Power Consolidation*

Mattias Polborn† Tigran Polborn‡

March 20, 2025

Abstract

We analyze a model of power consolidation in an authoritarian polity in which, initially, three agents share the spoils of office according to their power. Conflicts can be initiated either unilaterally or by a coalition of two agents; their outcome – removal of the losing side from power – is determined stochastically via a contest success function.

The model generates different interesting behaviors that are novel in the literature, namely: (i) Coalitions can form even though their members rationally foresee that they will turn on each other after a success; (ii) partial conflict, in which two agents fight each other while the third one stays out, may arise; (iii) stability of the initial power distribution is relatively rare and requires substantial power differences between all three agents, as well as a conflict technology that is neither too deterministic nor too random.

Key words: Autocracy; civil war; juntas
Word count: 10247

*We are very thankful to seminar audiences at the University of Zurich, Yale University, MPI Munich and especially Alexandre Debs, German Gieczewski, Kai Konrad, Marek Pycia and Milan Svolik.

†Vanderbilt University. E-mail: mattias.polborn@vanderbilt.edu

‡University of Southern California. E-mail: polborn@usc.edu.

1 Introduction

In any oligarchic group of individuals who share political power (such as a junta after a military coup; or a leadership committee after a popular revolution), the initial distribution of power among multiple individuals provides a semblance of collective decision-making and shared responsibility. However, as time progresses, conflicting interests and personal ambitions often lead to a struggle for supremacy, as members aim to increase their own power at the expense of others.

Understanding the process of power consolidation is important because dynamics of such power struggles often shape the course of nations and have far-reaching consequences, as the internal battles between individuals within the junta can result in political instability disrupting good governance, and the erosion of civil liberties. Additionally, understanding the process of power consolidation in juntas, in particular the identity of the player(s) most likely to initiate a power struggle, provides crucial insights into the nature of political alliances and the fragility of power-sharing arrangements.

To illustrate the types of situations that our model is designed to capture, consider the leadership of the Soviet Union after Lenin died in 1924. Formally, the leadership of party and government was collegially shared among the 7 members of the politburo, Stalin (S), Trotsky (T), Zinoviev (Z), Kamenev (K), Bukharin (B), Rykov (R) and Tomsky (To). Early on, S allied with K and Z against T, who was seen as the number-two man in the country and Lenin's heir apparent, and secured that he would not succeed Lenin outright, even though he remained a powerful player. In 1925/26, B, R and To allied with S after he broke with K and Z; they succeeded with expelling K and Z from the leadership of the Communist Party, with T not intervening. Soon thereafter, T was the next to be excluded from power, (expelled from the Soviet Union in 1929). Subsequently, S turned against B, R and To, expelling them from the politburo in 1929/1930. At this point, S was established as the clear leader of the Soviet Union, while everyone else mentioned was killed on S's orders over the following decade (or committed suicide just before they would have been killed).

Structurally, this episode displays several interesting features: (1) initially, no player is sufficiently powerful to marginalize everyone else; (2) coalitions form to exclude some opponents; (3) conflicts do not always include all active players at that stage, but rather, some players sometimes choose to stay out of a conflict; (4) after a coalition was successful, it may break apart, with its members fighting against each other; (5) the process unfolds over several stages, and some players must feel ex-post regret about their earlier actions.

Another historical example that displays some of these features in the Second Roman Triumvirate formed by Mark Antony (MA), Lepidus (L), and Octavian (O). Of these, L had clearly the smallest power base while MA and O appeared about equally powerful. After several years

of peace between them, O moved first against L in 36 BC, while MA did not intervene. After successfully excluding L, O moved against MA and emerged victoriously.

Our objective is to provide a model in which these features arise along the equilibrium path. A key difference to existing models is that the outcome of any conflict is stochastic in our model. This modification appears both realistic, and generates the rich dynamics just described.

Our work builds on the seminal work of Acemoglu et al. (2008, 2009) who consider the following model of intra-junta conflicts: Several members of the junta can form a "coalition" by mutual agreement. If this coalition is, in aggregate, more powerful than the opposing side (i.e., everyone outside the coalition), they can eliminate their opponents and remove them from power. After this, new coalitions could be formed until eventually either one player emerges as the last survivor, or there are no coalitions that can be formed with all coalition members willing to participate in an attack on the outsiders.

Crucially, all players in this game are rational and forward-looking; that is, no player will participate in an attack that would directly increase his power share if he foresees that his coalition partners would eventually turn against him and eliminate him in the future. Because there is always a coalition of players that would *initially* benefit from ousting their rivals, it is this property of looking forward further down the game tree that can ensure stability of an initial power distribution. In turn, if an initial power distribution is unstable, the property that players are forward-looking ensures that, whatever is the new stable power allocation, is reached in one step: Whoever would be eliminated in a second step is unwilling to join an attack coalition to begin with.

To understand Acemoglu, Egorov and Sonin's key stability argument, consider a junta of initially three members with different degrees of power, neither of whom has more power than the other two combined. While any of the three two-player coalitions that can be formed is able to eliminate the respective outsider, the coalition member who has less power than the other player in the coalition is justifiably nervous that, after the elimination of the outsider, his coalition partner will turn on him. Therefore, for the weaker prospective coalition partner, it is better to refuse to participate in rocking the boat and remain peaceful, rather than join in an attack that will be successful, but that ultimately would lead to his own elimination.

Interestingly, the lack of trust in potential partners is what keeps a triumvirate stable in their model.[1] This mutual interest in preventing one member to eliminate everyone else helps foster a sustainable governance structure where power is shared among the coalition members, reducing

[1] Juntas with more than three members can also be stable in their model. For example, consider a junta with 4 members with powers of 20, 11, 7 and 3. No coalition that includes the most powerful player can be stable (because, when any of the smaller players is eliminated, the most powerful player would eliminate everybody else). Furthermore, the coalition of the three small players against the biggest one is not stable either, because, subsequently, the 11-player would be able to eliminate the other two small players.

the likelihood of power concentration and preventing personal rule.

Our model builds on the same fundamental structure as Acemoglu et al. (2008, 2009), but focuses on settings with initially three players and departs in two key features. First, our model generalizes the function that determines the outcome in case of a power struggle. In their model, the bigger group wins with probability 1. In our model, if an attacking group of aggregate strength a is attacking a defending group with aggregate strength d, then the attacking group wins with probability $a^\alpha/(a^\alpha + d^\alpha)$, where $\alpha > 0$ is a given parameter. This Tullock contest success function has been analyzed in many articles of the literature on contests; see Konrad (2009) for a comprehensive review of this literature. Observe that the limit case of $\alpha \to \infty$ approximates the model of Acemoglu, Egorov, and Sonin, because then the larger group (almost) certainly wins.

Second, we allow for two ways to start a conflict. On the one hand, an attack coalition can form before the start of the conflict, if any member of such a coalition benefits in terms of their expected utility, relative to keeping the peace. On the other hand, we also allow for unilateral attack of one player against another one; in this case, the third player has the choice of joining either side, or waiting out the conflict between the two others.

We show the following main results. Once fighting erupts, power will eventually always consolidate into one person because juntas of size 2 are generically unstable for all $\alpha \neq 1$.[2]

The main case of interest is where the two weaker players together are stronger than the strongest player, focusing on the case that $\alpha > 1$. We show that there are two threats to stability, a coalition between the two weaker players, and a coalition between the two stronger ones.[3] For either type of coalition, there is a critical value α^* such that, for $\alpha < \alpha^*$, an attack coalition can form that improves the expected utility of all its members, while this is not possible if $\alpha > \alpha^*$. The critical value α^* depends on the initial distribution of power. In particular, the closer to equal the initial distribution of power is among the three players, the larger is α^* (i.e., the larger is the instability with respect to attack coalitions).

The key insight from allowing for unilateral attacks to start a conflict is that they may occur on the equilibrium path, with the third player either staying out, or joining the attacker. For example, by initiating a conflict against the most powerful player, the middle player can sometimes "force" the weakest player to join with him. This case arises for high values of α, when the weakest player would like to remain at peace (so, is unwilling to enter an explicit coalition with the middle player), but given that a conflict is initiated, prefers to join the middle player in order to avoid having to

[2] In knife-edge case that $\alpha = 1$, all players are always indifferent between fighting (in any coalition) and remaining peaceful.

[3] A coalition between the strongest and the weakest is possible, but only when both other coalitions are also feasible; in this sense, a coalition between the strongest and the weakest is not an additional threat to triumvirate stability.

fight the strongest player (which is the most likely outcome if he instead chooses to sit out the initial conflict). The weakest player's willingness to join makes it attractive for the middle player to launch a unilateral attack, and he is the most likely to emerge as the final winner.[4]

Overall, we show that, for most values of α, all initial triumvirate power distributions are unstable. As such, our model is consistent with the empirical observation that stable power distributions among a group of equals appear very rare (Tullock, 1987). However, we identify some intermediate values of α for which stable power distributions exist.

While our model focuses on intrastate conflict among the members of a ruling coalition, these same phenomena (coalition formation and breakup; partial conflict; unilateral conflict initiation) arise in inter-state wars with many potential participants. For example, one interpretation of the First World War is that the decision to start hostilities was made unilaterally by actors (Serbia, Austria) who preferred to start a war in order to involve other actors (Russia, Germany) in a war that they would enter once fighting started, but were not particularly keen on initiating if peace was the outside option. Similarly, there are historical examples of war in which a successful coalition breaks up to fight among themselves (e.g., German-Russian alliance at the start of World War II). Our model is not meant to directly address these questions of interstate conflict, but we hope that the methods we develop in this paper will be useful for future research in this area.

The paper proceeds as follows. We discuss related literature in the next section. The model is presented in Section 3, and the analysis follows in Sections 4 and 5. The final section discusses the results and concludes. All proofs of propositions are in the Appendix.

2 Related literature

The objective of our paper is to analyze how the power allocation in an "anarchical" society (i.e., one in which this question is not institutionally regulated) evolves over time when self-interested players can decide whether to try to exclude others from power in order to increase their own stake. As already mentioned in the introduction, Acemoglu et al. (2008) analyze a closely-related model in which any sufficiently large coalition can squeeze out the remaining agents, and where agents cannot commit to not attack their current partners in a future round. All conflicts in their model have a deterministic result, and any power consolidation (along the equilibrium path) occurs immediately. In contrast, conflict results are stochastic in our model that gives rise to a much wider spectrum of behavior.

[4]Interestingly, after Lenin's death in 1924, few observers would have picked Stalin, an ethnic Georgian with a heavy accent in Russian, as the politburo's most powerful member.

Other important models of anarchic power redistribution include Niou and Ordeshook (1990) and Jordan (2006). In these models, agents control resources and can form coalitions that can, if sufficiently powerful, redistribute the outsiders' resources among themselves. They allow for partial redistribution of a victim's assets so that, in contrast to Acemoglu et al. (2008, 2009) and our model, a conflict in Niou and Ordeshook (1990) and Jordan (2006) does not necessarily lead to the complete removal of the attacked. Furthermore, an attack coalition also has to decide how to redistribute the assets they gain between their members. In the Niou and Ordeshook (1990) model, stability requires that one agent controls exactly half of the resources.[5] In contrast to our model, conflict outcomes are deterministic in these models, and all conflicts are started by coalition agreements and involve all agents, either on the attacking or the defending side.

Conflict initiation is also a central topic in the formal international relations literature on war. Fearon (1995) and Powell (2006) provide models of conflict initiation between two countries that are based on expected shifts in their relative power. In particular, armed conflict is initiated by a strong player who fears an opponent's relative strengthening if peace prevails. In contrast, in our model, relative power is constant, so this is not what is driving conflict in our model. Furthermore, while conflict in two-player subgames is always initiated by the stronger player, in the initial three-player setting, conflict may be initiated by any player along the equilibrium path.

Our work is peripherally related to an important literature on the effects of threats of violence among an autocratic elite. For instance, Acemoglu et al. (2004), Konrad and Skaperdas (2007), Boix and Svolik (2013), and Persico (2021) analyze the problem of an autocrat who needs to maintain his rule by redistributing resources to prevent an attempt to overthrow him. Additionally, Debs (2016) deals with the power succession problems faced by military dictators. While conflict avoidance, often by paying off potential challengers, is a key consideration in this literature, players in our model cannot peacefully pay off their opponents in order to avoid conflict. Instead, the focus of our model lies on how players ally with and against other players.

There is some political economy literature that analyzes the dynamics of power and policy in formalized institutions where policy is set through voting (e.g., Messner and Polborn (2004); Roberts (2015); Gieczewski (2021)). While agents in these papers often try to affect power (and hence policy) in the future, they are limited, at least to some extant, by the existing institutions; for example, while they can enlarge the set of voters or change the majority rules, they are bound to have future decisions made through elections. In contrast, there are no institutional constraints in our paper – whoever wins in an open conflict can simply remove the losing side from further participation.

[5] An interesting application of a similar idea in the Acemoglu et al. (2008) framework is Fan (2019), which extends the model by allowing agents to destroy their own power, in order to be able to credibly commit to alliances.

Our model is also related to the literature on coalition formation (see Ray and Vohra (2015) for a recent review). However, the key points of interests of this literature are which coalition forms and how its benefits are distributed among the partners. In our model, players cannot make binding agreements on how to redistribute a vanquished enemy's or their own power among themselves. Furthermore, we do not impose any particular game form, and therefore cannot say which particular coalition will form among all those that are feasible; rather, we focus on the question for which parameters any given coalition is, in expectation, beneficial for all of its members.

Finally, there are several papers analyzing three-person conflicts known as "truels" (see, e.g., Kilgour and Brams (1997)) where each player is characterized by a target hitting probability, and players move in an exogenously given order. When it is his turn to move, a player decides which opponent (if any) to target. When all players are sufficiently good marksmen, triumvirates are relatively stable in truel models because a player who eliminates an opponent would set himself up as a target for the third opponent. Relative to the truel literature, our paper admits more actions than just a unilaterally attack (e.g., teaming up with one of the other players to take on the third), and there is no period of extreme vulnerability for a player who initiated an attack.

3 The Model

Consider the following model of authoritarian power politics. Initially, there are three players denoted X, Y and Z, characterized by their *power* x, y and z, such that $x \leq y \leq z$. Power is a positive real number between 0 and 1, and is normalized such that $x + y + z = 1$.

Power is both beneficial in peace and in power struggles, in the following way. If the players do not fight each other, each receives a share of the overall benefit from ruling equal to their respective power. The overall benefit from ruling is normalized to 1, so that the peace payoffs are x, y and z, respectively. Likewise, if, for example, X and Y successfully eliminate Z and then share power, X's payoff is $\frac{x}{x+y}$ and Y's payoff is $\frac{y}{x+y}$.

Players can initiate fights among each other, and the losing player(s) are removed before the distribution of benefits. There are two possible types of conflicts, two-sided conflicts and three-sided conflicts.

Two-sided conflicts. When a player initiates a fight against one other player, the third player can join either side of the fight, or stay neutral and not join the conflict. The attacking party wins the fight with probability $\frac{a^\alpha}{a^\alpha+d^\alpha}$, where α is a known constant, a is the sum of the powers of all attacking players, and d is the sum of the powers of all defending players; with the complementary

probability, the defender (or defending coalition) wins. All members of the losing side lose their power and are removed from the game without any benefit given to them.

If, say, a coalition of X and Y attacks Z and is successful, then each surviving player can choose whether to turn on his former coalition partner and attack him. This battle follows the same rules in terms of winning probabilities, and the ultimate survivor of this battle receives a payoff of 1. Of course, if both survivors of the first battle in this example choose to remain peaceful, then they simply share the overall payoff in proportion to their respective power.

Three-sided fights. If a player attacks both other players, they can either form a coalition and fight with their combined power (as described above), or each can remain separate, in which case X's probability of winning is $\frac{x^\alpha}{x^\alpha+y^\alpha+z^\alpha}$, and analogously for the other two players. The losers are eliminated, and the winner of the three-sided fight gets the entire payoff.

Stability concept. We will look at the question of stability of a triumvirate without imposing any particular sequence of play, by focusing on the incentives of a player who has the option to start conflict under the following two assumptions: First, the potential initiator has the choice to start conflict, or to remain at peace for the foreseeable future; that is, when a potential initiator chooses not to attack, then he (and everyone else) expects to receive their respective peace payoffs. Likewise, if an initiator proposes to one of the other players to form a coalition against the third player, and the partner decides on whether to accept the proposal based on whether his expected payoff from joining the coalition is larger than the peace payoff.

Second, once conflict has commenced, all players rationally continue to play without being bound by any agreement beyond the current round. In particular, no player can credibly promise to never attack another player after the current round, for example, in order to persuade that player to enter into a coalition with him, or to persuade a player to keep out of a conflict. Likewise, side-payments to avoid conflict are not feasible.

The first of these assumptions – namely that the initiator and the potential coalition partner compare their expected payoff to what they would get in peace – can be interpreted as a behavioral assumption, or as completely rational behavior in a game where there is exactly one opportunity to start a conflict.[6]

The advantage of not imposing a particular, more involved, sequence in which players move

[6]That is, initially, one player is randomly recognized and gets to choose whether to remain at peace, or start a conflict (either by proposing a coalition to another player, or through a unilateral attack). If the proposer chooses to remain at peace, or if the potential coalition partner rejects the proposed coalition, each player receives his peace payoff.

and decide whether or not to start a conflict is that there are many conceivable protocols. Furthermore, within each protocol, decisions on whether to attack (or join an attacker) depend on the players' expectations of what would happen off the equilibrium path, raising the possibility of multiple equilibria even for a given sequence of moves. For example, whether a particular player is willing to initiate an attack depends on whether they expect that, even if they chose peace, they would subsequently be attacked by some other players. This makes the analysis under any particular game protocol quite involved.

We will, however, show, in Proposition 14, that the set of parameters for which a triumvirate is stable under our approach is an upper limit (in the sense of set inclusion) of the set of parameters for which a triumvirate is stable under *any* given game protocol.

Discussion. A key parameter in our model is the fighting technology parameter α. Intuitively, $\alpha > 1$ describes a situation in which there are economies of scale in fighting, in the sense that any coalition of players, when fighting united, has a larger probability of winning than the sum of the winning probabilities of the members when these are fighting separately. While we will cover both cases – that is, $\alpha < 1$ and $\alpha \geq 1$ – in the following, casual observation suggests that $\alpha \geq 1$ may be the more realistic and relevant one.

Note that our model assumes that players are not risk averse, and that there are no costs of war in the sense that the aggregate value of winning the prize through conflict is smaller than the aggregate value if all agents remain peaceful. We do this both to keep the model tractable and clean, and to keep it comparable to Acemoglu et al. (2008, 2009) who do not consider risk aversion or cost of war either in their model.

Both risk-aversion and war-related partial destruction of the prize are intuitively likely to have stabilizing effects. For example, for a risk-averse agent, getting 45 percent of the spoils for sure may very well be preferable to having a 60 percent probability of getting it all (and getting nothing with the complementary probability). The effects of a cost of war (i.e., the effective prize of winning a civil war is less than 1) would be formally similar to those of risk aversion.

However, in Section 6, we discuss the effects of a cost of war. While costs of war do potentially pacify settings with two remaining players, their effect on settings with three players is ambiguous. This is because, by potentially stabilizing subgames with two remaining players, they potentially expand the set of mutually-beneficial attack coalitions.

4 Preliminary Analysis

4.1 Two player subgames are always unstable

We start the analysis by considering the situation after an initial round of two-sided conflict that was won by the coalition team.[7] Proposition 1 below shows that this situation is generically unstable, in the sense that (for any $\alpha \neq 1$) there is always one player who strictly prefers to attack the other.

Proposition 1. *Consider a subgame where only two players remain active. For any $\alpha > 1$, the more powerful player chooses to attack the less powerful player. For any $\alpha < 1$, the less powerful player chooses to attack the more powerful player. For $\alpha = 1$, both players are indifferent between attacking or staying at peace.*

Intuitively, with only two remaining players, all *strategic* effects on behavior after the current round are irrelevant, and only direct effects matter. In our model, strength (x, y, z) influences two different potential payoffs, the peace payoff and the conflict payoff. Whenever there is any difference between the two, one of the two players will initiate a fight.

4.2 The case of $\alpha < 1$

In the following Proposition 2, we show that a triumvirate is always unstable in case that $\alpha < 1$. Specifically, the weakest player always prefers to attack his two opponents over peace because, for $\alpha < 1$, his chances of winning are larger than his share of the prize if everyone remains at peace.

Proposition 2. *Assume that $\alpha < 1$.*

1. *If player X attacks both of his opponents Y and Z, then they will fight individually (i.e., the fight will be a three-way fight).*
2. *X would rather fight Y and Z than stay at peace.*

Note that it may also be the case that Y would attack X and Z (in the absence of an attack by X). This will clearly be the case if y is sufficiently close to x.

5 The main case: $\alpha > 1$

From now on, we focus exclusively on the case that $\alpha > 1$. Observe that Proposition 1 shows that any participant in a coalition fighting the third player needs to be aware that, once that fight is won,

[7] Clearly, if the coalition was defeated, the game ends and all power is consolidated in the single player.

he will have to fight his former partner. Therefore, when deciding whether to join a coalition, a player always needs to calculate the complete expected utility from both rounds of fighting.

For example, when X joins an attack coalition with Y against Z, X's overall expected utility is

$$\frac{(x+y)^\alpha}{(x+y)^\alpha + z^\alpha} \cdot \frac{x^\alpha}{x^\alpha + y^\alpha}. \tag{1}$$

The first fraction is the probability that the coalition of X and Y is victorious against Z, and the second fraction then is the probability that X wins the next round against his former partner Y. As explained in the model section, we assume that X compares this expected payoff as a coalition member with the payoff from peace.

5.1 Defensive coalitions

The first step in our analysis is to consider defensive coalitions, at the stage where all three players are active. The following Proposition 3 shows that, if one player (A) attacks both other players, D_1 and D_2 prefer to fight together. Intuitively, the two defenders are better off joining forces, because of the economy of scales logic when $\alpha > 1$.

Proposition 3. *Assume $\alpha > 1$, and suppose that player $A \in \{X, Y, Z\}$ attacks both other players (denoted D_1 and D_2). Then, D_1 and D_2 will form a coalition to fight against A.*

5.2 When one player is stronger than both opponents combined

We now analyze how different initial power distributions translate into conflict or stability. The first scenario is when one of the players is stronger than the other two combined.

Proposition 4. *Suppose $\alpha > 1$ and $z > 0.5$. Then Z prefers attacking X and Y to peace.*

Proposition 4 shows that, for an initial power distribution with one agent who is stronger than the two other players combined, the strongest player has an incentive to break the peace. As we will see below, this behavior contrasts markedly with the case that $z < 1/2$.

While Proposition 4 shows that triumvirates cannot be stable if the strongest player's strength exceeds the sum of the two other ones, there are two questions that the proposition does not speak to. First, is there ever a chance for Z to do better than immediately attack both opponents? For example, it could be that Z is better off teaming up with one of the two opponents first, or to attack only one (if the other player is willing to sit out the first fight). Second, is there also another player

(rather than Z) who would also initiate a fight? Intuitively, this can only be the case for Y, because X's chances to win an all-out fight are minimal.[8]

5.3 A non-dominant strongest player

We now turn to the case that $z < 1/2$, i.e., no player is stronger than both other players combined. Observe that, under this restriction, any triumvirate in which the two weaker players do not have exactly equal strength is stable under the contest success function of Acemoglu et al. (2008).[9] The following Proposition 5 shows that Z, in this case, does not want to start any fight in which he would eventually face a coalition of the two other players.

Proposition 5. *If $z < 1/2$ and $\alpha > 1$, Z never wants to fight a coalition of both other players.*

Going through the same steps as in the proof of Proposition 5, it is easy to show, for $\alpha > 1$ that no other player wants to attack both other players combined.

This leaves three potential scenarios for conflict. First, it could be that two players agree to gang up on the third; in this case, we need to check whether both coalition partners' expected utility from such cooperation is higher than if they stay peaceful.

Second, it could be that conflict is initiated by only one player against an opponent, but that the attacker is joined by the third player. In this case, the original attacker needs to have a higher utility (conditional on the third player joining him) than when staying peaceful, and the third player's utility when joining must be higher than when sitting out the first round of conflict.

Third, it could be that conflict is initiated by one player against an opponent, and the third player sits out the first round of conflict. In this case, the attacker's expected utility (after both rounds of conflict) must be higher than when staying at peace, and the third player's utility must be larger when sitting out the first round than when joining in (on either side).

[8] Because Z strictly prefers to fight and the game is zero-sum, we know that X and Y's aggregate payoff from a fight is negative. Thus, the expected payoff from a fight must be negative for at least one of them.

[9] In the non-generic case that the two weaker players are exactly equally strong, then a triumvirate is unstable even under the contest success function of Acemoglu et al. (2008). To see this, consider an example where $x = y = 0.3$ and $z = 0.4$. In the first round, X and Y will unite to take down Z; then, since X and Y are equally strong, they are indifferent between peace (where each of them gets one-half of the overall benefits), and fighting against each other, where each wins with probability $1/2$. In either case, X and Y get an expected payoff of 0.5, which is more than 0.3, which is what they would get when keeping the peace with Z.

5.4 Attack coalitions

When two partners team up against the third player, there are three different possible coalitions: X and Y (against Z); X and Z (against Y); and Y and Z (against X).

The following Proposition 6 considers a coalition between the two weak players X and Y. We show first that the stronger partner, Y, is always interested in attacking Z with X's support. In contrast, the weaker partner X has to be worried about being attacked by his partner after a successful conclusion of the first round, and is the constraining factor. Since, if $x < y$, X's disadvantage in the second round is increasing in α, an XY–coalition is more likely to be stable when α is small.

Proposition 6. *Assume $\alpha > 1$, $x < y < z$ and $z < x + y$.*

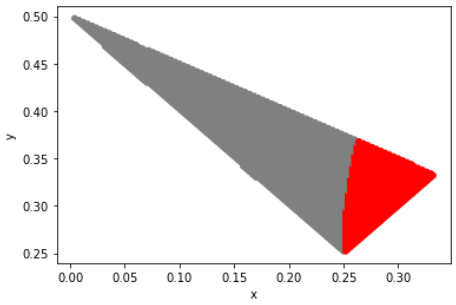
1. *Y's utility from entering a coalition with X to attack Z always exceeds his utility from remaining peaceful.*

2. *Consider the set $C_{x,y}(\alpha)$ of initial power distributions such that X prefers to form an attack coalition with Y to peace, as a function of α. Then, $C_{x,y}(\alpha)$ is weakly decreasing in the sense of set inclusion in α: If $\alpha_1 < \alpha_2$, then $C_{x,y}(\alpha_2) \subseteq C_{x,y}(\alpha_1)$.*

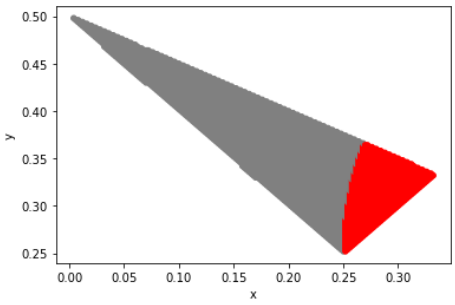
Figure 1 illustrates the second part of Proposition 6 and shows that the set of (x, y) constellations for which X is willing to join with Y for an attack on Z, for different values of α (the entire red-gray triangle consists of all points where $x \leq y \leq z = 1 - x - y$). The set of power distributions for which X is willing to join Y is shrinking in α and almost entirely disappears for very high values of α (except along the $x = y$ line). This is intuitive because, as $\alpha \to \infty$, Y will win the second round of fighting against his previous partner X (as long as $y > x$); therefore, X knows that if he joins Y to successfully gang up on Z, he is essentially signing his own death warrant.

On the other hand, if $x \approx y$ (the south-east side of the parameter space triangle in Figure 1), then an $XY-$coalition is stable for any finite α because then X receives approximately the same expected payoff as Y in the second round, and we know that the joint payoff of X and Y in case of an attack is always higher than their joint peace payoffs. This follows from Proposition 5 that shows that Z's expected utility is strictly lower when fighting against X and Y, and the fact that the game is a zero-sum game.

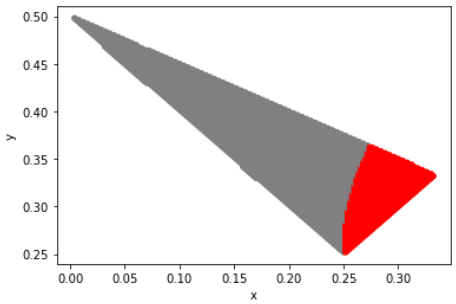
We now turn to the two other potential coalitions, which both include the strongest player, Z. It is easy to show that Z would always, if joined by a coalition partner, like to attack the third player. Again, the position of that potential coalition partner is more precarious as he knows that Z would turn against him after a first-round victory. This concern complicates Z's search for a coalition partner.



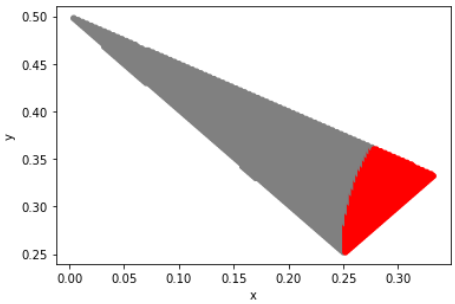
(a) $\alpha = 1.01$



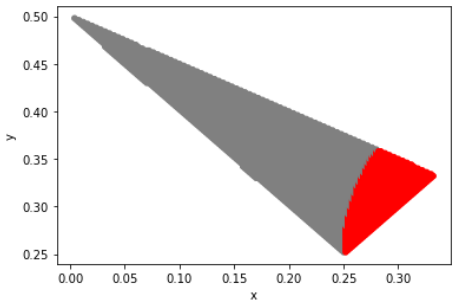
(b) $\alpha = 1.5$



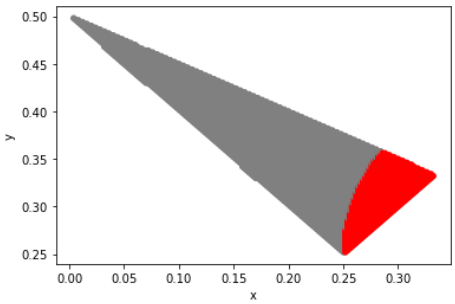
(c) $\alpha = 2$



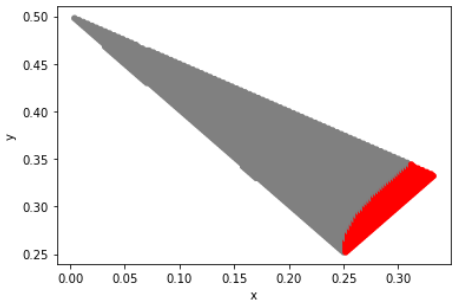
(d) $\alpha = 2.2$



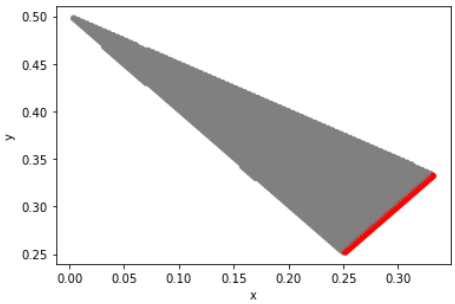
(e) $\alpha = 2.6$



(f) $\alpha = 3$



(g) $\alpha = 8$



(h) $\alpha = 100$

Figure 1: Red - X and Y prefer teaming up against Z to staying at peace.

Proposition 7. *1. Z is always willing to join with either X or Y into an agreement to jointly attack the third player.*

2. *X's utility from joining with Z against Y is weakly lower than X's utility from joining with Y against Z.*

3. *Y's utility from joining with Z against X is weakly lower than Y's utility from joining with X against Z.*

An implication of Proposition 7 is that Z-led coalitions are not the primary threat to the stability of a triumvirate: While Z would always be keen to have the support of one of his colleagues to take out the third one, he either cannot find a partner who is willing to join him (because the partner knows that Z would attack him afterwards, which makes the initial proposition unattractive), or, if there is a partner willing to join up with Z, that partner would be at least as eager to join up with the other small player against Z.

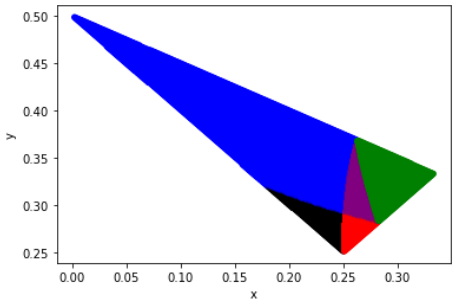
Just like Proposition 6, the following Proposition 8 shows that a $Y-Z-$coalition and a $X-Z-$coalition is more likely to be mutually beneficial for low values of α. For α too large, Z's advantage in the second round is too large so that his potential coalition partner's overall expected utility is eventually less than their utility from remaining at peace.

Proposition 8. *The set of cases in which coalitions between Y and Z, or between X and Z are stable, respectively, is decreasing in α.*

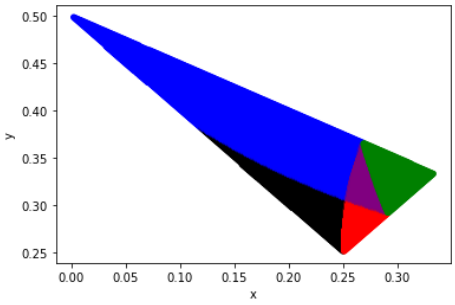
Figure 2 illustrates all stable attack coalitions. In the green area (corresponding to an initial power distribution with three approximately equally strong players), all three types of coalitions are stable; in general, this is the only region in which an $X-Z-$coalition is stable. In the red region, only an $X-Y-$coalition is stable, and in the blue region, only a $Y-Z-$coalition is stable. The purple region is the region where both an $X-Y-$coalition and a $Y-Z-$coalition are stable. Finally, the black area corresponds to regions in which all attack coalitions are unstable. Alternatively, going by the different types of coalition, an $X-Y-$coalition is stable in the red, purple and green regions; a $Y-Z-$coalition is stable in the blue, purple and green regions; finally, an $X-Z-$coalition is stable in the green area.

5.5 Stability against unilateral attacks

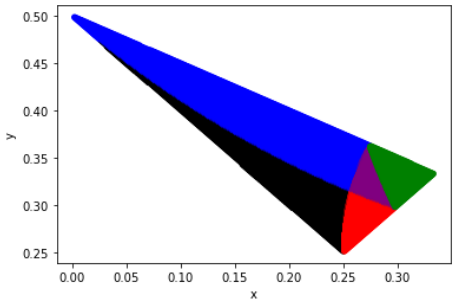
We now analyze the possibility of unilateral attacks against one opponent. Our first result is that, if the three players are approximately equally strong, none wants to unilaterally attack one of the other players.



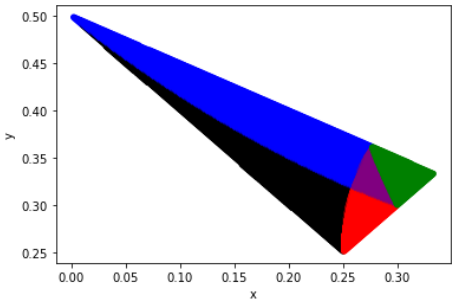
(a) $\alpha = 1.01$



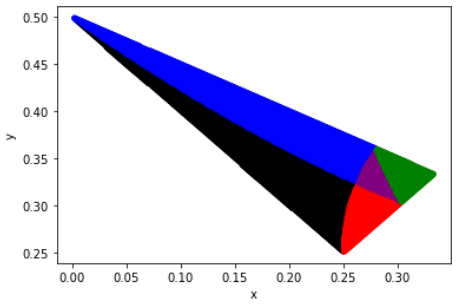
(b) $\alpha = 1.5$



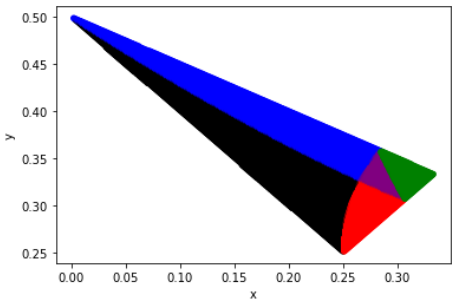
(c) $\alpha = 2$



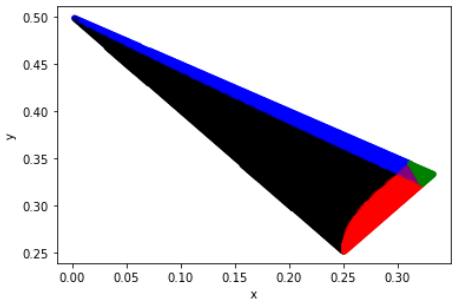
(d) $\alpha = 2.2$



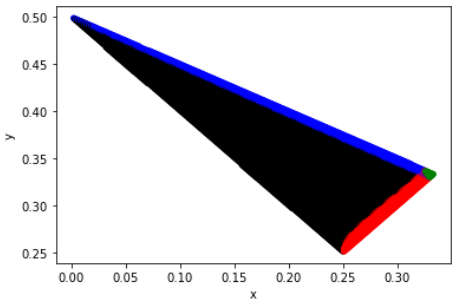
(e) $\alpha = 2.6$



(f) $\alpha = 3$



(g) $\alpha = 8$



(h) $\alpha = 100$

Figure 2: Black - There are no mutually beneficial attack coalitions. Red – Only $X - Y-$coalition is stable. Blue – Only $Y - Z-$coalition is stable. Purple – Both $X - Y-$coalition and $Y - Z-$coalition are stable. Green – All three possible coalitions are stable.

Proposition 9. *For any $\alpha < \infty$, there exists $\varepsilon^*(\alpha) > 0$ such that, if $z \geq y \geq x \geq z - \varepsilon$, then no player chooses to unilaterally attack one of the other players.*

Observe, though, that cases of approximately equal strength are cases where two players will form a coalition against the third one; in Figures 1 and 2, these are the points close to the eastern-most corner of the parameter triangle. Thus, Proposition 9 implies that, in cases of approximately equal strength, the resulting conflict will uniquely be of the type that two players form a coalition and agree to gang up on the third.

For the case of significantly different players, we now analyze, when one agent attacks only one of his opponents, the incentives of the third player. In this case, the third player needs to decide which side, if any, to join. The first part of the following Proposition 10 shows that the third player will never join the stronger one of the other two. Intuitively, joining the stronger player is dominated by just sitting out the first round: If joining the stronger player in first round fighting, one could get killed in that round. Furthermore, if successful in the first round, one faces the stronger player for sure, while, if sitting out the first round, there is some probability that the weaker player wins an upset victory and the ultimate fight is then easier.

The second part derives a necessary and sufficient condition for the third player to join the fight on the side of the weaker contestant. Furthermore, it shows that this condition is never satisfied for sufficiently weak players, i.e., these always sit out a fight between the two large players. The third part then shows that a higher α enlarges the set of parameters for which the weak player joins the fight.

Proposition 10. *Suppose that player $A \in \{X, Y, Z\}$ attacks player $B \neq A$ in the first round. Let s denote the strength of the stronger of these players, and let w denote the other player's strength.*

1. *The third player, C will either join the weaker player of A and B, or sit out the fight without joining in, but will never join the stronger player. If $s = w$ (i.e., both initial fighters have the same strength), then C sits out their fight.*

2. *The necessary and sufficient condition for C to join the weaker of the two players is given by*

$$\frac{(w+c)^\alpha - w^\alpha}{(w+c)^\alpha + s^\alpha} \geq 2sw\frac{w^\alpha + c^\alpha}{s^\alpha + c^\alpha}$$

   This condition is always violated at $c = 0$: Sufficiently weak players sit out a fight between the two other players.

3. *If, for a given power distribution, C joins the weaker player for a particular value of $\hat{\alpha}$, then C also joins the weaker player for all higher α.*

The second part of Proposition 10 shows that sufficiently weak players better stay out of the fight between two heavyweights. Intuitively, by sitting out the first fight, the weak player only needs to survive one fight, albeit more likely against the strong players.

In contrast, the third part shows that C is less likely to sit out a fight for higher α. Intuitively, a higher α reduces C's risk in the first fight (in which $w + c > s$), while also making it (relatively) more attractive to fight against the weaker rather than stronger player in the second round.

Proposition 11 considers Y's incentive to unilaterally attack Z. Why might this be an attractive course of action for Y? Remember that Y is always happy to attack Z if he receives X's support (see Proposition 6). The crucial question therefore is whether, by starting a fight with Z, Y can induce X to join him, possibly even in cases in which X would rather remain at peace. With a unilateral attack, peace is not anymore an option that X can secure because whoever wins the fight between Y and Z will then fight X. Intuitively, X prefers to fight Y rather than Z in the final round, and that makes it attractive to join Y.

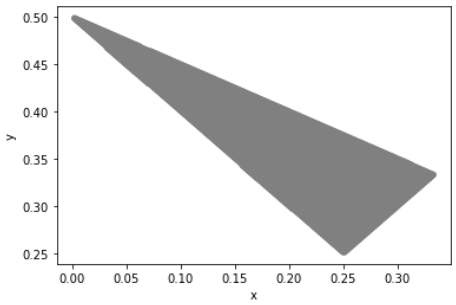
Proposition 11. *For any initial power distribution where $x < y < z < 1/2$ and $x + y > z$, if α is sufficiently large, then Y unilaterally attacks Z and is joined by X.*

While Proposition 6 showed that an increase in α increased the stability against attack coalitions, Proposition 11 shows that it also decreases stability against *unilateral attacks*.

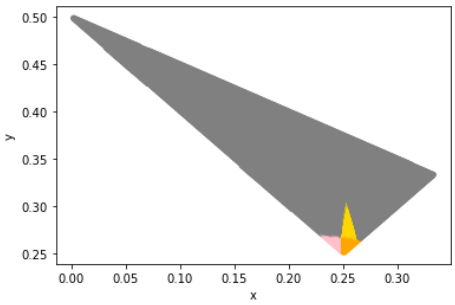
Figure 3 illustrates Proposition 11. In the pink and orange areas, X is willing to join a unilateral attack by Y on Z. For low values of α (about $1 \leq \alpha \leq 2$), there are no visible areas in which X is willing to join Y in a potential attack on Z. Beyond this level of α, the set of parameters where such a unilateral attack by Y is supported by X grows rapidly, taking up most of the parameter space by $\alpha = 8$ and essentially all of it by $\alpha = 20$. In particular, a comparison with Figure 2 for the cases $\alpha \in \{2.2, 2.6, 3\}$ shows a rapid increase of the unilateral attack area while the parameter space that supports XY-coalitions shrinks rather slowly.

In addition, the gold and orange areas in Figure 3 show the sets of parameters where X prefers a unilateral attack on Z over peace, and Y prefers joining in on that attack over sitting it out. Specifically, in the gold area, only a unilateral attack by X can occur, while in the orange region both unilateral attacks (i.e., X on Z and Y on Z) are profitable for the attacker and are joined by the third player.

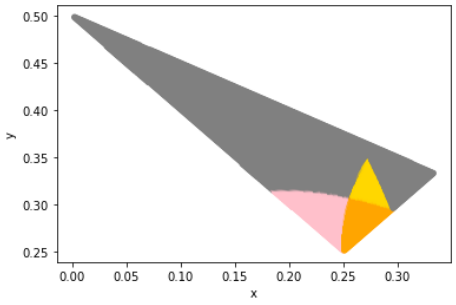
Observe that, along the equilibrium path, X's and Y's payoffs are exactly the same as in a XY-coalition areas (red and purple) in Figure 2. However, for the XY-coalition, the outside option is "staying at peace," while here, the initially uninvolved party instead has the option to sit out the first round. This may be better or worse than the "staying at peace" option. For example, there are some areas in which X's expected payoff from joining a Y-attack on Z is less than the peace payoff



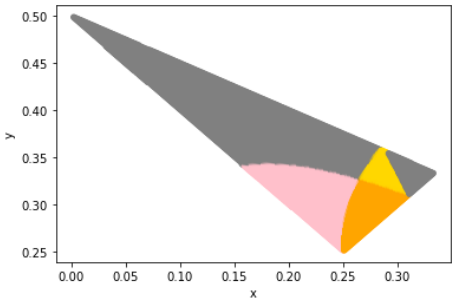
(a) $\alpha \in \{1.01, 1.5, 2\}$



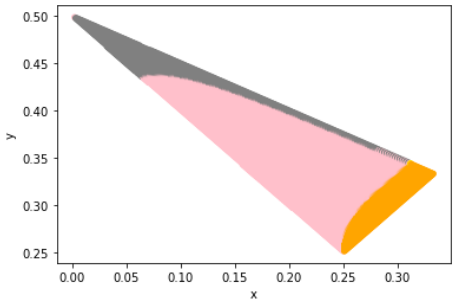
(b) $\alpha = 2.2$



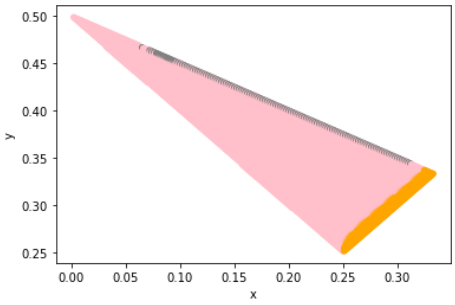
(c) $\alpha = 2.6$



(d) $\alpha = 3$



(e) $\alpha = 8$



(f) $\alpha = 20$

Figure 3: Unilateral attacks on Z by either Y or X:
Pink – Unilateral Y–attack on Z, with X joining in on Y's side.
Gold – Unilateral X–attack on Z, with Y joining in on X's side.
Orange – Both conditions are true

x, but is more than the payoff from sitting out the first round; thus, this case generates additional instability.

On the other hand, when all players are almost equally strong, an XY-coalition is stable for any α, but if Y were to attack unilaterally, it would be more attractive for X to sit out the first round because the strength of his second round opponent does not change very much whether he faces Y or Z, and sitting out the first round eliminates the risk of getting killed in that round.[10]

We now analyze other potential unilateral attacks. Remember that, if Z faces X and Y as unified opposition, his expected utility is lower than when remaining at peace. Therefore, a necessary condition for Z to unilaterally attack, say, X is that Y prefers to stay out of such a fight.

Proposition 12. *1. For Z to be willing to unilaterally attack Y, the following conditions are necessary and sufficient:*

$$\frac{z^\alpha}{z^\alpha + x^\alpha}\frac{z^\alpha}{z^\alpha + y^\alpha} > z \tag{2}$$

$$\frac{(x+y)^\alpha - y^\alpha}{(x+y)^\alpha + z^\alpha} < 2yz\frac{x^\alpha + y^\alpha}{z^\alpha + x^\alpha} \tag{3}$$

2. For Z to be willing to unilaterally attack X, the necessary conditions are (2) *and*

$$\frac{(x+y)^\alpha - x^\alpha}{(x+y)^\alpha + z^\alpha} < 2xz\frac{x^\alpha + y^\alpha}{z^\alpha + y^\alpha}. \tag{4}$$

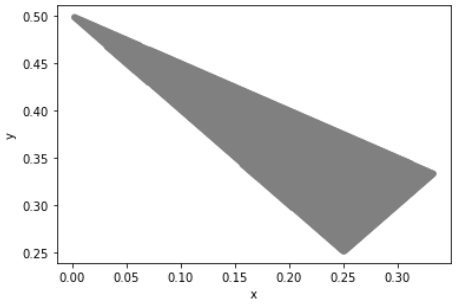
The set of parameters where Z is willing to unilaterally attack X is a strict subset of the set of parameters where Z is willing to unilaterally attack Y.

Figure 4 illustrates Proposition 12. In both colored regions (red and purple), Z's expected utility when attacking X increases relative to peace, and Y chooses to stay out of the first round fight. In the purple region, Z's utility would also increase from attacking Y (and X stays out of the first round). Observe that both sequences (first X and then Y, or the other way around) give the same utility for Z as a victory for Z involves fighting the same opponents, just in a different sequence. Thus, in the purple area, Z is indifferent between attacking X and attacking Y.

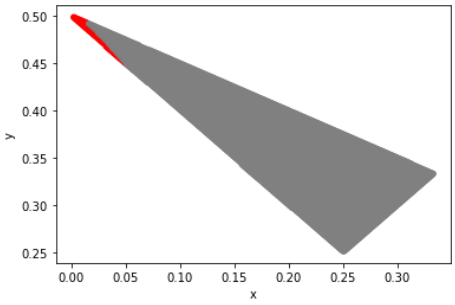
Intuitively, the red area is much larger than the purple one because Y is more willing to stay out of the first round than X, as Y has better chances to survive in the second round against Z than the weaker X has.

Interestingly, in contrast to the other pictures, the regions in Figure 4 do not monotonically grow or shrink in α, but rather move through the parameter space. Also, a unilateral attack by Z

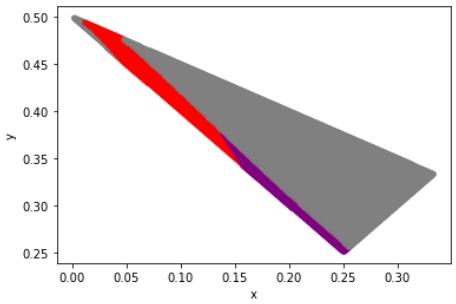
[10] In the figures for $\alpha = 8$ or 20, the gray area around $(x, y) = (1/3, 1/3)$ cannot be seen because it is too small.



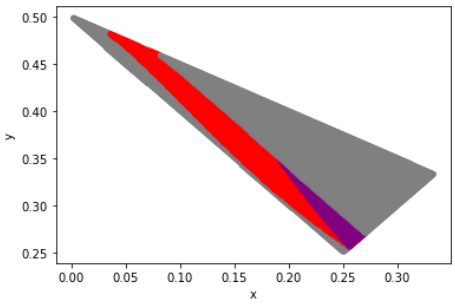
(a) $\alpha = 1.01$



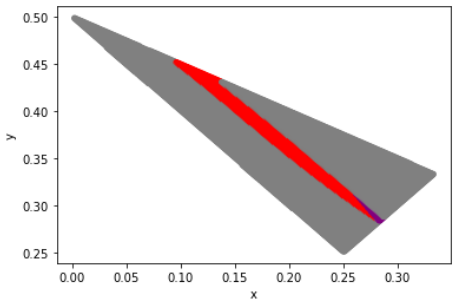
(b) $\alpha = 1.2$



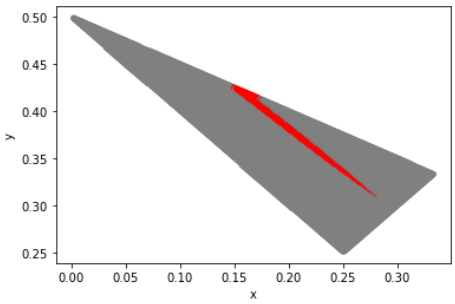
(c) $\alpha = 1.3$



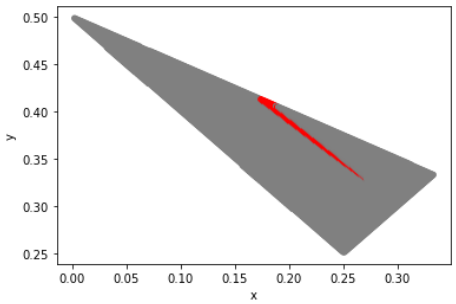
(d) $\alpha = 1.4$



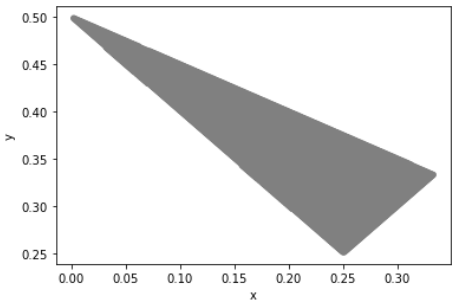
(e) $\alpha = 1.6$



(f) $\alpha = 1.8$



(g) $\alpha = 1.9$



(h) $\alpha = 2.2$

Figure 4: Red – Z attacks X, and Y chooses to stay out.
Purple – Z can attack either X or Y, and the third player stays out.

can only occur for values of α between about 1.2 and 2. This is a considerably smaller range than for many other coalitions and unilateral attacks.

Finally, Proposition 13 below shows that unilateral attacks by X on one of the other players would always decrease X's expected utility if the third player sits out the fight. Only in case that the third player joins on X's side may a unilateral attack by X occur. But in those cases, an XY– or XZ–coalition is anyway stable, so these scenarios do not add any cases of triumvirate instability.

Proposition 13. *X does not unilaterally attack either Y or Z if he would not be joined by the third player.*

Intuitively, if X has to fight alone, he is disadvantaged in both fights because his opponents are stronger and $\alpha > 1$. In addition, he would have to win both rounds of fighting in order to survive. This cannot be more attractive to X than peace.

5.6 Overall stability

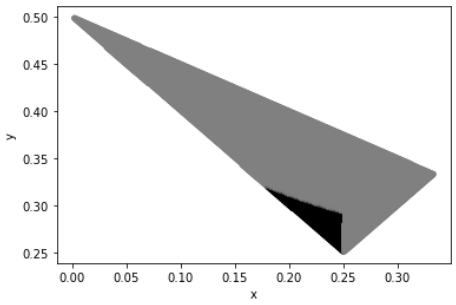
Figure 5 summarizes the cases in which a triumvirate is stable against any internal conflict, whether initiated by an attack coalition of two members against the third one, or initiated by a lone attacker.

Remember that, for $\alpha < 1$, there is no stable power constellation. Once α rises above 1, there is limited stability, primarily in a region where X and Y are approximately equally strong, and combined are only slightly stronger than Z.

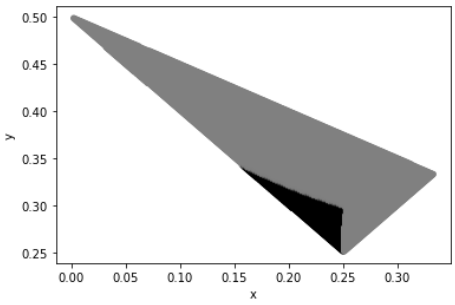
There are no stable triumvirates when α is high (for $\alpha > 20$, there are practically no stable power configurations). While attack coalitions become less of a problem when α is large (see Figure 2), unilateral attacks threaten stability more as α increases (see Figure 3). Thus, stability occurs only for intermediate values of α.

Interestingly, the area of stability changes significantly with α, but is always very small. None of the power distributions that are stable when $\alpha = 1.01$ or $\alpha = 1.2$ are stable when $\alpha = 3$, and vice versa. Moreover, the case of $\alpha = 1.4$ shows that the set of stable power configurations for a given α does not even need to be connected.

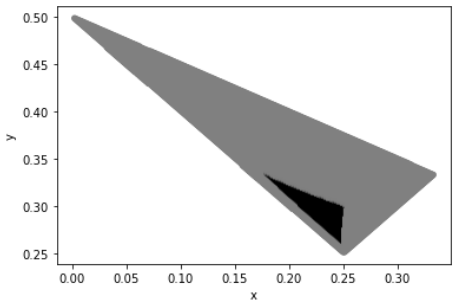
Stability is most likely to occur when X and Y together are barely stronger than Z alone (i.e., along the southwestern boundary of the parameter space). For low values of α, stability occurs close to the southern end where also X and Y are almost equally strong. In contrast, for higher values of α such as $\alpha = 3$, stability occurs when X and Y have substantially different levels of strength.



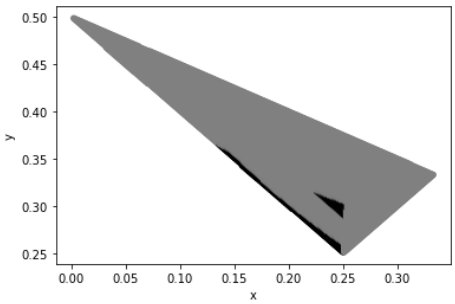
(a) $\alpha = 1.01$



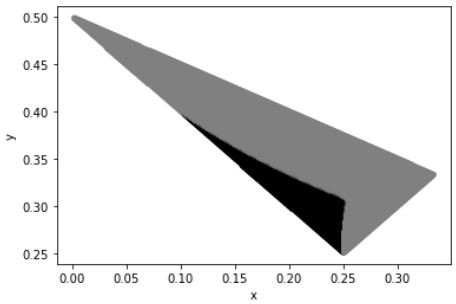
(b) $\alpha = 1.2$



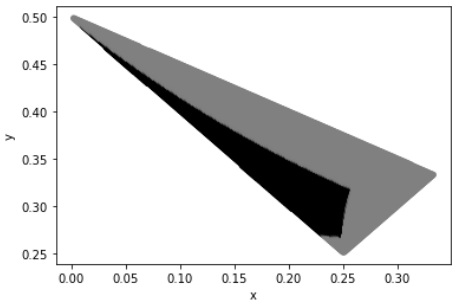
(c) $\alpha = 1.3$



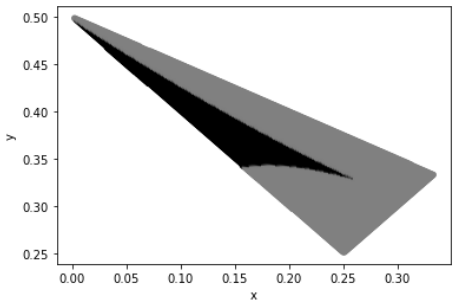
(d) $\alpha = 1.4$



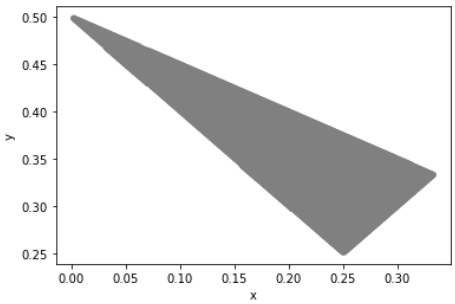
(e) $\alpha = 1.6$



(f) $\alpha = 2.2$



(g) $\alpha = 3$



(h) $\alpha = 20$

Figure 5: Black – Peace; Gray – Conflict

5.7 Specific game forms and stability

As explained in the model section, our concept of stability has players decide whether to start conflict based on a comparison between their expected utility in case of conflict, and their expected utility when peace prevails.

An alternative approach would have been to specify the exact sequence in which each player has the opportunity to initiate conflict or propose to a potential coalition partner (followed by different protocols of what happens after a coalition offer is rejected). Of course, there are many conceivable game trees that this process could follow, and there is no clear "most realistic" one.

In a particular game form, the decision to start a conflict, or to join a proposed coalition is affected by what would happen in the subgame where the conflict is not initiated, as this defines the players' continuation utilities. For example, a player who has received an offer to join a coalition might be willing to do so even if the expected payoff is lower than his peace share if he fears that, if he were to reject the proposal, he would be attacked by a coalition of the two other players (and that would give him an even lower utility).

For example, consider a scenario in which X receives an expected payoff of less than x when joining with Y in a coalition to oust Z, but where he would join a unilateral attack by Y on Z. Suppose the structure of the game is such that Y first can make a coalition offer, and then (if rejected) can choose to attack unilaterally. In this game, if Y proposes a coalition to X in the first stage, then X might as well accept that proposal because he knows that, if he were to reject, Y would attack nevertheless and essentially force X to join him at that point. Thus, the set of stability for XY-coalitions expands in this game form, relative to a game form where a rejection of the coalition proposal ends the game. Conversely, if, in a different game form, X expects a more favorable continuation after a rejection than his peace payoff, then the XY-coalition stability set could shrink.

However, Proposition 14 below shows that the set of parameters for which a triumvirate is stable under our approach is an upper limit (in the set of set inclusion) of the set of parameters for which a triumvirate is stable under any given game protocol that gives each player at least one chance to initiate conflict.

Definition 1. *A conflict initiation game is a game in which, as long as there is peace, there is exactly one player in each period t who can either pass (i.e., maintain peace), unilaterally initiate conflict, or propose to another player to form an attack coalition against the third player; in the latter case, the second player responds (in the same period, with conflict starting iff the second player agrees to the proposal). Once conflict is initiated, it proceeds exactly like in our model.*

The move sequence is defined by a finite sequence of triplets σ, with the interpretation that σ_t

gives the probability of each player to be the initiator in period t. For example,

$$((1,0,0),(0,1,0),(0,0,1),(1/3,1/3,1/3))$$

is a four period conflict initiation game in which X gets to move in period 1, Y gets to move in period 2, Z in period 3, and in the fourth period, one of the three players is randomly chosen as the potential initiator.

Let Γ *be the set of all possible initiation games with the property that, for each player p, there is at least one stage t at which* $\sigma_{t,p} = 1$*. That is, if peace prevails until stage t, player p gets the option to decide whether to initiate conflict.*

Proposition 14. *Fix* α*, and suppose that* $\gamma \in \Gamma$*, and the initial power distribution* (x, y, z) *is stable under game form* γ *(i.e., at each stage, along the equilibrium path, the acting player either plays pass, or makes a coalition proposal that is not accepted). Then the power distribution* (x, y, z) *is also classified as stable in our model (i.e., is in the black area in Figure 5).*

Proposition 14 shows that, for any game form that gives each player at least one chance to start a conflict, the set of stable initial power distributions is a subset of the set of stable initial power distributions in our model. Thus, our model provides an upper limit to stability under *any* reasonable description of how conflict could be initiated.

6 Discussion and Conclusion

6.1 The fragility of strategic stability

Acemoglu et al. (2008, 2009) develop an important theory of *strategic stability* of a power allocation between rival partners. In their theory, costs of war or other reasons to maintain peace are absent, and so their stability result is purely driven by strategic considerations: In particular, there are some potential coalitions that could win but are not formed because some members of these coalitions are rightfully concerned that, after a successful partnership phase, their coalition partners would turn against them and eliminate them next.

However, our model in which conflict outcome is stochastic suggests that this form of strategic stability is relatively fragile. For intermediate values of α, there are indeed some triumvirate power distributions that are stable against both unilateral and coalition attacks, but in the majority of cases, there is at least one player who prefers to break the peace.

Moreover, our model can sometimes provide indications about the constellations in which open

conflict arises. For example, a scenario with three essentially equally strong partners is unstable because (any) two partners can gang up on the third one. Even though the partners are aware that they will ultimately turn against each other, their equal strength gives each of them a substantial chance of winning, and thus an ex-ante expected payoff that is better than what they have when keeping the peace. In this situation, any of the three potential coalitions is stable.

In contrast, in situations with 3 differently-strong players, both the weakest and the strongest antagonists may prefer peace, but the middle player is likely to find it beneficial to start a conflict against the strongest player, effectively forcing the hand of the weak player: Once conflict has started, peace is no longer an option, and so the weakest player finds himself forced to choose between sitting out the initial conflict and then (most likely) facing the strongest player, or allying himself with the attacker and then having a better chance to win in the final round.

In yet another scenario where the two top players are relatively closely matched and substantially stronger than the third player, the strongest player may attack the weakest player, with the third player sitting out this conflict. The breakdown of the Second Roman Triumvirate between Octavian, Marc Antony and Lepidus (as described in the introduction) fits this scenario quite well.

6.2 Cost of war and risk aversion

In our model, the size of the prize is not diminished by conflict. Likewise, all players are indifferent between a *probability p* of receiving the entire prize, and a *share p* of the prize for sure. How would costs of war or risk aversion change our results?

In subgames with two remaining players, costs of war potentially stabilize the situation. The reason is that the stronger player now faces a trade-off: It is still the case that the stronger player's winning probability is higher (for $\alpha > 1$) than his share of the prize when there is peace. However, if conflict leads to a smaller aggregate size of the prize, then gaining a larger expected share of a smaller prize may not be worthwhile for the stronger player.

Interestingly, in initial subgames with three active players, this stabilization of two-player subgames may be destabilizing because weaker players are now more willing to enter attack coalitions.

For an example, consider the following initial power distribution: $x = 0.1$, $y = 0.4$, $z = 0.5$. Let $\alpha = 2.2$. Inspection of Figure 5 shows that this power configuration is stable in our model.[11]

Let us add a cost of war in the following way: After each battle, the size of the prize decreases by $\frac{C_1^\alpha}{C_1^\alpha + C_2^\alpha}$, where C_1 is the combined power of all the players on the (ex-ante) less powerful side and

[11] For ease of computation, the specific power configuration is on the boundary of the parameter space. However, none of the results in the following is a knife-edge case, so the same qualitative results would hold true for all power distributions in a sufficiently small neighborhood.

C_2 is the combined power of all the players on the (ex-ante) more powerful side. Intuitively, a conflict between two equally strong camps is likely to be drawn out and destroy a lot of infrastructure or economic potential, significantly diminishing the value of winning, while a conflict between two players of very different strengths is very likely over quickly and less destructive.[12]

In the Appendix, we show that, for these parameter values, it is now profitable for Z to attack X in the first round, and Y will sit out this attack. If Z successfully eliminates X, then Y and Z remain at peace (because a conflict between these rather evenly matched players would be very costly), and this justifies Y sitting out the first-round conflict rather than joining X. Furthermore, since Z is much stronger than X, the cost of the initial round of fighting is rather low.

6.3 Conclusion

In this paper, we have developed a new model of power consolidation. There are two main innovations: First, conflict outcomes are stochastic functions of strength. Second, conflict can be initiated either by a formateur forming an attack coalition or unilaterally attacking an opponent.

We show that this model, relative to existing ones, allows for an expanded range of interesting behavior. Conflict is not limited to one stage, and in the first stage, two players may form a mutually-beneficial coalition even though they know that they will eventually turn against each other. Furthermore, partial conflict (i.e., a round in which one player sits out the conflict between the two others) and unilateral conflict initiation are possible in our model.

Clearly, our model is very simple in certain aspects, and future research to generalize it would be very useful. In addition to the topics of risk aversion and costs of fighting addressed above, there are the following dimensions for generalizations. First, we have only analyzed initial power distributions with exactly three players. Second, if a fight is won by coalition partners in our model, their relative power remains the same; likewise, a one-on-one fight does not change the power of the winner relative to the power of the third player. In principle, there are two conceivable countervailing effects: winning the first round might provide additional resources to the winner, increasing his power in a future conflict, but on the other hand, casualties incurred during the fight might diminish the winner's power in a future conflict. Evidently, these generalizations complicate the analysis substantially so that they are left for future research.

[12] Clearly, there are different ways in which we could model the cost of war, such as a fixed cost per round of conflict, or a fixed percentage of the remaining prize that is lost in each round of conflict. The point of this example is only to show a possibility result, and it should be fairly clear that similar examples can be obtained for the other modeling choices as well.

Online-Appendix (not for publication)

Lemma 1

Lemma 1. *1. Suppose that f is a strictly concave function, and $f(0) = 0$. Then, for any $s \neq t$, $f(s) + f(t) > f(s + t)$.*

2. Suppose that f is a strictly convex function, and $f(0) = 0$. Then, for any $s \neq t$, $f(s) + f(t) < f(s + t)$.

3. In particular, if $\alpha < 1$, then $x^\alpha + y^\alpha > (x + y)^\alpha$. And if $\alpha > 1$, then $x^\alpha + y^\alpha < (x + y)^\alpha$.

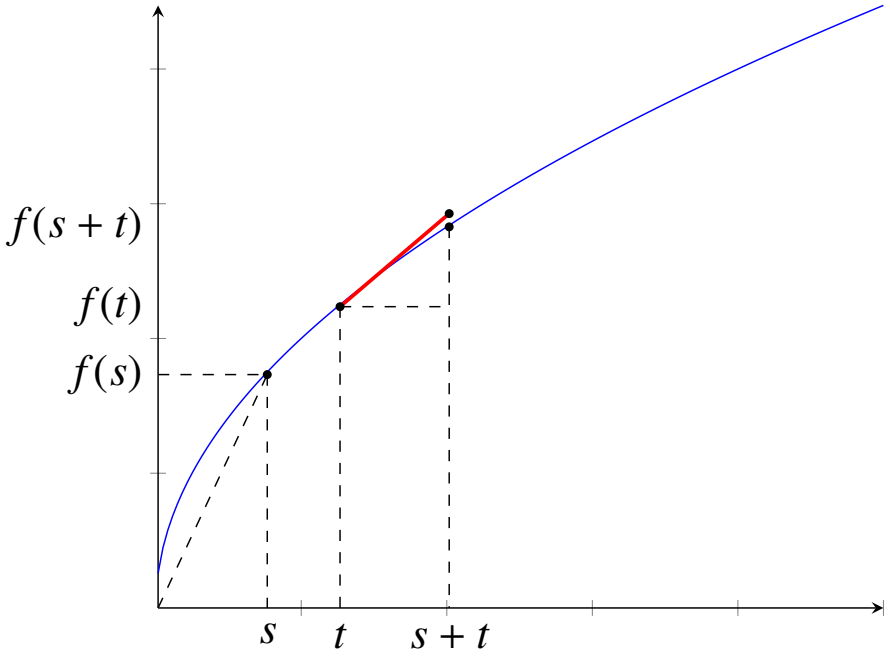


Figure 6: Illustration of Lemma 1

Proof. 1. Without loss of generality, let $s < t$. Consider the tangent to $f(x)$ at t. The point on this line at $x = s + t$ will have a value greater than $f(s + t)$ due to the definition of concavity.

Thus we get

$$f'(t) * s + f(t) > f(s + t).$$

We also know that $f'(t)$ is less than the slope of the line from the origin to $(s, f(s))$.

Thus,

$$f'(t) < \frac{f(s)}{s}.$$

This can be rearranged as

$$f'(t) \cdot s < f(s).$$

We can substitute this in the first inequality, thus

$$f(s) + f(t) > f'(t) \cdot s + f(t) > f(s + t).$$

This proves the claim.

2. By analogous arguments, a corollary to this is that when $f(x)$ is convex and goes through the origin, $f(s) + f(t) < f(s + t)$.

3. When $\alpha < 1$, x^α is a concave function. Thus, the first part of the Lemma may be applied, with $s = x$ and $t = y$ and $f(x) = x^\alpha$, proving that $x^\alpha + y^\alpha > (x + y)^\alpha$.

   When $\alpha > 1$, x^α is a convex function. Thus, the second part of the Lemma may be applied, with $s = x$ and $t = y$ and $f(x) = x^\alpha$, proving that $x^\alpha + y^\alpha < (x + y)^\alpha$. □

Proofs of Propositions

Proof of Proposition 1. Let the two players be named L and S (for "large" and "small"), with powers ℓ and s respectively, where $\ell > s$.

If the two players were to stay at peace, L and S would receive payoffs of $\frac{\ell}{\ell+s}$ and $\frac{s}{\ell+s}$ respectively. If the two players fight each other, L and S receive expected payoffs of $\frac{\ell^\alpha}{\ell^\alpha+s^\alpha}$ and $\frac{s^\alpha}{\ell^\alpha+s^\alpha}$ respectively.

We want to show that

$$\frac{\ell^\alpha}{\ell^\alpha + s^\alpha} \gtreqless \frac{\ell}{\ell + s} \iff \alpha \gtreqless 1. \tag{5}$$

Cross-multiplying and simplifying, this inequality is equivalent to

$$\ell^\alpha s \gtreqless \ell s^\alpha \iff \left(\frac{\ell}{s}\right)^\alpha \gtreqless \frac{\ell}{s} \iff \alpha \gtreqless 1. \tag{6}$$

Since $\frac{\ell}{s} > 1$, equation (6) is clearly satisfied. □

Proof of Proposition 2. 1. We will focus on Z (since both defenders would have to agree to a joint defense, this is without loss of generality). We need to show that Z's payoff from fighting separately exceeds his payoff from fighting X together with Y, and then fighting Y;

thus, we want to show that

$$\frac{z^\alpha}{x^\alpha + y^\alpha + z^\alpha} > \frac{(z+y)^\alpha}{x^\alpha + (y+z)^\alpha} \cdot \frac{z^\alpha}{y^\alpha + z^\alpha} \tag{7}$$

Canceling z^α and rearranging, this is equivalent to

$$\frac{y^\alpha + z^\alpha}{x^\alpha + y^\alpha + z^\alpha} = \frac{1}{1 + \frac{x^\alpha}{y^\alpha + z^\alpha}} > \frac{(z+y)^\alpha}{x^\alpha + (y+z)^\alpha} = \frac{1}{1 + \left(\frac{x}{y+z}\right)^\alpha}. \tag{8}$$

Cross-multiplying, this is equivalent to

$$\left(\frac{x}{y+z}\right)^\alpha > \frac{x^\alpha}{y^\alpha + z^\alpha} \iff y^\alpha + z^\alpha > (y+z)^\alpha, \tag{9}$$

which always holds for $\alpha < 1$ (see Lemma 1 in the Appendix).

2. We want to show that X prefers fighting both opponents over staying at peace. By the first claim, we know that, in the case of $\alpha < 1$, X's opponents will fight separately. X's expected payoff from attacking is thus

$$\frac{x^\alpha}{x^\alpha + y^\alpha + z^\alpha} = \frac{1}{1 + \left(\frac{y}{x}\right)^\alpha + \left(\frac{z}{x}\right)^\alpha} > \frac{1}{1 + \left(\frac{y}{x}\right) + \left(\frac{z}{x}\right)} = \frac{x}{x+y+z} = x, \tag{10}$$

where the inequality follows from the fact that y/x and z/x are greater than 1, and $\alpha < 1$. □

Proof of Proposition 3. The expected payoff for D_1 if he defends jointly with D_2 is $\frac{(d_1+d_2)^\alpha}{(d_1+d_2)^\alpha + a^\alpha} \cdot \frac{d_1^\alpha}{d_1^\alpha + d_2^\alpha}$. In a three-sided fight instead, his expected payoff is $\frac{d_1^\alpha}{d_1^\alpha + d_2^\alpha + a^\alpha}$.

We therefore need to show that:

$$\begin{aligned} \frac{(d_1+d_2)^\alpha}{(d_1+d_2)^\alpha + a^\alpha} \cdot \frac{d_1^\alpha}{d_1^\alpha + d_2^\alpha} &> \frac{d_1^\alpha}{d_1^\alpha + d_2^\alpha + a^\alpha} \\ \frac{1}{1 + \frac{a^\alpha}{(d_1+d_2)^\alpha}} &> \frac{1}{1 + \frac{a^\alpha}{d_1^\alpha + d_2^\alpha}} \iff \\ (d_1+d_2)^\alpha &> d_1^\alpha + d_2^\alpha. \end{aligned} \tag{11}$$

By Lemma 1, this holds for all $\alpha > 1$. The argument for the other player is analogous. Therefore, two players who are both fighting against the same player will always team up with each other when $\alpha > 1$. □

Proof of Proposition 4. By Proposition 3, we know that, if Z attacks both X and Y, they will join

forces. Therefore, Z's expected utility is

$$\frac{z^\alpha}{z^\alpha + (x+y)^\alpha} = \frac{1}{1 + \left(\frac{x+y}{z}\right)^\alpha}, \tag{12}$$

and we need to show that this is larger than z, Z's utility when remaining at peace. Observe that the assumption that $z > 1/2$ implies that $\frac{x+y}{z} < 1$, and thus, for $\alpha > 1$, we have $\left(\frac{x+y}{z}\right)^\alpha < \frac{x+y}{z}$. Therefore, the right hand side of (12) is larger than

$$\frac{1}{1 + \left(\frac{x+y}{z}\right)} = \frac{1}{\left(\frac{x+y+z}{z}\right)} = \frac{z}{x+y+z} = z,$$

and thus Z prefers attacking both opponents over staying at peace. □

Proof of Proposition 5. From Proposition 3, we know that if Z attacks both X and Y, they will join a defensive alliance. Thus, Z's expected utility is $\frac{z^\alpha}{z^\alpha+(x+y)^\alpha}$, and we need to show that

$$\frac{z^\alpha}{z^\alpha + (x+y)^\alpha} = \frac{1}{1 + \left(\frac{x+y}{z}\right)^\alpha} < z. \tag{13}$$

The facts that $x + y > z$ and $\alpha > 1$ imply $\left(\frac{x+y}{z}\right)^\alpha > \frac{x+y}{z}$. Thus, the left-hand side of (13) is smaller than

$$\frac{1}{1 + \left(\frac{x+y}{z}\right)} = \frac{1}{\left(\frac{x+y+z}{z}\right)} = z,$$

as claimed. □

Proof of Proposition 6. 1. We need to show that

$$U_y(\alpha; x, y, z) \equiv \frac{(x+y)^\alpha}{(x+y)^\alpha + z^\alpha} \cdot \frac{y^\alpha}{x^\alpha + y^\alpha} = \frac{1}{1 + \left(\frac{z}{x+y}\right)^\alpha} \cdot \frac{1}{1 + \left(\frac{x}{y}\right)^\alpha} > y. \tag{14}$$

Since $z < x + y$, we have $\left(\frac{z}{x+y}\right)^\alpha < \frac{z}{x+y}$, and because $x < y$, we have $\left(\frac{x}{y}\right)^\alpha < \frac{x}{y}$. Thus,

$$U_y(\alpha; x, y, z) > \frac{1}{1 + \left(\frac{z}{x+y}\right)} \cdot \frac{1}{1 + \left(\frac{x}{y}\right)} = \frac{x+y}{x+y+z} \cdot \frac{y}{x+y} = y.$$

2. Observe that X prefers to cooperate with Y whenever

$$U_x(\alpha; x, y, z) \equiv \frac{(x+y)^\alpha}{(x+y)^\alpha + z^\alpha} \cdot \frac{x^\alpha}{x^\alpha + y^\alpha} > x \tag{15}$$

Observe that $U_x(1;x,y,z) = x$ and $\lim_{\alpha\to\infty} U_x(\alpha;x,y,z) = 0$. Our objective is to show that U_x is quasi-concave in α, which then implies the claim.

Taking the natural logarithm of U_x, we get

$$\begin{aligned}\ln U_x &= \ln((x+y)^\alpha) - \ln((x+y)^\alpha + z^\alpha) + \ln(x^\alpha) - \ln(x^\alpha + y^\alpha)\\ &= \alpha\ln(x+y) - \ln((x+y)^\alpha + z^\alpha) + \alpha\ln(x) - \ln(x^\alpha + y^\alpha) \qquad (16)\end{aligned}$$

Differentiating (16) with respect to α, we get

$$\begin{aligned}\frac{d\ln U_x}{d\alpha} &= \ln(x+y) - \frac{\ln(x+y)(x+y)^\alpha + \ln(z)z^\alpha}{(x+y)^\alpha + z^\alpha} + \ln(x) - \frac{\ln(x)x^\alpha + \ln(y)y^\alpha}{x^\alpha + y^\alpha}\\ &= \ln(x+y)\left(1 - \frac{(x+y)^\alpha}{(x+y)^\alpha + z^\alpha}\right) - \frac{\ln(z)z^\alpha}{(x+y)^\alpha + z^\alpha} + \ln(x)\left(1 - \frac{x^\alpha}{x^\alpha + y^\alpha}\right) - \frac{\ln(y)y^\alpha}{x^\alpha + y^\alpha}\\ &= \ln\left(\frac{x+y}{z}\right)\frac{z^\alpha}{(x+y)^\alpha + z^\alpha} + \ln\left(\frac{x}{y}\right)\frac{y^\alpha}{x^\alpha + y^\alpha}\\ &= \ln\left(\frac{x+y}{z}\right)\frac{1}{(\frac{x+y}{z})^\alpha + 1} + \ln\left(\frac{x}{y}\right)\frac{1}{(\frac{x}{y})^\alpha + 1} \qquad (17)\end{aligned}$$

Differentiating another time yields

$$\begin{aligned}\frac{d^2\ln U_x}{d\alpha^2} &= -\ln\left(\frac{x+y}{z}\right)\frac{\ln\left(\frac{x+y}{z}\right)\left(\frac{x+y}{z}\right)^\alpha}{\left(\left(\frac{x+y}{z}\right)^\alpha + 1\right)^2} - \ln\left(\frac{x}{y}\right)\frac{\ln\left(\frac{x}{y}\right)\left(\frac{x}{y}\right)^\alpha}{\left(\left(\frac{x}{y}\right)^\alpha + 1\right)^2}\\ &= -\frac{\left[\ln\left(\frac{x+y}{z}\right)\right]^2\left(\frac{x+y}{z}\right)^\alpha}{\left(\left(\frac{x+y}{z}\right)^\alpha + 1\right)^2} - \frac{\left[\ln\left(\frac{x}{y}\right)\right]^2\left(\frac{x}{y}\right)^\alpha}{\left(\left(\frac{x}{y}\right)^\alpha + 1\right)^2} < 0,\end{aligned} \qquad (18)$$

because all terms in the two fractions are positive. Thus, U_x is quasiconcave. Together with $U_x(1;x,y,z) = x$, this implies that, for any (x,y,z), one of two cases must hold: Either there is no α for which (15) holds; or (15) holds for small values of α up to a critical value, and does not hold beyond the critical value. Which case obtains depends on whether $\frac{d\ln U_x(\alpha;x,y,z)}{d\alpha}$ is negative (in which case there is no α such that (15) holds) or positive (corresponding to the second case). Both cases are possible. In particular, for $x \approx 0$, and $y \approx z$, (17) is clearly negative, and for $x \approx y > z/2$, (17) is positive. □

Proof of Proposition 7. 1. Consider first an $Z - X$ coalition against Y. Z's expected utility in such a coalition is

$$\frac{(x+z)^\alpha}{(x+z)^\alpha + y^\alpha}\cdot\frac{z^\alpha}{x^\alpha + z^\alpha} = \frac{1}{1 + \left(\frac{y}{x+z}\right)^\alpha}\cdot\frac{1}{1 + \left(\frac{x}{z}\right)^\alpha}. \qquad (19)$$

Since both $\frac{y}{x+z}$ and $\frac{x}{z}$ are smaller than 1, this expected utility is increasing in α. Moreover, for $\alpha = 1$, the expected utility is z. This implies that for any $\alpha > 1$, Z's expected utility from the coalition is greater than his utility from keeping the peace, z.

2. X's utility from joining with Z against Y is

$$\frac{(x+z)^\alpha}{(x+z)^\alpha + y^\alpha} \cdot \frac{x^\alpha}{x^\alpha + z^\alpha}. \tag{20}$$

X's utility from joining with Y against Z is

$$\frac{(x+y)^\alpha}{(x+y)^\alpha + z^\alpha} \cdot \frac{x^\alpha}{x^\alpha + y^\alpha}. \tag{21}$$

Thus the claim is equivalent to showing that

$$\frac{(x+y)^\alpha}{(x+y)^\alpha + z^\alpha} \cdot \frac{x^\alpha}{x^\alpha + y^\alpha} \geq \frac{(x+z)^\alpha}{(x+z)^\alpha + y^\alpha} \cdot \frac{x^\alpha}{x^\alpha + z^\alpha}. \tag{22}$$

Rearranging, we get

$$\begin{aligned} &\frac{(x+y)^\alpha}{(x+y)^\alpha + z^\alpha} \cdot (x^\alpha + z^\alpha) \geq \frac{(x+z)^\alpha}{(x+z)^\alpha + y^\alpha} \cdot (x^\alpha + y^\alpha) \iff \\ &(x+y)^\alpha \cdot (x^\alpha + z^\alpha) \cdot [(x+z)^\alpha + y^\alpha] \geq (x+z)^\alpha \cdot (x^\alpha + y^\alpha) \cdot [(x+y)^\alpha + z^\alpha] \end{aligned} \tag{23}$$

we then distribute and combine the resulting like terms to get

$$\begin{aligned} (x+y)^\alpha \cdot (z^\alpha - y^\alpha) \cdot (x+z)^\alpha &\geq (x+z)^\alpha \cdot (x^\alpha + y^\alpha) \cdot z^\alpha - (x+y)^\alpha \cdot (x^\alpha + z^\alpha) \cdot y^\alpha \iff \\ z^\alpha - y^\alpha &\geq \frac{(x^\alpha + y^\alpha) \cdot z^\alpha}{(x+y)^\alpha} - \frac{(x^\alpha + z^\alpha) \cdot y^\alpha}{(x+z)^\alpha} \iff \\ \frac{1}{y^\alpha} - \frac{1}{z^\alpha} &\geq \frac{x^\alpha + y^\alpha}{(x+y)^\alpha \cdot y^\alpha} - \frac{x^\alpha + z^\alpha}{(x+z)^\alpha \cdot z^\alpha} \iff \\ \frac{1 - \frac{x^\alpha + y^\alpha}{(x+y)^\alpha}}{y^\alpha} &\geq \frac{1 - \frac{x^\alpha + z^\alpha}{(x+z)^\alpha}}{z^\alpha} \end{aligned} \tag{24}$$

Since the inequality is homogeneous in (x, y, z), we can normalize $x = 1$ (so that $z > y \geq 1$). Therefore, we now need to show that

$$\frac{1 - \frac{1+y^\alpha}{(1+y)^\alpha}}{y^\alpha} \geq \frac{1 - \frac{1^\alpha + z^\alpha}{(1+z)^\alpha}}{z^\alpha} \tag{25}$$

whenever $y \in [1, z]$. Clearly, this is equivalent with showing that the function

$$f(y) \equiv \frac{1 - \frac{1+y^\alpha}{(1+y)^\alpha}}{y^\alpha} = \frac{(1+y)^\alpha - 1 - y^\alpha}{y^\alpha \cdot (1+y)^\alpha} \tag{26}$$

is non-increasing in y. Taking the derivative with respect to y yields $f'(y) =$

$$\frac{[\alpha(1+y)^{\alpha-1} - \alpha y^{\alpha-1}] \cdot [y^a(1+y)^\alpha] - [(1+y)^\alpha - 1 - y^\alpha] \cdot [\alpha(1+y)^{\alpha-1} y^\alpha + (1+y)^\alpha \alpha y^{\alpha-1}]}{[y^\alpha \cdot (1+y)^\alpha]^2} =$$
$$\frac{\alpha(1+y)^{\alpha-1} y^{\alpha-1}}{[y^\alpha \cdot (1+y)^\alpha]^2} \cdot \Big([(1+y)^{\alpha-1} - y^{\alpha-1}] \cdot y \cdot (1+y) - [(1+y)^\alpha - 1 - y^\alpha] \cdot [y + (1+y)] \Big). \tag{27}$$

Because $\frac{\alpha(1+y)^{\alpha-1} y^{\alpha-1}}{(y^\alpha \cdot (1+y)^\alpha)^2}$ is positive, $f' \leq 0$ if and only if

$$\begin{aligned} [(1+y)^{\alpha-1} - y^{\alpha-1}] \cdot y \cdot (1+y) - [(1+y)^\alpha - 1 - y^\alpha] \cdot [y + (1+y)] \leq 0 \iff \\ y(1+y)^\alpha - y^\alpha(1+y) - 2y(1+y)^\alpha + 2y + 2y^{\alpha+1} - (1+y)^\alpha + 1 + y^\alpha \leq 0 \iff \\ 1 + 2y + y^{\alpha+1} - (1+y)^{\alpha+1} \leq 0. \end{aligned} \tag{28}$$

To show that this holds, we now show that (i) (28) holds for $\alpha = 1$ and (ii) the derivative of (28) is negative for all $\alpha > 1$. To show the first step, substituting $\alpha = 1$ yields

$$1 + 2y + y^2 - (1+y)^2 = 0. \tag{29}$$

For the second step, the derivative of (28) with respect to α is

$$\ln(\alpha+1) y^{\alpha+1} - \ln(\alpha+1)(1+y)^{\alpha+1} = \ln(\alpha+1)[y^{\alpha+1} - (1+y)^{\alpha+1}] \leq 0. \tag{30}$$

Since $\alpha > 1$ and $1 \leq y$, $\ln(\alpha+1)$ is positive and $y^{\alpha+1} - (1+y)^{\alpha+1}$ is negative.

3. We need to show that Y's utility from joining with X against Z is greater or equal to Y's utility from joining with Z against X; thus, we have to show that

$$\frac{(y+x)^\alpha}{(y+x)^\alpha + z^\alpha} \cdot \frac{y^\alpha}{x^\alpha + y^\alpha} \geq \frac{(y+z)^\alpha}{(y+z)^\alpha + x^\alpha} \cdot \frac{y^\alpha}{y^\alpha + z^\alpha}. \tag{31}$$

Proceeding in the same way as in the proof of number 2 above, this is equivalent to

$$\frac{1 - \frac{x^\alpha + y^\alpha}{(x+y)^\alpha}}{x^\alpha} \geq \frac{1 - \frac{y^\alpha + z^\alpha}{(y+z)^\alpha}}{z^\alpha}. \tag{32}$$

Again, we can normalize this by setting $x \equiv 1$ so that $1 \leq y \leq z$ in the following. Substituting in (32) and rearranging, we get

$$1 - \frac{1}{z^\alpha} - \frac{1+y^\alpha}{(1+y)^\alpha} + \left(\frac{y}{z(y+z)}\right)^\alpha + \frac{1}{(y+z)^\alpha} \geq 0. \tag{33}$$

We next show that (33) is increasing in z. Differentiating (33) with respect to z yields

$$\begin{aligned} &\alpha z^{-\alpha-1} - \alpha\left(\frac{y}{y+z}\right)^{\alpha-1} \frac{y(y+2z)}{z^2(y+z)^2} - \alpha(y+z)^{-\alpha-1} \\ &= \frac{\alpha}{z^{\alpha+1}(y+z)^{\alpha+1}}\left[(y+z)^{\alpha+1} - z^{\alpha+1} - y^\alpha(y+2z)\right] \end{aligned} \tag{34}$$

The sign of this expression is clearly determined by the sign of the term in square brackets. For $\alpha = 1$, the term in square brackets vanishes by the binomial formula, and differentiating it with respect to α yields

$$\ln(y+z)(y+z)^{\alpha+1} - \ln(y)y^{\alpha+1} - \ln(z)z^{\alpha+1} - 2z\ln(y)y^\alpha > ln(z)\left[(y+z)^{\alpha+1} - z^{\alpha+1} - (y+2z)y^\alpha\right]. \tag{35}$$

Observe that the term in square brackets in (35) is exactly the same as the term in square brackets in (34). Thus, whenever (34) is (close to) zero, the derivative with respect to α is positive, and thus the value of (34) can never become negative.[13]

Thus, we have shown that (33) is increasing in z. Therefore, if (33) holds for the lowest possible value of z, $z = y$, then it also holds for all other values of z. Substituting $z = y$ in (33), we obtain

$$1 - \frac{1}{y^\alpha} - \frac{1+y^\alpha}{(1+y)^\alpha} + 2\left(\frac{1}{2y}\right)^\alpha, \tag{36}$$

and need to show that this is non-negative for all $(\alpha, y) \geq (1, 1)$.

Differentiating (36) with respect to y yields

$$(1 - 2^{1-\alpha})\frac{\alpha}{y^{\alpha+1}} + \alpha\frac{(1+y^\alpha) - (1+y)}{(1+y)^{\alpha+1}} > 0, \tag{37}$$

as $y \geq 1$ and $\alpha \geq 1$.

Thus, (36) is smallest when $y = 1$, and substituting this in (36) yields

$$-\frac{2}{(2)^\alpha} + 2\left(\frac{1}{2}\right)^\alpha = 0$$

[13] Obviously, the expression in (34) is continuous in all variables.

In conclusion, this shows that (33) is always satisfied. □

Proof of Proposition 8. Observe that Y prefers to cooperate with Z whenever

$$U_y(\alpha; x, y, z) \equiv \frac{(y+z)^\alpha}{(y+z)^\alpha + x^\alpha} \cdot \frac{y^\alpha}{y^\alpha + z^\alpha} > y \tag{38}$$

Observe that $U_y(1; x, y, z) = y$ and $\lim_{\alpha\to\infty} U_y(\alpha; x, y, z) = 0$ because the first fraction goes to 1 and the second fraction goes to zero. Like in the proof of Proposition 6, our objective is to show that U_y is quasi-concave in α, which then implies the claim.

By going thorough the same steps as in the proof of Proposition 6, it is easy to show that

$$\frac{d \ln U_y}{d\alpha} = \ln\left(\frac{y+z}{x}\right) \frac{1}{(\frac{y+z}{x})^\alpha + 1} + \ln\left(\frac{y}{z}\right) \frac{1}{(\frac{y}{z})^\alpha + 1} \tag{39}$$

Differentiating another time yields

$$\frac{d^2 \ln U_y}{d\alpha^2} = -\frac{\left[\ln\left(\frac{y+z}{x}\right)\right]^2 \left(\frac{y+z}{x}\right)^\alpha}{\left(\left(\frac{y+z}{x}\right)^\alpha + 1\right)^2} - \frac{\left[\ln\left(\frac{y}{z}\right)\right]^2 \left(\frac{y}{z}\right)^\alpha}{\left(\left(\frac{y}{z}\right)^\alpha + 1\right)^2} < 0 \tag{40}$$

because all terms in the two fractions are positive. Thus, U_y is quasiconcave. Together with $U_y(1; x, y, z) = y$, this implies that, for any (x, y, z), one of two cases must hold: Either there is no α for which (38) holds; or (38) holds for small values of α up to a critical value, and does not hold beyond the critical value.

The proof for the $X - Z-$coalition is analogous and omitted. □

Proof of Proposition 9. Consider what happens when $x = y = z = 1/3$ and one player (say, Z) unilaterally attacks one of the other players (say, Y). If X sits out this fight (i.e., does not join anyone), Z wins the first round with probability $1/2$, and then wins against X with probability $1/2$, for an overall winning probability of $1/4 < 1/3$.

If, instead, X joins the fight on Y's side, Z's overall winning probability is

$$\frac{(1/3)^\alpha}{(1/3)^\alpha + (2/3)^\alpha} = \frac{1}{1 + 2^\alpha} < \frac{1}{3}.$$

Thus, independent of what X does, Z's payoff is less than when remaining at peace. Since these inequalities are strict and the winning probabilities are continuous in strength for any $\alpha < \infty$, the claim follows. □

Proof of Proposition 10. 1. The expected payoff for C if he joins with the weaker player is

$$\frac{(w+c)^\alpha}{(w+c)^\alpha + s^\alpha} \cdot \frac{c^\alpha}{w^\alpha + c^\alpha}. \tag{41}$$

If, instead, he joins with the stronger player, his expected payoff is

$$\frac{(s+c)^\alpha}{(s+c)^\alpha + w^\alpha} \cdot \frac{c^\alpha}{s^\alpha + c^\alpha}. \tag{42}$$

Finally, if staying out of the fight between the two others, C will fight the winner of their fight, and his expected payoff is therefore

$$\frac{s^\alpha}{s^\alpha + w^\alpha} \cdot \frac{c^\alpha}{s^\alpha + c^\alpha} + \frac{w^\alpha}{s^\alpha + w^\alpha} \cdot \frac{c^\alpha}{w^\alpha + c^\alpha}. \tag{43}$$

Our objective is to show that (43) is greater than (42). Forming the difference, we have

$$\begin{aligned}
&\left[\frac{s^\alpha}{s^\alpha + w^\alpha} - \frac{(s+c)^\alpha}{(s+c)^\alpha + w^\alpha}\right] \cdot \frac{c^\alpha}{s^\alpha + c^\alpha} + \frac{w^\alpha}{s^\alpha + w^\alpha} \cdot \frac{c^\alpha}{w^\alpha + c^\alpha} \\
&= \frac{w^\alpha c^\alpha}{s^\alpha + w^\alpha}\left[\frac{s^\alpha - (s+c)^\alpha}{(s+c)^\alpha + w^\alpha} + \frac{1}{w^\alpha + c^\alpha}\right] \\
&= \frac{w^\alpha c^\alpha}{(s^\alpha + w^\alpha)((s+c)^\alpha + w^\alpha)(w^\alpha + c^\alpha)}\left[(s^\alpha - (s+c)^\alpha)(w^\alpha + c^\alpha) + (s+c)^\alpha + w^\alpha\right] \\
&= \frac{w^\alpha c^\alpha}{(s^\alpha + w^\alpha)((s+c)^\alpha + w^\alpha)(w^\alpha + c^\alpha)}\left[(s+c)^\alpha(1 - w^\alpha - c^\alpha) + w^\alpha + s^\alpha(w^\alpha + c^\alpha)\right] > 0,
\end{aligned}$$

since $1 - w^\alpha - c^\alpha > 0$ when $\alpha > 1$, and all other terms are also positive.

The last claim (for $s = w$) follows immediately.

2. We want to see under which conditions (41) is greater than (43). This is the case if and only if

$$\begin{aligned}
\frac{(w+c)^\alpha}{(w+c)^\alpha + s^\alpha} \cdot \frac{c^\alpha}{w^\alpha + c^\alpha} &\geq \frac{s^\alpha}{s^\alpha + w^\alpha} \cdot \frac{c^\alpha}{s^\alpha + c^\alpha} + \frac{w^\alpha}{s^\alpha + w^\alpha} \cdot \frac{c^\alpha}{w^\alpha + c^\alpha} \iff \\
\left[\frac{(w+c)^\alpha}{(w+c)^\alpha + s^\alpha} - \frac{w^\alpha}{s^\alpha + w^\alpha}\right] \cdot \frac{1}{w^\alpha + c^\alpha} &\geq \frac{s^\alpha}{s^\alpha + w^\alpha} \cdot \frac{1}{s^\alpha + c^\alpha} \iff \\
\frac{(w+c)^\alpha(s^\alpha + w^\alpha) - w^\alpha[(w+c)^\alpha + s^\alpha]}{(w+c)^\alpha + s^\alpha} &\geq \frac{s^\alpha(w^\alpha + c^\alpha)}{s^\alpha + c^\alpha} \iff \\
\frac{(w+c)^\alpha - w^\alpha}{(w+c)^\alpha + s^\alpha} &\geq 2sw\frac{w^\alpha + c^\alpha}{s^\alpha + c^\alpha}.
\end{aligned} \tag{44}$$

At $c = 0$, the left-hand side of this inequality is equal to $-w^\alpha(s^\alpha + w^\alpha) < 0$, and (44) therefore

does not hold.

3. Suppose, by contradiction, that the claim is false. Then, there must be a power distribution (defined by w, s and c) and a value of α where the left hand side of the last line in (44) is zero and where the derivative of the left-hand side with respect to α is negative.

   Differentiating with respect to α yields

$$(w+c)^\alpha s^\alpha \ln(s(w+c)) - (w+c)^\alpha w^\alpha \ln(w(w+c)) - 2s^\alpha w^\alpha \ln(sw) - s^\alpha c^\alpha \ln(sc) - w^\alpha c^\alpha \ln(wc). \quad (45)$$

   Because $s(w+c) > \max(w(w+c), sw, sc, wc)$, this is larger than

$$\ln(s(w+c))((w+c)^\alpha s^\alpha - (w+c)^\alpha w^\alpha - 2s^\alpha w^\alpha - s^\alpha c^\alpha - w^\alpha c^\alpha) = 0,$$

   contradicting the assumption that the claim is false. □

Proof of Proposition 11. Applying equation (44) from Proposition 10, setting $w = y$, $s = z$ and $c = x$ yields

$$\frac{(x+y)^\alpha - y^\alpha}{(x+y)^\alpha + z^\alpha} \geq 2yz \frac{y^\alpha + x^\alpha}{z^\alpha + x^\alpha} \quad (46)$$

as the condition for X to join in Y's attack. Observe that the left-hand side of (46) can be written as

$$\frac{1 - \left(\frac{y}{x+y}\right)^\alpha}{1 + \left(\frac{z}{x+y}\right)^\alpha} \to 1$$

for $\alpha \to \infty$ because both fractions in numerator and denominator go to zero as $\alpha \to \infty$.

The fraction on the right-hand side goes to zero as $\alpha \to \infty$ because $z > y > x$. Thus, (46) holds for all sufficiently large values of α.

Proposition 6 implies that, if (46) holds so that he is guaranteed to be joined by X, Y prefers attacking Z over remaining at peace. □

Proof of Proposition 12. Z's utility when facing a unified opposition of X and Y is

$$\frac{z^\alpha}{z^\alpha + (x+y)^\alpha} = \frac{1}{1 = \left(\frac{x+y}{z}\right)^\alpha}.$$

At $\alpha = 1$, this is equal to z, and the expression is decreasing in α. Thus, for any $\alpha > 1$ and $x+y > z$, Z's expected utility when fighting both X and Y is lower than when remaining at peace.

For Z to unilaterally attack Y therefore requires two conditions to hold. First, Z's expected utility when attacking first X and then (in case of victory in the first round) Y must be larger than

his utility from remaining at peace, so (2) must hold. Second, it has to be true that X does not want to join Y, so (3) must hold.

For Z to unilaterally attack X also requires two conditions to hold. Since expected utility (2) is the same whether Z fights X first and then Y, or the other way around, this condition is unchanged. The second condition is that Y does not want to join X, so (3) must hold.

Rearranging (3) and (4) so that identical terms are on the right-hand side gives

$$\begin{aligned} \frac{[(x+y)^\alpha - y^\alpha][z^\alpha + x^\alpha]}{y} &< 2z[x^\alpha + y^\alpha][(x+y)^\alpha + z^\alpha] \\ \frac{[(x+y)^\alpha - x^\alpha][z^\alpha + y^\alpha]}{x} &< 2z[x^\alpha + y^\alpha][(x+y)^\alpha + z^\alpha], \end{aligned} \tag{47}$$

respectively. As $x < y$, the two terms in the numerator on the left-hand side of the first equation are smaller than the corresponding terms in the numerator on the left-hand side of the second equation. Furthermore, the denominator on the left-hand side of the first equation, y, is larger than in the second equation, x. It follows that the left-hand side of the first equation is smaller than the left-hand side of the second equation. Thus, whenever (4) is satisfied, so is (3). □

Proof of Proposition 13. Suppose, to the contrary of the claim, that X attacks Y, and Z sits out this fight. In this case, X's expected utility is

$$\frac{x^\alpha}{x^\alpha + y^\alpha} \cdot \frac{x^\alpha}{x^\alpha + z^\alpha} = \frac{1}{1 + \left(\frac{y}{x}\right)^\alpha} \cdot \frac{1}{1 + \left(\frac{z}{x}\right)^\alpha}. \tag{48}$$

We want to show that this is less than x (X's utility from peace) for any initial power distribution satisfying $x \leq y \leq z$ and $\alpha > 1$. Inspection of the right-hand side of (48) shows that the expression is non-increasing in α, so showing that (48) is less than x at $\alpha = 1$ is sufficient to show the claim. Thus, we need to show that

$$\frac{x^2}{(x+z)(x+y)} \leq x.$$

Rearranging, this is equivalent to

$$x \leq (x+z)(x+y) = (x+z)(1-z) = x + z - xz - z^2 \iff z(1-x-z) = zy \geq 0,$$

which is always satisfied. □

Proof of Proposition 14. We prove the claim by contradiction. Thus, suppose that (x, y, z) is stable under game form γ, but that it is classified as unstable in our model. That is, in our model, there is either a player p who gets a higher expected payoff when unilaterally initiating conflict than his

peace payoff, or there is a coalition in which both partners receive a higher expected utility than their peace payoffs.

Consider first the first case. In game γ, when player p has to move. Playing his equilibrium strategy, he can expect his peace payoff (as, by assumption, all other players will also remain peaceful), but he could unilaterally deviate and receive a higher payoff, a contradiction to the assumption that we are in an equilibrium.

Consider now the second case, in which a coalition between p and q is mutually beneficial, but is not implemented in the equilibrium of game γ. Let t denote the period in which p is the proposer. In a stage t subgame where player q is offered a coalition by p he must reject (otherwise p would actually make such a proposal, and receive a higher utility than in equilibrium, the desired contradiction).

The only reason for q to reject the proposal is that q gets an even higher payoff in the subgame that follows after a rejection, starting in $t + 1$. However, since the stage $t + 1$ game has exactly the same structure as the stage $t + 1$ subgame on the equilibrium path, it has the same payoffs for all players, the desired contradiction. □

6.4 Proof of the claim in section 6.2

Consider the initial power distribution $x = 0.1$, $y = 0.4$, $z = 0.5$, and let $\alpha = 2.2$. The claim is as that, with a cost of fighting that diminishes the prize by $\frac{C_1^\alpha}{C_1^\alpha + C_2^\alpha}$, where C_1 is the combined power of all the players on the (ex-ante) less powerful side and C_2 is the combined power of all the players on the (ex-ante) more powerful side, the equilibrium looks as follows.

In the first stage, Z unilaterally attacks X, and Y will stay out. If Z wins the first stage, then peace will prevail between Z and Y.

Consider the subgame where Z has attacked and won against X, so that the prize after the first round is diminished to $\left(1 - \frac{x^\alpha}{z^\alpha + x^\alpha}\right)$.

If Z now attacks Y, he wins with probability $\frac{z^\alpha}{z^\alpha + y^\alpha}$, and the prize is further diminished by $\frac{y^\alpha}{z^\alpha + y^\alpha}$. Thus, Z's expected payoff when attacking Y in the second round is

$$\frac{z^\alpha}{z^\alpha + y^\alpha} \cdot \left(1 - \frac{x^\alpha}{z^\alpha + x^\alpha} - \frac{y^\alpha}{z^\alpha + y^\alpha}\right) \approx 0.367.$$

In contrast, if Z stays at peace with Y, he receives a share $\frac{z}{z+y}$ of the (diminished) prize of ruling, so that Z's expected payoff is

$$\frac{z}{z + y}\left(1 - \frac{x^\alpha}{z^\alpha + x^\alpha}\right) \approx 0.540$$

Since this is larger, it follows that Z prefers no to attack Y in the second round. Furthermore, in this situation, Y will also want to stay at peace with Z. This follows from Proposition 1 because even without taking into account the additional cost of war, Y does not want to attack the stronger Z when $\alpha > 1$.

We now turn to the first stage. Z's expected payoff when fighting X alone, and then staying at peace with Y is

$$\frac{z^\alpha}{z^\alpha + x^\alpha} \times 0.54 \approx 0.525,$$

which is larger than 0.5. Thus, Z in fact prefers to attack X (given that Y does not join X in the first round).

It remains to be shown that Y does not want to join X in the first round. If Y stays out, his payoff is (at least[14])

$$\frac{z^\alpha}{z^\alpha + x^\alpha}\frac{y}{z+y}\left(1 - \frac{x^\alpha}{z^\alpha + x^\alpha}\right) + \frac{x^\alpha}{z^\alpha + x^\alpha}\frac{y}{x+y}\left(1 - \frac{x^\alpha}{z^\alpha + x^\alpha}\right) \approx 0.442.$$

Alternatively, Y could join X in the first round fight (and then, if victorious keep the peace with X). Since $x + y = z = 0.5$, each side's winning probability as well as the cost of war is $1/2$, so that this yields an expected payoff of

$$\frac{1}{2} \times \left(1 - \frac{1}{2}\right) \times \frac{4}{5} = 0.2.$$

Clearly, this is less attractive for Y than sitting out the first round, as claimed.

[14]The following formula calculates what Y can expect when keeping the peace with either winner. It is possible that Y can do even better by attacking X in the second round should X be the winner of the first round conflict with Z.

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin, "Coalition formation in non-democracies," *Review of Economic Studies*, 2008, *75* (4), 987–1009.

_ , _ , and _ , "Do juntas lead to personal rule?," *American Economic Review*, 2009, *99* (2), 298–303.

_ , Thierry Verdier, and James A Robinson, "Kleptocracy and divide-and-rule: A model of personal rule," *Journal of the European Economic Association*, 2004, *2* (2-3), 162–192.

Bai, Jinhui H and Roger Lagunoff, "On the faustian dynamics of policy and political power," *The Review of Economic Studies*, 2011, *78* (1), 17–48.

Boix, Carles and Milan W. Svolik, "The foundations of limited authoritarian government: Institutions, commitment, and power-sharing in dictatorships," *Journal of Politics*, 2013, *75* (2), 300–316.

Debs, Alexandre, "Living by the Sword and Dying by the Sword? Leadership Transitions in and out of Dictatorships," *International Studies Quarterly*, 2016, *60* (1), 73–84.

Fan, Xinyu, "Elite Persistence, Power Struggles, and Coalition Dynamics," 2019. Working paper, available at `https://papers.ssrn.com/abstract=4497191`.

Fearon, James D., "Rationalist explanations for war," *International Organization*, 1995, *49* (3), 379–414.

Gieczewski, Germán, "Policy persistence and drift in organizations," *Econometrica*, 2021, *89* (1), 251–279.

Jordan, James S., "Pillage and property," *Journal of Economic Theory*, 2006, *131* (1), 26–44.

Kilgour, D. Marc and Steven J. Brams, "The Truel," *Mathematics Magazine*, 1997, *70* (5), 315–326.

Konrad, Kai A., *Strategy and Dynamics in Contests*, Oxford University Press, 2009.

Konrad, Kai A and Stergios Skaperdas, "Succession rules and leadership rents," *Journal of Conflict Resolution*, 2007, *51* (4), 622–645.

Messner, Matthias and Mattias K Polborn, "Voting on majority rules," *The Review of Economic Studies*, 2004, *71* (1), 115–132.

Niou, Emerson and Peter Ordeshook, "Stability in anarchic international systems," *American Political Science Review*, 1990, *84* (4), 1207–1234.

Persico, Nicola, "A theory of non-democratic redistribution and public good provision," Technical Report, Working paper 2021.

Powell, Robert, "War as a commitment problem," *International Organization*, 2006, *60* (1), 169–203.

Ray, Debraj and Rajiv Vohra, "Coalition formation," *Handbook of game theory with economic applications*, 2015, *4*, 239–326.

Roberts, Kevin, "Dynamic voting in clubs," *Research in Economics*, 2015, *69* (3), 320–335.

Tullock, G, "Autocracy. Hingham, Mass," 1987.