2}(3H - w + \sqrt{4d^2 + (-3H + w)^2 + 4d(H + w)}), 3H + d - w)$  and will not initiate conflict for  $a > 3H + d - w$ .*

**Proposition 2.B.** *If the dictator is weakly advantaged ( $d < H - w$ ) and there are two possible rounds of conflict, the dictator will target H in the first round and L in the second*

round for

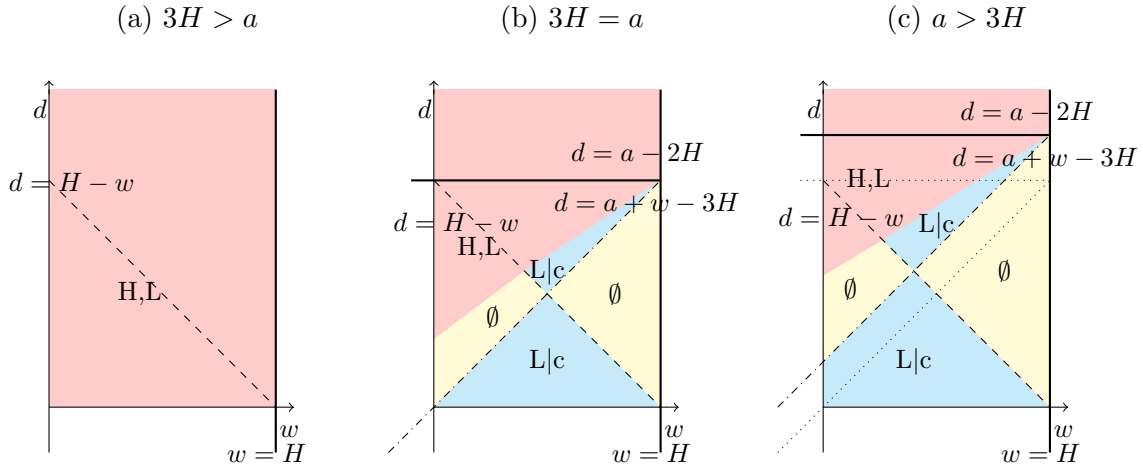
$$a \in (H + d + w, \frac{1}{(2(3d+H+w))}((d + 3H - w)(2d + H + w) - \sqrt{(d + 3H - w)(16d^3 + (3H - w)(H + w)^2 + 16d^2(2H + w) + d(H + w)(17H + w))}))$$

not initiate conflict for

$$a \in ((\frac{1}{(2(3d+H+w))}((d + 3H - w)(2d + H + w) - \sqrt{(d + 3H - w)(16d^3 + (3H - w)(H + w)^2 + 16d^2(2H + w) + d(H + w)(17H + w))})), 3H + d - w)$$

and target L and fight the elite coalition for  $a > 3H + d - w$

Figure 4: Coalitions, Two Rounds Conflict



Much like his multi-round targeting strategy without coalitions, targeting H and then using that additional power advantage to fight L in the second round is the dictator's equilibrium strategy when uncertainty is low. For higher levels of conflict uncertainty, the dictator's power advantage relative to the elites matters for his optimal strategy. A dictator who is weak relative to the elites ( $d < H - w$ ) will target H then L, initiate no conflict, or target L and fight the coalition depending on his strength and the difference between the elites' power,  $w$ . For the weakest dictator ( $d$  close to 0), higher uncertainty gives him the opportunity to take on a powerful elite coalition in a conflict where the elites' advantage is counteracted by

stochastic noise.<sup>15</sup> As his advantage increases (though he is still weak with  $d < H - w$ ), the dictator will avoid conflict altogether, then take on his sequential targeting strategy of H then L. A strong dictator ( $d > H - w$ ) prefers the same three strategies, but the parameter spaces in which he pursues them differ. When the benefits of conflict are low (as L's power decreases), even a relatively strong dictator will avoid initiating conflict. In an intermediate area when the dictator is of middling strength and the elites' power differential is not too great, the dictator will target L and fight the coalition.

Note that this sequential type of conflict, which occurs with and without coalitions when multiple conflicts are possible, in which elites are picked off one by one, is similar to the empirical patterns of early purges under Stalin in the Soviet Union. Further, the sequential targeting in this model, in which the high-powered elite is targeted first, is quite opposite to the “encircling” strategy suggested by Keller and Wang (N.d.) in which weak, peripheral members of the regime are removed first. This dictator targets H first because (a) it increases his chances of consolidating full power by the second round and (b) targeting L first could result in an elite coalition.

## Discussion

Who should the dictator target for a purge? In general, strong dictators will go after strong targets that will yield the most rewards. Who constitutes the strongest target depends on the ability of the elites (namely H) to form a coalition against the dictator: elites together in a coalition are a more lucrative, and therefore desirable, target for a strong dictator. This drive is mitigated by uncertainty as a relatively strong dictator will maintain his status quo power instead of initiating conflict when uncertainty over the conflict is too high, especially if the low-powered elite is not bringing much reward. Weak dictators are similarly conflict

---

<sup>15</sup>He is actually indifferent between fighting the conflict in the first round or the second round: the gains from the conflict and probability of beating the coalition is the same regardless of the timing.

prone, but not in order to preemptive strike or coup proof.<sup>16</sup> The power hungry dictator will utilize an uncertain conflict environment to his advantage to take on a strong elite coalition.

Most importantly for empirical work on purges and elite-dictator conflict, the relationship between the dictator's power advantage and whether any conflict will be observed is non-linear in the dictator's advantage. When conflict is uncertain, it is not the case that more elite purges are an indication of the personalist strength of the dictator. Indeed, for high levels of uncertainty, the observation of an attempt at a mass purge would indicate the *weakness* of the dictator vis-a-vis elites: he is willing to risk his small power advantage for the possibility of a huge payoff if he is able to defeat the elite coalition. A lack of conflict between the dictator and his elite allies is, likewise, not an indication of weakness. While a lack of observed conflict may indicate a relatively weak dictator in a semi-uncertain conflict environment, it may also be an extremely strong, advantaged dictator who is not willing to risk losing power in an uncertain conflict. The uncertainty of the conflict, as well as the difference in power between the possible purge targets, must be considered in order to understand the observed relationship between a dictator's power and purges.

### **Elite Heterogeneity and Coalitions**

While the ability of the elites to form coalitions makes the dictator more conflict-prone, particularly at higher levels of conflict uncertainty due to the higher possible benefit of removing both elites, the affect on the elites is heterogeneous. The higher powered elite's ability to join L in a coalition benefits him: he can join a conflict where he may not have before, use L's power with his own against the dictator, and reap the potential benefits if the two elites are able to oust the dictator. As the only coalition joiner in equilibrium, H will only form a coalition with L and join the conflict when, in response to (or in anticipation of,

---

<sup>16</sup>These incentives are outside the scope of the model as elites here cannot initiate conflict. Even without these concerns (which we may expect will make weak dictators more likely to initiate a purge), a weak dictator is willing to initiate conflict.

when there are two rounds of possible conflict) the dictator's targeting choice, his expected utility is greater.

This is not the case for L, who is unable to reject H as a coalition partner, and ends up being targeted because of H's joining. In the single conflict benchmark, L both benefits from and is hurt by H's ability to join him in different parts of the parameter space. When uncertainty over conflict outcomes is relatively low, the dictator switches from targeting L to targeting H because H would join L in a coalition, so L gets to avoid a conflict which he is very likely to lose.<sup>17</sup> In this case, L benefits from the possibility of a coalition because he avoids the conflict maintains his status quo power. For higher levels of uncertainty, however, L is always worse off. Instead of avoiding conflict, he becomes the target and H, who joins him in a coalition, reaps the benefits of the fight. Similarly when two rounds of conflict are possible, L is weakly worse off<sup>18</sup> from the possibility of elite coalitions. While for high levels of uncertainty being targeted in the second round might be preferred for L as it gives him the opportunity to win all of the power of the regime, this will not occur in equilibrium.<sup>19</sup> Instead, there is now incentive for the dictator to target L, knowing the coalition will form and he will fight both elites, whereas before there was no conflict. These differential effects of coalitions formation on different elites depends on who gains the most from winning a conflict as a coalition. Even if H and L split the gains of the dictator's power if they beat him, that is not always sufficient compensation for L when, all else equal, he would not have been targeted without the coalition.<sup>20</sup> The inner workings of elite coalitions formed against a dictator's purge attempt warrant greater study and the inclusion of heterogeneity in elite power is an important aspect of understanding the different incentives and rewards of coalition behavior.

---

<sup>17</sup>In this parameter space of  $a < 2H + d - w$ , L would prefer not being targeted to fighting the dictator; but note that for higher uncertainty, when the dictator's advantage is overcome by noise, L would like to be targeted for the chance to win the dictator's power. In equilibrium, however, he will not be.

<sup>18</sup>Either indifferent or strictly worse, depending on the parameter space

<sup>19</sup>L prefers the being targeted in the second round to no conflict when  $a > 3H + d - w$ , but the dictator's targeting strategy  $H, L$  is not on path in that parameter space.

<sup>20</sup>For some parameter space.

## Conclusions

In an environment in which decisions are made, and disagreements are settled, under the shadow of extreme violence, even the simplest of personnel and staffing decisions can become uncertain conflicts (Svolik, 2012). This uncertainty is exacerbated when elite “allies,” with independent bases of power, wealth, and arms, are the regime officials being demoted or eliminated. By introducing heterogeneity in power to the ruling coalition, I have shown how the dictator will target his strongest possible opponent, whether the individual or an elite coalition, when he has a strong power advantage. While conflict uncertainty makes a strong dictator more conflict-averse, greater uncertainty induces a weak dictator into more conflict. Even without informational issues in which the dictator can hide his intent to personalize power, the balance of power among the dictator and among his allies can explain when the dictator will initiate conflict, who will be targeted for elimination, and the type of conflict observed. The differences in elite power yield differences in elite incentives to join one another in coalitions against the dictator. While a more powerful elite uses coalitions to his benefit to join the conflict against the dictator, a lower-powered elite is often hurt by the possibility of a coalition because the dictator’s anticipation of the coalition makes the weaker elite a more attractive target. The process by which power is consolidated into the hands of one leader not only has life and death consequences for the elites being targeted, but repercussions for the makeup of the ruling coalition, with downstream effects on lower personnel, policy implementation, and the distribution of public goods and services to citizens.

## References

- Acemoglu, Daron, Georgy Egorov and Konstantin Sonin. 2008. “Coalition formation in non-democracies.” *The Review of Economic Studies* 75(4):987–1009.
- Arriola, Leonardo, Jed DeVaro and Anne Meng. 2017. “Elite Cooptation and Opposition Fragmentation in Electoral Authoritarian Regimes.” Working Paper.
- Barros, Robert. 2002. *Constitutionalism and dictatorship: Pinochet, the Junta, and the 1980 constitution*. Vol. 4 Cambridge University Press.
- Barros, Robert. 2016. “On the outside looking in: Secrecy and the study of authoritarian regimes.” *Social Science Quarterly* 97(4):953–973.
- Bas, Muhammet A and Robert J Schub. 2016. “How uncertainty about war outcomes affects war onset.” *Journal of Conflict Resolution* 60(6):1099–1128.
- Bas, Muhammet A and Robert Schub. 2017. Theoretical and Empirical Approaches to Uncertainty and Conflict in International Relations. In *Oxford Research Encyclopedia of Politics*.
- Boix, Carles and Milan W Svobik. 2013. “The foundations of limited authoritarian government: Institutions, commitment, and power-sharing in dictatorships.” *The Journal of Politics* 75(2):300–316.
- Bueno De Mesquita, Bruce, Alastair Smith, Randolph M Siverson and James D Morrow. 2003. *The logic of political survival*. MIT press.
- Carter, Brett L and Mai Hassan. 2021. “Regional governance in divided societies: evidence from the Republic of Congo and Kenya.” *The Journal of Politics* 83(1):000–000.
- Coughlin, Con. 2005. *Saddam: His Rise and Fall*. Harper Collins.
- Cuttner, Allison. 2021. “Empowering Elite Rivals under Contested Dictatorship.”.

- Gandhi, Jennifer. 2008. *Political Institutions under Dictatorship*. Cambridge University Press.
- Gandhi, Jennifer and Jane Lawrence Sumner. 2020. "Measuring the consolidation of power in nondemocracies." *The Journal of Politics* 82(4):1545–1558.
- Geddes, Barbara, Joseph Wright and Erica Frantz. 2014. "Autocratic breakdown and regime transitions: A new data set." *Perspectives on Politics* 12(2):313–331.
- Geddes, Barbara, Joseph Wright and Erica Frantz. 2018. *How dictatorships work: Power, personalization, and collapse*. Cambridge University Press.
- Gehlbach, Scott, Konstantin Sonin and Milan W Svobik. 2016. "Formal models of nondemocratic politics." *Annual Review of Political Science* 19:565–584.
- Getty, John Archibald. 1987. *Origins of the great purges: the Soviet Communist Party reconsidered, 1933-1938*. Number 43 Cambridge University Press.
- Goldring, Edward. 2020. "The Effects of Elite Purges on Attempted and Successful Coups."
- Haber, Stephen. 2006. Authoritarian Government. In *The Oxford Handbook of Political Economy*, ed. Barry R. Weingast and Donald A. Wittman. Oxford University Press pp. 693–707.
- Hassan, Mai. 2017. "The strategic shuffle: Ethnic geography, the internal security apparatus, and elections in Kenya." *American Journal of Political Science* 61(2):382–395.
- Hornsby, Charles. 2013. *Kenya: A history since independence*. IB Tauris.
- Keller, Franziska Barbara and Yuhua Wang. N.d. "Farewell Comrades! Power Struggle and Elite Removal during Authoritarian Leadership Succession." <http://scholar.harvard.edu/files/yuhuawang/files/farewell-comrades.pdf>.
- Khlevniuk, Oleg V. 2008. *Master of the house: Stalin and his inner circle*. Yale University Press.



- Konrad, Kai A. 2009. *Strategy and Dynamics in Contests*. Oxford University Press.
- Kosterina, Svetlana. 2017. "Ambition, personalist regimes, and control of authoritarian leaders." *Journal of Theoretical Politics* 29(2):167–190.
- Kramon, Eric and Daniel N Posner. 2016. "Ethnic favoritism in education in Kenya." *Quarterly Journal of Political Science* 11(1):1–58.
- Little, Andrew T. 2017. "Coordination, learning, and coups." *Journal of Conflict Resolution* 61(1):204–234.
- Lu, Xi and Peter L Lorentzen. 2016. "Rescuing autocracy from itself: China's anti-corruption campaign." *SSRN Electronic Journal* 6.
- Luo, Zhaotian and Arturas Rozenas. 2019. "Ruling the Ruling Coalition: Information Control and Authoritarian Power-Sharing." *Available at SSRN 2978254* .
- Meng, Anne. 2019. "Accessing the state: Executive constraints and credible commitment in dictatorship." *Journal of Theoretical Politics* .
- Montagnes, B Pablo and Stephane Wolton. 2019. "Mass Purges: Top-Down Accountability in Autocracy." *American Political Science Review* 113(4):1045–1059.
- Myerson, Roger B. 2008. "The autocrat's credibility problem and foundations of the constitutional state." *American Political Science Review* 102(1):125–139.
- Singh, Naunihal. 2014. *Seizing power: The strategic logic of military coups*. JHU Press.
- Sudduth, Jun Koga. 2017. "Strategic logic of elite purges in dictatorships." *Comparative Political Studies* 50(13):1768–1801.
- Svolik, Milan W. 2009. "Power sharing and leadership dynamics in authoritarian regimes." *American Journal of Political Science* 53(2):477–494.
- Svolik, Milan W. 2012. *The politics of authoritarian rule*. Cambridge University Press.

Taubman, William. 2003. *Khrushchev: the man and his era*. WW Norton & Company.

Van Dam, Nickolas. 1996. *Struggle for Power in Syria*. St. Martin's Press.

Woldense, Josef. 2018. "The ruler's game of musical chairs: Shuffling during the reign of Ethiopia's last emperor." *Social Networks* 52:154–166.

# Online Appendix

## No Coalitions Possible

### Round 2: Three Players

Because there is no coalition stage, the only action is the dictator's targeting decision

$\epsilon$  is distributed uniform, so

$$F_\epsilon(x) = \begin{cases} 0 & x < -a \\ \frac{x+a}{2a} & x \in [-a, a) \\ 1 & x \geq a \end{cases}$$

Assume  $a > H + d + w$  such that all conflicts are uncertain. Thus the probability that the dictator wins against H is  $\frac{d+a}{2a}$  and the probability the dictator wins against L is  $\frac{d+w+a}{2a}$

$$U_D(H) = \left(\frac{d+a}{2a}\right)(2H + d) + \left(1 - \frac{d+a}{2a}\right)(0)$$

$$U_D(L) = \left(\frac{d+w+a}{2a}\right)(2H + d - w) + \left(1 - \left(\frac{d+w+a}{2a}\right)\right)(0)$$

$$U_D(\emptyset) = H + d$$

H versus no conflict

$$\left(\frac{d+a}{2a}\right)(2H + d) > H + d$$

$2H + d > a$  if this holds, the dictator prefers targeting H relative to no conflict (equivalently

$$H > \frac{a-d}{2})$$

L versus no conflict

$$\left(\frac{d+w+a}{2a}\right)(2H + d - w) > H + d$$

$2H + d - w > a$  if this holds, the dictator prefers targeting L relative to no conflict (equiva-

$$\text{lently, } H > \frac{a+w-d}{2})$$

H versus L

$$\left(\frac{d+a}{2a}\right)(2H + d) > \left(\frac{d+w+a}{2a}\right)(2H + d - w)$$

$a > 2H - w$  if this holds, the dictator prefers targeting H relative to targeting L (equivalently

$$H < \frac{a+w}{2})$$

## Round 2: Two Players

If there are two players remaining, there was a conflict in the first round between D and H or D and L.

If the first round conflict was between D and H, the winner (and second round dictator) now has a power endowment of  $2H + d$ . Should he target L?

$$\frac{H+d+w+a}{2a}(3H + d - w) > 2H + d$$

$3H + d - w > a$  if this holds, the second round dictator should target L relative to no conflict (equivalently  $H > \frac{a+w-d}{3}$ )

If the first round conflict was between D and L, the winner (and second round dictator) now has a power endowment of  $2H + d - w$ . Should he target H?

$$\frac{H+d-w+a}{2a}(3H + d - w) > 2H + d - w$$

$3H + d - w > a$  if this holds, the second round dictator should target H relative to no conflict (equivalently  $H > \frac{a+w-d}{3}$ )

## Round 1 Targeting

Possible paths for two round targeting:

$$U_D(\emptyset, H) = \frac{d+a}{2a}(2H + d)$$

$$U_D(\emptyset, L) = \frac{d+w+a}{2a}(2H + d - w)$$

$$U_D(H, L) = \left(\frac{d+a}{2a}\right)\left(\frac{H+d+w+a}{2a}\right)(3H + d - w)$$

$$U_D(L, H) = \left(\frac{d+w+a}{2a}\right)\left(\frac{H+d-w+a}{2a}\right)(3H + d - w)$$

$$U_D(H, \emptyset) = \frac{d+a}{2a}(2H + d)$$

$$U_D(L, \emptyset) = \frac{d+w+a}{2a}(2H + d - w)$$

$$U_D(\emptyset, \emptyset) = H + d$$

H,L versus L,H

$$\left(\frac{d+a}{2a}\right)\left(\frac{H+d+w+a}{2a}\right)(3H + d - w) > \left(\frac{d+w+a}{2a}\right)\left(\frac{H+d-w+a}{2a}\right)(3H + d - w)$$

$a > H - d - w$  if this holds, H,L is preferred to targeting L, H (equivalently  $H < a + d + w$ )

H,L versus H alone (note  $H, \emptyset$  and  $\emptyset, H$  are the same expected utilities)

$$\left(\frac{d+a}{2a}\right)\left(\frac{H+d+w+a}{2a}\right)(3H + d - w) > \frac{d+a}{2a}(2H + d)$$

$3H + d - w > a$  if this holds, targeting H and L is preferred to H alone (equivalently,

$$H > \frac{a+w-d}{3})$$

L,H versus L alone (note  $L, \emptyset$  and  $\emptyset, L$  are the same expected utilities)

$$\left(\frac{d+w+a}{2a}\right)\left(\frac{H+d-w+a}{2a}\right)(3H + d - w) > \frac{d+w+a}{2a}(2H + d - w)$$

$3H + d - w > a$  if this holds, targeting L and H is preferred to L alone (equivalently,

$$H > \frac{a+w-d}{3})$$

H,L versus L alone

$$\left(\frac{d+a}{2a}\right)\left(\frac{H+d+w+a}{2a}\right)(3H + d - w) > \left(\frac{d+w+a}{2a}\right)(2H + d - w)$$

$$\frac{2a(a+d)(a+d+H+w)}{(a+d+w)} > \frac{2H+d-w}{3H+d-w}$$

As  $3H + d - w > 2H + d - w$ , the right side is less than 1

$$\frac{2a(a+d)(a+d+H+w)}{(a+d+w)}$$

H,L is preferred if  $2a(a+d)(a+d+H+w) > (a+d+w)$  which holds for sufficiently high

$a, d$

(see round two for comparisons of H and L alone and relative to no conflict)

Where does  $H = \frac{a+w-d}{3}$  fall relative to the other cutpoints?

$$\frac{a+w-d}{3} < \frac{a+w-d}{2}$$

Is  $\frac{a+w-d}{3} > \frac{a-d}{2}$ ?

$$2(a+w-d) > 3(a-d)$$

$$d+2w > a$$

This may be possible, but by assumption  $a > H + d + w$

$$d+2w > a > H + d + w$$

$w > H$  this is a contradiction by definition so it must be the case that  $\frac{a-d}{2} > \frac{a+w-d}{3}$

## Coalitions Possible

### Round 2: Three Players

#### Coalition Formation

**Lemma 4.A.** *L will never join a coalition with either the dictator or elite H.*

**Lemma 4.B.** *H will never join a coalition with the dictator, but will always join L in a coalition if L is targeted.*

For L, not joining a coalition weakly dominates joining either D or H. H will not join a coalition with D. If  $d > H - w$ , for every vector  $d, H, w, a$ , there exists a unique threshold  $a_{Coal}(d, H, w) > 0$  such that H is indifferent between joining a coalition with L or staying out of the conflict and joining L against D is preferred for all  $a > a_{Coal}$ . If  $d < H - w$ , H will always join a coalition with L.

Assume H is targeted, what does L do?

$$E[u_L(\text{join } D)] = P(\text{coalition wins})(H - w) + P(H \text{ wins})(0) = F_\epsilon(H + d - w)(H - w)$$

$$E[u_L(\neg \text{join})] = H - w$$

$$E[u_L(\text{join})] = P(D \text{ win})(0) + P(D \text{ lose})(H - w) = (1 - F_\epsilon(d + w - H))(H - w)$$

$L$  join iff  $F_\epsilon(H + d - w)(H - w) > H - w$ . This will never occur as  $F_\epsilon$  is a proper CDF.  $L$  would, at most, be indifferent between joining the dictator and staying out of the conflict.  $L$  will join  $H$  in a coalition if and only if  $(1 - F_\epsilon(d + w - H))(H - w) > H - w$ , which is never the case. Even if the probability that the dictator wins goes to 0,  $L$  is at most indifferent.

Assume  $L$  is targeted. What should  $H$  do? Recall the CDF of the Uniform distribution, with symmetric bounds  $[-a, a]$ , is

$$F_\epsilon(x) = \begin{cases} 0 & x < -a \\ \frac{x+a}{2a} & x \in [-a, a] \\ 1 & x \geq a \end{cases}$$

$$E[u_H(\text{join } L)] = P(D \text{ wins})(0) + P(D \text{ loses})(2H + d) = (1 - F_\epsilon(d + w - H))(2H + d)$$

$$E[u_H(\neg \text{join})] = H$$

$$E[u_H(\text{join } D)] = P(\text{coalition wins})(H) + P(L \text{ wins})(0) = F_\epsilon(H + d + w)(H)$$

$H$  join iff  $F_\epsilon(H + d + w)(H) > H$ . This will never occur as  $F_\epsilon$  is a proper CDF.  $H$  would, at most, be indifferent between joining the dictator and staying out of the conflict.

$H$  should join the coalition with  $L$  if  $(1 - F_\epsilon(d + w - H))(2H + d) > H$

If  $a \geq d + w - H > 0$ ,  $F_\epsilon(d + w - H) = \frac{d+w-H+a}{2a}$

Join coalition iff  $1 - \frac{d+w-H+a}{2a} \geq \frac{H}{2H+d}$

$a \geq d + w + H + \frac{2H(w-H)}{d}$ . Denote the point at which  $H$  is indifferent between joining the conflict in a coalition with  $L$  and staying out as  $a_{Coal} = d + w + H + \frac{2H(w-H)}{d}$ . For all  $a > d + w + H + \frac{2H(w-H)}{d}$   $H$  will join the coalition with  $L$  against the dictator. Note that

$a_{Coal} < H + d + w$ , the lower bound on  $a$  such that all conflicts will be uncertain. Thus  $H$  will always join a coalition with  $L$  in this parameter space.

If  $d < H - w$

$$\text{If } a \geq H - d - w, F_\epsilon(d + w - H) = \frac{d+w-H+a}{2a}$$

Join coalition iff  $1 - \frac{d+w-H+a}{2a} \geq \frac{H}{2H+d}$   $a \geq d + w + H + \frac{2H(w-H)}{d}$ . Denote the point at which  $H$  is indifferent between joining the conflict in a coalition with  $L$  and staying out as  $a_{Coal} = d + w + H + \frac{2H(w-H)}{d}$ . For all  $a > d + w + H + \frac{2H(w-H)}{d}$ ,  $H$  will join the coalition with  $L$  against the dictator. (note that  $a_{Coal}$  is the same as above). As above,  $H$  will always join a coalition with  $L$  in this parameter space.

## Targeting Behavior

**Proposition 1.A.** *If the dictator is strongly advantaged ( $d > H - w$ ) and there is a single round of conflict, he will target  $H$  for  $a \in (H + d + w, 3H - w)$ , target  $L$  and fight the elite coalition for  $a \in (3H - w, 3H + d - w)$ , and not initiate conflict for  $a > 3H + d - w$ .*

$d > H - w$

$$F_\epsilon(d + w - H) = \frac{d+w-H+a}{2a}, F_\epsilon(d + w) = \frac{d+w+a}{2a}, F_\epsilon(d) = \frac{d+a}{2a}.$$

The noise distribution is disperse enough that there are no certain conflict outcomes. Even against his weakest opponent, the dictator is uncertain about whether he will win the conflict. Note as  $a \rightarrow \infty$ , all of the win probabilities go to  $\frac{1}{2}$ . For sufficiently high  $a$ , the dictator will prefer the status quo and will not initiate conflict.

Crossing points:  $E[u_D(\emptyset)] > E[u_D(L|\neg c)]$  if  $a > 2H + d - w$

$E[u_D(\emptyset)] > E[u_D(H)]$  if  $a > 2H + d$

$E[u_D(\emptyset)] > E[u_D(L|c)]$  if  $a > 3H + d - w$

Note that  $3H + d - w > 2H + d > 2H + d - w$ , thus if the coalition forms, fighting the



coalition will be the preferred conflict for  $a$  around  $3H + d - w$ , then no conflict is preferred for all  $a > 3H + d - w$

The coalition will form across the entire range. For  $a < 3H + d - w$ , the dictator is choosing between targeting  $H$  alone or  $L$  knowing that the coalition will form (recall from above that for  $a > 3H + d - w$ , the dictator does not initiate conflict). The dictator is indifferent between targeting  $H$  and fighting the coalition at  $a = 3H - w$ .

If  $H - w < d < 2(H - w)$ ,  $H + d + w < 3H - w$ , then for  $a \in (H + d + w, 3H - w)$  the dictator targets  $H$ ; for  $a \in (3H - w, 3H + d - w)$ , the dictator targets  $L$  and the coalition forms; for  $a > 3H + d - w$ , the dictator does not initiate conflict. If  $d > 2(H - w)$ ,  $3H - w$  is out of the relevant range of uncertainty. For  $a \in (H + d + w, 3H + d - w)$ , the dictator targets  $L$  and fights the coalition. For  $a > 3H + d - w$ , the dictator does not initiate conflict.

$$d < H - w$$

$$E[u_D(H)] = F_\epsilon(d)(2H + d)$$

$$E[u_D(L)] = F_\epsilon(d + w - H)(3H + d - w)$$

$$E[u_D(\emptyset)] = H + d$$

$H$  alone is preferred to no target if  $a < 2H + d$

$H$  alone is preferred to fighting the coalition if  $a < 3H - w$

$L(\text{coal})$  is preferred to no target if  $a > 3H + d - w$

Thus targeting  $H$  is preferred for  $a \in (H + d + w, 2H + d)$ , no target is preferred for  $a \in [2H + d, 3H + d - w)$  targeting  $L$  and fighting the coalition is preferred for  $a > 3H + d - w$

## Round 2: Two Players Remaining

**Lemma 5.** *The round 2 dictator will target the remaining elite for conflict if the dictator's advantage is sufficiently high relative to the conflict uncertainty ( $d > a + w - 3H$ ), otherwise he will not initiate conflict.*

Case 1: Dictator (whether identity was D or H in rd 1) has  $2H + d$ . Denote round 2 dictator as  $D'$

$$E[u_{D'}(\emptyset)] = 2H + d$$

$$E[u_{D'}(L)] = F_\epsilon(H + d + w)(3H + d - w) \text{ as } a > H + d + w$$

$$\text{Target } L \text{ if } F_\epsilon(H + d + w)(3H + d - w) > 2H + d$$

$$a < 3H + d - w$$

Note that  $3H + d - w > H + d + w$ . Thus given this history,  $D'$  will target  $L$  in round 2 for all  $a < 3H + d - w$

in terms of  $d$ , target if  $d > a + w - 3H$

Case 2: Dictator has  $2H + d - w$ . Denote round 2 dictator as  $D''$

$$E[u_{D''}(\emptyset)] = 2H + d - w$$

$$E[u_{D''}(H)] = F_\epsilon(H + d - w)(3H + d - w) \text{ as } a > H + d - w$$

$$\text{Target } H \text{ if } F_\epsilon(H + d - w)(3H + d - w) > 2H + d - w$$

$$a < 3H + d - w$$

Note that  $3H + d - w > H + d - w$ . Thus given this history,  $D''$  will target  $H$  in round 2 for all  $a < 3H + d - w$

in terms of  $d$ , target if  $d > a + w - 3H$

## Round 1

### Coalition Behavior

**Lemma 6.A.** *H will always join L in an elite coalition.*

Denote the expected utility of H not joining any coalition in round 1, staying out of the conflict, as  $E[u_H(\neg join_1)]$ . If H does not join, the winner of the conflict between D and L will choose to target H or not in accordance with the round 2 strategies defined above, which depend on the relative uncertainty (a) of the conflict.

Denote the expected utility of H joining a coalition with the dictator in round 1 as  $E[u_H(join_{1D})]$ . Recall that if the dictator and H together defeat L, all of the gains will go to the dictator and he will retain his position. He will target the remaining player, H, in accordance with the above strategies.

Denote the expected utility of H joining L in a coalition against the dictator in round 1 as  $E[u_H(join_{1L})]$ . If the elite coalition defeats the dictator, all of the gains will go to H and he will target L in the second round in accordance with the above strategies.

$$a \in [H + d + w, 3H + d - w)$$

$$E[u_H(\neg join_1)] = (1 - F_\epsilon(H + d - w))(3H + d - w)$$

$$E[u_H(join_{1D})] = F_\epsilon(H + d + w)(1 - F_\epsilon(H + d - w))(3H + d - w)$$

$$E[u_H(join_{1L})] = (1 - F_\epsilon(d + w - H))F_\epsilon(H + d + w)(3H + d - w)$$

Note joining D is dominated by not joining

$$E[u_H(join_{1L})] - E[u_H(\neg join_1)] = 0$$

a at the indifference point

$$= \frac{(2H + d - w)(3H + d - w) \pm \sqrt{(3H + d - w)^2(4Hd + 5H^2 - 4dw - 4Hw)}}{3H + d - w}$$

$$= 2H + d - w \pm \frac{\sqrt{(3H+d-w)^2(4Hd+5H^2-4dw-4Hw)}}{3H+d-w}$$

recall the bounds of this case are  $H + d + w$  and  $3H + d - w$

Where do the roots fall relative to the case bounds?

$$\text{let } \eta = \frac{\sqrt{(3H+d-w)^2(4Hd+5H^2-4dw-4Hw)}}{3H+d-w}$$

Is  $2H + d - w - \eta < H + d + w$ ?

$H - 2w - \eta < 0$ ? We know that  $-2w \in (0, -2H)$  by definition of  $w$ , so whether or not the inequality holds depends on the relative size of  $H$  and  $\eta$  i.e. if  $\eta > H$ , the expression is always less than 0 and the lower root is less than the lower bound.

$$\sqrt{(3H + d - w)^2 (4Hd + 5H^2 - 4dw - 4Hw)} > H(3H + d - w)$$

$$(4Hd + 5H^2 - 4dw - 4Hw) > H^2 \text{ (square both sides and divide by } (3H + d - w)^2) Hd + H^2 - dw - Hw > 0$$

$H(H + d) - w(H + d) > 0$  this is always true as  $H > w$  by definition

Therefore  $\eta > H$ , therefore the lower root is less than the lower bound.

Compare upper root and upper bound:

$$2H + d - w + \eta > 3H + d - w$$

this holds as  $\eta > H$

$$\frac{\partial}{\partial a} E [u_H(\neg \text{ join } )] - E [u_H(\text{ join }_L)] = -\frac{(3H+d-w)((H+d+w)(d+w-H)+a(w-d-2H))}{2a^3}$$

Note if  $d < H - w$ ,  $d + w - H$  and  $w - 2H - d$  are both negative, making the entire first derivative positive for all  $a$ . Further note that when  $d < H - w$ , the lower root discussed above does not exist. The expression of utility difference crosses zero only once, at  $2H + d - w + \eta$  and is increasing everywhere. Therefore it the expression is less than zero for all  $a < 3H + d - w < 2H + d - w + \eta$  and greater than zero for all  $a > 2H + d - w + \eta$ .

Thus H prefers to join L in this range if  $d < H - w$

If  $d > H - w$ , both roots exist and the expression of utility difference is non-monotonic.

$$\text{Note the sign of } \frac{\partial}{\partial a} E [u_H(\neg \text{ join } )] - E [u_H(\text{ join }_L)] = -\frac{(3H+d-w)((H+d+w)(d+w-H)+a(w-d-2H))}{2a^3} \text{ depends—}$$

on the sign of  $(H + d + w)(d + w - H) + a(w - d - 2H)$  as  $3H + d - w$  is always positive and  $a$  is always positive by definition.

sign( $H + d + w$ )( $d + w - H$ ) +  $a(w - d - 2H)$  evaluated at the upper and lower roots of  $E[u_H(\neg \text{join})] - E[u_H(\text{join}_L)] = 0$

Upper root:  $d^2 - H^2 + 2dw + w^2 + (2H + d - w + \eta)(w - d - 2H) > 0$   $-d(4H - 4w + \eta) + \eta(w - 2H) + H(4w - 5H) > 0$

As  $H > w$  and  $d, \eta > 0$  by definition, this is a contradiction. Therefore  $-d(4H - 4w + \eta) + \eta(w - 2H) + H(4w - 5H) < 0$ . Therefore  $\frac{\partial}{\partial a} E[u_H(\neg \text{join})] - E[u_H(\text{join}_L)] > 0$ . The expression is increasing at the upper root.

Lower root:  $d^2 - H^2 + 2dw + w^2 + (2H + d - w - \eta)(w - d - 2H) > 0$  (recall  $d > H - w$ )  $\eta(2H + d - w) > H^2 + 4(H + d)(H - w)$  while we know from above  $\eta > H$ , the actual magnitude matters for signing the derivative here. So plug in the actual value of  $\eta = \frac{\sqrt{(3H+d-w)^2(4Hd+5H^2-4dw-4Hw)}}{3H+d-w}$   $(2H + d - w)\sqrt{(3H + d - w)^2(4Hd + 5H^2 - 4dw - 4Hw)} > (H^2 + 4(H + d)(H - w))(3H + d - w)$

Square both sides:

$(2H + d - w)^2(3H + d - w)^2(4Hd + 5H^2 - 4dw - 4Hw) > (H^2 + 4(H + d)(H - w))^2(3H + d - w)^2$

$(2H + d - w)^2 > 4Hd + 5H^2 - 4dw - 4Hw$

$d^2 + w^2 + 2dw - H^2 > 0$

$(H + d + w)(d + w - H) > 0$

Note this holds as  $d > H - w$ . Thus  $(H + d + w)(d + w - H) + (2H + d - w - \eta)(w - d - 2H) > 0$  so the derivative evaluated at the lower root is negative.

As  $E[u_H(\neg \text{join})] - E[u_H(\text{join}_L)]$  is decreasing at the lower root and increasing at the upper root, the expression is negative between the two roots. As the case bounds are within the roots as shown above, the expression is negative for the full range of  $a$  in this case. Thus H prefers to join L in a coalition for this case.

$$a > 3H + d - w$$

$E[u_H(\neg \text{join}_1)] = H$  as there will be no round 2 conflict

$$E[u_H(\text{join}_{1D})] = F_\epsilon(H + d + w)(H)$$

$$E[u_H(\text{join}_{1L})] = (1 - F_\epsilon(d + w - H))(2H + d)$$

Note don't join dominates joining D as  $F$  is a proper CDF

Join L is preferred to don't join iff  $(1 - F_\epsilon(d + w - H))(2H + d) > H$

$$a > \frac{(2H+d)(d+w-H)}{d}$$

Note  $d + w - H > 0$  if  $d > H - w$ , so if strong coalition, join L for whole range

$$\text{Is } \frac{(2H+d)(d+w-H)}{d} < 3H + d - w?$$

$$2w(H + d) < 2H(H + d) \text{ true by definition}$$

Therefore H will join L across this entire range of  $a$ .

**Lemma 6.B.** *L will never join a coalition*

Denote the expected utility of L not joining any coalition in round 1, staying out of the conflict, as  $E[u_L(\neg\text{join}_1)]$ . If L does not join, the winner of the conflict between D and H will choose to target L or not in accordance with the round 2 strategies defined above, which depend on the relative uncertainty ( $a$ ) of the conflict.

Denote the expected utility of L joining a coalition with the dictator in round 1 as  $E[u_L(\text{join}_{1D})]$ .

Recall that if the dictator and L together defeat H, all of the gains will go to the dictator and he will retain his position. He will target the remaining player, L, in accordance with the above strategies.

Denote the expected utility of L joining H in a coalition against the dictator in round 1 as  $E[u_L(\text{join}_{1H})]$ . If the elite coalition defeats the dictator, all of the gains will go to H and he will target L in the second round in accordance with the above strategies.

$$a \in [H + d + w, 3H + d - w)$$

$$E[u_L(\neg\text{join}_1)] = (1 - F_\epsilon(H + d + w))(3H + d - w)$$

$$E[u_L(\text{join}_{1D})] = F_\epsilon(H + d - w)(1 - F_\epsilon(H + d + w))(3H + d - w)$$

$$E[u_L(\text{join}_{1H})] = (1 - F_\epsilon(d + w - H))(1 - F_\epsilon(H + d + w))(3H + d - w)$$

Note that not joining dominates joining either coalition as  $F_\epsilon$  is a proper CDF

$$a > 3H + d - w$$

$E[u_L(\neg join_1)] = H - w$  as there will be no round 2 conflict

$E[u_L(join_{1D})] = F_\epsilon(H + d - w)(H - w)$  as there will be no round 2 conflict.

$E[u_L(join_{1H})] = (1 - F_\epsilon(d + w - H))(H - w)$  as there will be no round 2 conflict

Not joining dominates either coalition as  $F_\epsilon$  is a proper CDF.

## Targeting Behavior

**Proposition 2.A.** *If the dictator is strongly advantaged ( $d > H - w$ ) and there are two possible rounds of conflict, he will target  $H$  in the first round and  $L$  in the second for  $a \in (H + d + w, \frac{1}{2}(3H - w + \sqrt{4d^2 + (-3H + w)^2 + 4d(H + w)}))$*

*will target  $L$  and fight the coalition for*

*$a \in (\frac{1}{2}(3H - w + \sqrt{4d^2 + (-3H + w)^2 + 4d(H + w)}), 3H + d - w)$  and will not initiate conflict for  $a > 3H + d - w$ .*

**Proposition 2.B.** *If the dictator is weakly advantaged ( $d < H - w$ ) and there are two possible rounds of conflict, the dictator will target  $H$  in the first round and  $L$  in the second round for*

$$a \in (H + d + w, (\frac{1}{(2(3d+H+w))}((d+3H-w)(2d+H+w) - \sqrt{(d+3H-w)(16d^3 + (3H-w)(H+w)^2 + 16d^2(2H+w) + d(H+w)(17H+w)})))$$

*not initiate conflict for*

$$a \in ((\frac{1}{(2(3d+H+w))}((d+3H-w)(2d+H+w) - \sqrt{(d+3H-w)(16d^3 + (3H-w)(H+w)^2 + 16d^2(2H+w) + d(H+w)(17H+w)}))), 3H + d - w)$$

and target  $L$  and fight the elite coalition for  $a > 3H + d - w$

$$a \in [H + d + w, 2H + d) \text{ assume } d < H - w$$

$$E[u_D(H_1)] = F_\epsilon(d)F_\epsilon(H + d + w)(3H + d - w)$$

$$E[u_D(L_1)] = F_\epsilon(d + w - H)(3H + d - w)$$

$$E[u_D(\emptyset)] = F_\epsilon(d)(2H + d)$$

H alone is preferred to no target for all  $a < 3H + d - w$ , which is true in this range.

$$E[u_D(H_1)] - E[u_D(L_1)] = \frac{(3H+d-w)(-a^2+a(3H-w)+d(H+d+w))}{4a^2}$$

$$\frac{\partial}{\partial a} E[u_D(H_1)] - E[u_D(L_1)] = \frac{-(3H+d-w)(2d^2+a(3H-w)+2d(H+w))}{4a^3} \text{ which is always negative by definition}$$

of  $H, w, d, a$ . Thus the expression is monotonically decreasing everywhere.

$$\lim_{a \rightarrow H+d+w} E[u_D(H_1)] - E[u_D(L_1)] = \frac{(H-w)(3H+d-w)}{2(H+d+w)} \text{ which is always positive}$$

$$\lim_{a \rightarrow 2H+d} E[u_D(H_1)] - E[u_D(L_1)] = \frac{H(H-w)(3H+d-w)}{2(2H+d)^2} \text{ which is always positive.}$$

Therefore targeting H alone is preferred to L and the subsequent coalition in this range.

$$a \in [H + d + w, 3H + d - w) \text{ assume } d > H - w$$

$$E[u_D(H_1)] = F_\epsilon(d)F_\epsilon(H + d + w)(3H + d - w)$$

$$E[u_D(L_1)] = F_\epsilon(d + w - H)(3H + d - w)$$

$$E[u_D(\emptyset)] = F_\epsilon(d + w - H)(3H + d - w)$$

Compare H and the other option (target L and fight coalition now, or wait until next round and fight coalition).

$$E[u_D(H_1)] - E[u_D(L_1)] = \frac{(3H+d-w)(-a^2+a(3H-w)+d(H+d+w))}{4a^2}$$

$$\frac{\partial}{\partial a} E[u_D(H_1)] - E[u_D(L_1)] = \frac{-(3H+d-w)(2d^2+a(3H-w)+2d(H+w))}{4a^3} \text{ which is always negative by definition of } H, w, d, a. \text{ Thus the expression is monotonically decreasing everywhere.}$$

$$\lim_{a \rightarrow H+d+w} E[u_D(H_1)] - E[u_D(L_1)] = \frac{(H-w)(3H+d-w)}{2(H+d+w)} \text{ which is always positive}$$

$$\lim_{a \rightarrow 3H+d-w} E[u_D(H_1)] - E[u_D(L_1)] = \frac{d(w-H)}{2(3H+d-w)} \text{ which is always negative by definition of } H > w.$$

By the intermediate value theorem, there exists an  $a^{**} \in (H + d + w, 3H + d - w)$  at which



the dictator is indifferent all his options. For  $a < a^{**}$ , targeting H alone is preferred. For  $a > a^{**}$ , the dictator is indifferent between targeting L and fighting the coalition in this round or choosing no target in round 1 and fighting the coalition in round 2.

In terms of  $d$ ,

$$\lim_{d \rightarrow a-H-w} E[u_D(H_1)] - E[u_D(L_1)] = \frac{(a+2H-2w)(H-w)}{2a} \text{ which is always positive}$$

$$\lim_{d \rightarrow a+w-3H} E[u_D(H_1)] - E[u_D(L_1)] = -\frac{(H-w)(a-3H+w)}{2a} \text{ which is always negative}$$

By the intermediate value theorem, there exists a  $d^{**} \in (a+w-3H, a-H-w)$  at which the dictator is indifferent all his options. For  $d > d^{**}$  targeting H alone is preferred. For  $d < d^{**}$ , the dictator is indifferent between targeting L and fighting the coalition in this round or choosing no target in round 1 and fighting the coalition in round 2.

$$a \in [2H+d, 3H+d-w) \text{ and } d < H-w$$

$$E[u_D(H_1)] = F_\epsilon(d)F_\epsilon(H+d+w)(3H+d-w)$$

$$E[u_D(L_1)] = F_\epsilon(d+w-H)(3H+d-w)$$

$$E[u_D(\emptyset)] = H+d \text{ as there would be no conflict in round 2}$$

No target is preferred to L if  $a < 3H+d-w$ , which is true.

Compare No target and targeting H: no target preferred if  $E[u_D(\emptyset)] - E[u_D(H_1)] > 0$   
 $\frac{\partial}{\partial a} H+d - (F_\epsilon(d)F_\epsilon(H+d+w)(3H+d-w)) = \frac{(3H+d-w)(2d(H+d+w)+a(2d+H+w))}{4a^3}$  which is always positive and therefore monotone.

$$\lim_{a \rightarrow 2H+d} E[u_D(\emptyset)] - E[u_D(H_1)] = -\frac{(H+d)(H-w)(H+d+w)}{2(2H+d)^2} \text{ which is always negative.}$$

$$\lim_{a \rightarrow 3H+d-w} E[u_D(\emptyset)] - E[u_D(H_1)] = \frac{d(H-w)}{2(3H+d-w)} \text{ which is always positive.}$$

By the intermediate value theorem, there exists an  $\tilde{a} \in (2H+d, 3H+d-w)$  such that for all  $a > \tilde{a}$  no target is preferred to targeting H while H is preferred for  $a < \tilde{a}$

In terms of  $d$ ,

$$\lim_{d \rightarrow a-2H} E[u_D(\emptyset)] - E[u_D(H_1)] = -\frac{(a-H)(H-w)(a-H+w)}{2a^2} \text{ which is always negative.}$$

$\lim_{d \rightarrow a+w-3H} E[u_D(\emptyset)] - E[u_D(H_1)] = \frac{(H-w)(a+w-3H)}{2a}$  which is always positive.

By the intermediate value theorem, there exists an  $\tilde{d} \in (a+w-3H, a-2H)$  such that for all  $d < \tilde{d}$  no target is preferred to targeting H while H is preferred for  $d > \tilde{d}$

$$a > 3H + d - w$$

$$E[u_D(H_1)] = F_\epsilon(d)(2H + d) \text{ as there will be no conflict in round 2}$$

$$E[u_D(L_1)] = F_\epsilon(d+w-H)(3H + d - w)$$

$$E[u_D(\emptyset)] = F_\epsilon(d+w-H)(3H + d - w) \text{ if } d < H - w$$

$$E[u_D(\emptyset)] = H + d \text{ if } d > H - w$$

If the dictator is weak relative to the coalition, the dictator is indifferent between fighting L now (as the coalition will form) and not initiating conflict now and instead fighting the coalition in round 2

Targeting H is preferred  $a < 3H - w$ , however  $3H - w < 3H + d - w$ , therefore targeting H is dominated. Either no target or targeting L and fighting to coalition is preferred in this range, the dictator is indifferent between them.

If the dictator is strong relative to the coalition ( $d > H - w$ ), no target is preferred in this range and there will be no second round conflict.