The Instability of Power Sharing

Brams, Steven J. and Kilgour, D. Marc

New York University, Wilfrid Laurier University

January 2007

The Instability of Power Sharing

Steven J. Brams
Department of Politics
New York University
New York, NY 10003
USA
steven.brams@nyu.edu

D. Marc Kilgour
Department of Mathematics
Wilfrid Laurier University
Waterloo, Ontario N2L 3C5
CANADA
mkilgour@wlu.ca

January 2007

Abstract

Three models are presented in which two players agree to share power in a particular ratio, but either player may subsequently "fire" at the other, as in a duel, to try to eliminate it. The players have positive probabilities of eliminating each other by firing. If neither is successful, the agreement stays in place; if one is successful, that player obtains all the power; if each eliminates the other, both players get nothing.

In Model I, the game is played once, and in Model II it is repeated, with discounting of future payoffs. Although there are conditions under which each player would prefer not to shoot, satisfying these conditions for one player precludes satisfying them for the other, so at least one player will always have an incentive to shoot. In anticipation, its rival would prefer to shoot, too, so there will be a race to preempt.

In Model III, a damage factor caused by shooting, whether successful or not, is introduced into Model II. This mitigates the incentive to shoot but does not eliminate it entirely. The application of the models to conflicts, especially civil wars, is discussed.

The Instability of Power Sharing

1. Introduction

When two factions clash within a committee, a company, or a country, a natural question to ask is: Why don't they share power or responsibilities and try to reach a compromise that, while leaving neither faction in control, at least leaves neither too aggrieved.

At the interpersonal level, couples, sometimes aided by marriage counseling or couples therapy, resolve their differences and avoid divorce. In some families, the torch is passed on smoothly to new generations, especially in the creation and development of businesses. Father, son, and grandson shared responsibilities in the hugely successful IBM. Father and daughter now run Playboy Enterprises, and Rupert Murdoch's family has successfully divided responsibilities in their vast multimedia empire.

In organizations, conflicting parties frequently do find room to maneuver, make deals, and save face. At the national level, countries like Belgium and Switzerland have managed to avoid break-up despite their language and ethnic differences. And the most diverse large country in the world today, the United States, has remained one entity for over two hundred years. Indeed, the fact that it suffered, but survived, a major civil war may have contributed to its subsequent stability, as we will suggest later.

But these successes are probably the exception. Family quarrels, especially, are frequent and bitter. As a case in point, three Koch brothers split into two factions, in which identical twins took two different sides, in a fight over control of Koch Industries, Inc., an energy-exploration and trading company that is the largest private company in the United States. And while some countries have split peacefully, like the former Soviet Union into several republics in 1990 (with Chechnya the notable exception), and Czechoslovakia into the Czech Republic and Slovakia in 1993, it is civil wars that are the norm.[1] Since World War II, they have led to far more deaths and destruction than international wars and have a mixed record of leading to a sustainable peace.[2]

What is possibly most surprising is the number of voluntary mergers that go sour. At the international level, Egypt and Syria formed the United Arab Republic in 1958, only to see it dissolve in acrimony three years later. In business, voluntary "mergers of equals" are frequent, but in reality they hardly ever turn out that way. For example, while DaimlerChrysler and Citigroup started out as such mergers, one of the principals in each quickly became the dominant figure and ousted his erstwhile partner.

To try to understand the difficulties of power sharing, we begin with Model I (section 2), in which two players agree, initially, on how they will share the assets of their merged enterprise. Either player may then choose to break the agreement by "shooting" at its erstwhile partner—now its opponent—as in a duel. If this shot hits its mark, which we assume occurs with a specific probability, the shooter eliminates its opponent and acquires all the assets. If neither player is eliminated, the original sharing agreement stays in place.

Each player must worry that its opponent will fire first. Even if it is not rational for, say, player P to fire first, we show that it will be rational for player Q to do so. But

---

[1] Conditions that lead to civil wars are analyzed in, among other places, Collier *et al.* (2003) and Fearon and Laitin (2003).

[2] Conditions that produce sustainable peace after civil wars are discussed in, among other places, Rothchild and Roeder (2005) and Walter (2002).

anticipating a shot by Q, which may or may not be successful, P, in turn, can do better by getting in the first shot.

This result is extremely robust—it does not depend on the sharing agreement or the probability of success of either player. We show that neither player will be deterred from shooting, however large its initial share of the assets, and however small its probability of eliminating its opponent.[3]

This is true even with the discounting of future payoffs in a multiperiod game, as we show in Model II (section 3). No matter how small the discount factor and, therefore, how large the shadow of the future looms, each player will try to preempt its opponent in every period until one player or both are eliminated.

In Model III (section 4), we add a damage factor to Model II that renders the strategic situation somewhat more auspicious. In particular, we posit that when a shot is fired—whether it hits its mark or not—the assets to the players will be reduced.

Put another way, there is a cost, in reduced assets, when one or both players attack their opponents. This cost must be weighed against the benefit of acquiring, with some probability, all the assets, which, of course, would be reduced by the shooting. (If both

[3] There is a huge literature on modeling deterrence, but for the most part it ignores sharing agreements and an analysis of their stability. In international relations, see, for example, Zagare and Kilgour (2000), Powell (1999), and Brams and Kilgour (1988). Perhaps the models most relevant to our analysis are those of "predation and conflict" developed by Jack Hirshleifer and his associates, in which one side attempts to appropriate what others have produced. See, for example, Hirshleifer (2001), the articles in the edited collection of Garfinkel and Skarpedas (1996), and Garfinkel and Skarpedas (2007). Whereas these models are rooted in economic comparisons of cost and gain, our game-theoretic models of duels abstract from those details to focus on the simple question of whether or not to attack an opponent in a 2-person game. Thus, we do not address the question of the stability of coalitions in *n*-person games, nor do we consider the possibility of there being a higher authority or institution to enforce an agreement. We believe our main finding—that the damage caused by shooting is essential in halting it, but it is hard to make this factor salient—justifies our using "instability" rather than "stability" in the title. The other justification is empirical: Power struggles that provoke violent conflict are ubiquitous (Kadera, 2001).

players succeed in eliminating their opponents in the same period, this factor does not come into play, because both players receive nothing.)

When the assets are reduced by the act of shooting as well as the passage of time, the players will be deterred if the probabilities of eliminating their opponents are sufficiently low. But if, say, player P has a high probability of eliminating player Q, it will shoot, and so will Q, even if Q is not such a good shot.

This is perhaps why gunfighters in westerns invariably try to shoot each other. Either they are too good to do otherwise, or they worry that their rival will beat them to the draw. Consequently, they try to get in the first shot before the rival can respond. This metaphor, however, is not so apt today, because gunfighters usually do not affect the lives of others in the way, for example, that duelists in a multinational corporation or a country can ruin the lives of thousands or even millions of people when they battle for control.

We conclude (section 5) that the instability of power sharing, at least as modeled by the duel we analyze, cannot be easily overcome. Evidence from the recent deadly civil wars in Lebanon and the former Yugoslavia, in which competing factions vied for power at a cost of hundreds of thousand of lives, reinforces this conclusion.

But the damage factor in Model III suggests that there may be a brighter side—if not now, then later. The destruction wrought by the combatants in the Lebanese and Yugoslavian civil wars was horrific, but, ironically, one benefit was to make the combatants reluctant to renew hostilities. In addition, realizing that a quick and decisive blow against the enemy is probably not in the cards and more assets, instead, will be destroyed if there is further shooting, the duelists have been deterred. We briefly discuss what affects the strength of this damage factor in different settings.

2. Model I: One-Period Play

Assume there are two players, P and Q, who have probabilities, *p* and *q,* of eliminating their opponents when they fire at them. We assume $0 < p < 1$ and $0 < q < 1$, so while P and Q are not perfect shots, they each have some positive probability of eliminating their opponents when they shoot. While we assume that the players know *p* and *q*, they do not know when their shots will eliminate their opponents, making the outcomes resulting from their actions probabilistic rather than deterministic. But we assume that the players do know when a shot has been fired, rendering the duel "noisy" and the games to be described next ones of perfect information.[4]

The conflict between P and Q is over how they will divide their total assets, which we assume sum to 1. Suppose, initially, that both players agree to divide their assets so that P receives *a* units and Q receives $1 - a$ units.[5] After reaching this agreement, each player then must decide whether to fire (F) or not fire ($\overline{\mathrm{F}}$) at its opponent to try to gain all the assets and thereby not have to share power. There are four possible outcomes:

1. $\overline{\mathbf{F}}\,\overline{\mathbf{F}}$. Both P and Q hold their fire. The distribution of the assets to (P, Q) remains $(a, 1-a)$.

---

[4] By contrast, in a "silent" duel, the shots are no less deadly, but the players do not know when they have been fired upon unless the shot is successful. For a general analysis of noisy and silent duels, see Karlin (1959, chs. 5 and 6).

[5] We assume that assets are a surrogate for power, although we recognize that assets in the form of money or other resources are only one component of what is normally considered to be power. Note that the division of the assets into *a* and 1 – *a* might be an expectation of how the assets will be divided, not necessarily a formal agreement, and might be proposed, or even imposed, by some outside power. Whether there is a formal agreement or not, we assume that P and Q can make the choices we describe next.

2. $\mathbf{F\bar{F}}$. P fires first. Q is eliminated with probability p, in which case P receives all the assets and the distribution is $(1,0)$. But Q, if not eliminated, then Q fires at P, eliminating P with probability q, in which case the distribution is $(0,1)$.[6] If neither P nor Q is eliminated, the distribution of assets remains $(a,1-a)$.

3. $\mathbf{\bar{F}F}$. Q fires first. P is eliminated with probability q, in which case Q receives all the assets and the distribution is $(0,1)$. But P, if not eliminated, then fires at Q, eliminating Q with probability p, in which case the distribution is $(1,0)$. If neither P nor Q is eliminated, the distribution of assets remains $(a,1-a)$.

4. **FF**. P and Q fire simultaneously. Q is eliminated with probability p and, independently, P is eliminated with probability q. If there is one survivor, that survivor receives all the assets. If both players survive, the distribution of assets remains $(a,1-a)$. If neither player survives, both players receive 0, in which case the distribution is $(0,0)$.

In a standard (noisy) duel, the players can observe (or hear) each other's actions. Consequently, each knows if its opponent gets off the first shot (in which case it may be too late for the second player to respond). If the players cannot directly observe each other's actions, which will be true in many situations that we seek to model, we need only assume that the players act independently, without knowledge of each other's actions.

We suppose throughout that P and Q are rational and make choices to maximize their expected share of assets, which we use as a surrogate measure for their expected share of power. When will the players be deterred from shooting?

Assume that Q is docile and does not attempt to eliminate P—it chooses $\bar{\mathrm{F}}$. We compare P's expected share of power when it decides to share power with Q at $\bar{\mathrm{F}}\bar{\mathrm{F}}$ with its

[6] This is an assumption of the model, but it will be justified later as always a rational choice in Models I, II, and III whenever one player (P or Q) fires first.

expected share when it attempts to eliminate Q at $F\overline{F}$.

If Q is docile and does not fire at P, even after surviving P's attempt to eliminate it, P's expected share of power will be

$$EP^{\mathrm{I}} = p[1] + (1-p)[a] = a + p(1-a) \geq a$$

for all values of p (the superscript ,"I," signifies Model I). Therefore, P is always better off trying to eliminate a docile Q.

More realistically, suppose that Q is not docile but merely slower than P. Then if P attempts to eliminate Q and does not succeed, Q responds by trying to eliminate P. If neither player's attempt succeeds, the power-sharing arrangement stays in place, with no loss to either player (we modify this assumption in Model III). In this case, P's expected share of power[7] is

$$EP_1^{\mathrm{I}} = p[1] + q(1-p)[0] + (1-p)(1-q)[a] = p + a(1-p)(1-q).$$

P will accept the power-sharing arrangement iff $a \leq EP_1^{\mathrm{II}}$, or

$$a \geq p + a(1-p)(1-q),$$

which is equivalent to

$$p \leq \frac{aq}{1-a+aq}. \qquad (1)$$

The region in which power sharing is rational for P is labeled **1** in Figure 1; this figure is drawn under the assumption that $a = 0.6$. For all values of (q, p) in region **1**, P will not attack Q and accept power sharing; outside this region, P does better firing at Q.

[7] Here and later in the text we use subscripts on *EP* to indicate the conditions under which P's expected value is calculated: Subscript "1" means that P fires first; subscript "2" that P fires second, and subscript "0" that both players fire simultaneously.

Figure 1 about here

A similar analysis shows that if P does not attack Q first but retaliates if Q attacks first and P survives—the "realistic scenario" postulated above—then Q maximizes its expected share of power by not attacking P iff

$$p \geq \frac{aq}{(1-a)(1-q)}. \tag{2}$$

For all values of (q, p) in the region labeled **2** in Figure 1, Q will not attack P and accept power sharing; outside this region, Q does better firing at P.

To summarize, given that P is assured that Q will not attack first but will retaliate if attacked, P will be deterred from attacking first only if (q, p) lies in region **1**. Similarly, if Q is assured that P will not attack first but will retaliate, A will be deterred from attacking first only if (q, p) lies in region **2**. It follows that both P and Q will accept the power-sharing arrangement iff the point (q, p) lies in *both* regions.

But as suggested by Figure 1, these regions do not overlap. To prove this, suppose that (q, p) lies in both regions. Then

$$\left(\frac{aq}{1-a+aq} - p\right) + \left(p - \frac{aq}{(1-a)(1-q)}\right) \geq 0$$

because, by (1) and (2), the left side is the sum of two non-negative terms. But this sum equals

$$\frac{aq}{1-a+aq} - \frac{aq}{(1-a)(1-aq)} = \frac{-aq^2}{(1-a+aq)(1-a)(1-q)} < 0.$$

This contradiction implies that (q, p) cannot satisfy both (1) and (2) simultaneously.[8]

---

[8] In fact, it can be verified that the curves defining (1) and (2) have a common tangent at the origin. The fact that the deterrence regions are disjoint means that the players do not need to know the values of any

Thus, it is never rational for both players to accept power-sharing. Note that this conclusion does not depend on the specific values of *p*, *q*, or *a*.

We can say more about what will happen by determining P's best course of action if Q attempts to eliminate P initially. If Q attempts to eliminate P and P, if not eliminated, attacks Q, then P's expected share of power is

$$EP_2^{\mathrm{I}} = q[0] + (1-q)p[1] + (1-q)(1-p)[a] = p - pq + a(1-p)(1-q).$$

If P and Q attack each other simultaneously, P's expected share of power is

$$EP_0^{\mathrm{I}} = p(1-q)[1] + q[0] + (1-p)(1-q)[a] = p - pq + a(1-p)(1-q).$$

Because $EP_0^{\mathrm{I}} = EP_2^{\mathrm{I}}$, P will indifferent between attacking Q at the same time that Q attacks P, or attacking Q only if Q attacks P unsuccessfully. This means that shooting simultaneously with Q, and shooting after Q, are equally good for P.[9]

It is easy to verify that

$$EP_1^{\mathrm{I}} - EP_2^{\mathrm{I}} = EP_1^{\mathrm{I}} - EP_0^{\mathrm{I}} = pq > 0.$$

Therefore, for any values of *p* and *q*, P prefers to attack Q before Q attacks P; P is indifferent between attacking simultaneously and allowing Q to attack first (with P firing in

---

parameters of the game: Knowing that both can never, simultaneously, be deterred is sufficient to incite them to try to preempt each other. But the specific values of the probabilities, *p* and *q*, of eliminating an opponent—and new parameters to be introduced—do become important in Model III, wherein a region exists in which both players can benefit from power-sharing.

[9] Alternatively, we might assume that if P is eliminated and therefore gets nothing, it would prefer to do so when it eliminates Q (who would also get nothing) than when Q survives and gets everything. In other words, dying alone is worse than dying with one's opponent. The difficulty with this assumption is that we are not modeling literally dying in a duel but, instead, being prevented from sharing power with another player. If P, because of *Schadenfreude* or other reasons, prefers that Q also be cut out of power if P is, what payoff greater than 0 should P receive? And aren't there P's that would have the opposite preference, wanting an opponent to win control rather than see a new player take over? Because the preferences of the eliminated players are by no means apparent in this kind of situation, we have not included them as a parameters in the game. Note, incidentally, that the game is not constant-sum, because when P and Q shoot simultaneously and succeed in eliminating each other, they both end up losers with 0—as compared with other outcomes at which their power shares sum to 1.

return if it is not eliminated). Of course, exactly the same conclusions apply to Q.

Even when P would prefer to accept the power-sharing arrangement if it knew that Q would accept it too, P knows that Q maximizes its expected share of power by attacking P first. Exactly the same conclusion applies to Q. Ineluctably, P and Q are led to try to be the first to attack in a preemptive race that applies the anti-golden rule: "Do to your opponent what he would do to you, but do it first."

Earlier we argued that power-sharing would not be accepted by both sides in any region of the (q, p) unit square of Figure 1. Now we claim, more ominously, that both players will attempt to preempt throughout the unit square.

More specifically, in region **1** P prefers not to attack (on the assumption that if P does not attack, the power-sharing arrangement will be implemented), whereas Q prefers to attack. As we have just seen, the knowledge that Q prefers to attack will induce P to try to attack before Q.

Because P is indifferent between attacking simultaneously with Q or allowing Q to attack first, P has good reason to preempt in region **1**. By a similar argument, we can also expect a race to preempt in region **2**. Finally, in region **3**, neither side will be willing to wait under any conditions, because each player is always better off attacking first. In this sense, region **3** is the most unstable region, because each player has a dominant strategy of attacking.

Because of the rationality of preemption by one player in regions **1** and **2**, and by the other player in anticipation of this preemption, these regions are hardly more stable than region **3**. We conclude that no matter what the values of p and q are, no matter what the value of a—including $a = 1/2$, when power is shared equally—is, and no matter what knowledge the players have of each other's intentions, each player maximizes its expected share of power by racing to preempt.

3. Model II: Multiple-Period Play with Discounting of Assets

In Model II, Model I is played repeatedly: In discrete time periods 1, 2, 3, . …, each player may fire or not fire at its opponent. The players discount their future payoffs, using discount factor r. Thus, receiving k units of assets t periods in the future is worth the same as kr^t units received now. As in Model I, P and Q are assumed to make choices so as to maximize their expected total assets, but now over all time in which one or both players survive.

We search for stationary (or Markov) equilibrium behavior, whereby a player's choice of a strategy depends only on the situation that the player faces at the moment, not on the history of play.[10] With this understanding, we analyze the situation at time 0.

Suppose P believes that Q will not fire first at time 0, and in all subsequent periods because of stationarity. If P also does not fire, so both players survive into the infinite future, P's expected current value of its stream of revenues is

$$a + ar + ar^2 + ar^3 + \ldots = \frac{a}{1-r}. \tag{3}$$

On the other hand, if P tries to eliminate Q in every period, and if Q responds by trying to eliminate P, then the expected value of P's revenue stream at time t is

$$EP_1^{\mathrm{II}} = p[1 + r + r^2 \ldots] + q[0] + (1-p)(1-q)[a + rEP_1^{\mathrm{II}}],$$

where the first term on the right reflects P's success in period 0 by firing first and eliminating Q, the second term that Q eliminates P (whether Q is eliminated or not) and the third term that both players survive period t and the game continues, giving P a payoff of a from period t plus its expectation in period $t+1$, discounted by factor r. The stationarity condition implies that $EP_1^{\mathrm{II}}(t) = EP_1^{\mathrm{II}}(t+1)$. The recursion can then be solved directly; at

[10] This restriction, which facilitates calculations, does limit the players' behavior in that it rules out history-dependent strategies like tit-for-tat. We assess next in the text whether an analogue of a trigger strategy can stabilize power-sharing by assuming that, if either player attempts to eliminate an opponent, the players enter into a duel, firing alternately until one of them is eliminated. For now we note that a trigger strategy constitutes a greater threat than tit-for-tat, so if tit-for-tat can stabilize power-sharing, so can a suitable trigger strategy. Conversely, if a trigger strategy cannot stabilize power-sharing, then the analogous tit-for-tat strategy cannot stabilize it either.

any time, the expected value of P's future revenue stream is

$$EP_1^{\mathrm{II}} = \frac{p + (1-p)(1-q)(1-r)a}{(1-r)[1-(1-p)(1-q)r]}. \tag{4}$$

Note that EP_1^{II} is analogous to EP_1^{I} in Model I.

Subtracting (4) from (3), P will be at least as well off not firing at Q as firing at it, and thereby sharing in the income stream, iff

$$\frac{a}{1-r} - \frac{p + (1-p)(1-q)(1-r)a}{(1-r)[1-(1-p)(1-q)r]} = \frac{a - p - (1-p)(1-q)a}{(1-r)[1-(1-p)(1-q)r]} \geq 0. \tag{5}$$

Because the denominator on the right side of equation (5) must be positive, the inequality will be satisfied iff the numerator on the right is non-negative, which is equivalent to inequality (1).

In other words, the condition under which P will share power—assuming Q is also willing to share power—is the same as in Model I. Moreover, while the expected current values from the two courses of action depend on the parameter r in Model II, whether or not inequality (5) holds is in fact independent of the value of r.

As in Model I, it is easy to verify that it is in Q's interest to respond to P's firing first by firing at P if it survives. If Q does not respond, then P should still attempt to eliminate Q subsequently, regardless of the values of the parameters p, q, a, and r.

It can also be shown that if Q attempts to eliminate P, as in Model I, then P should respond by firing itself. Furthermore, P's expected payoff is the same whether, in each period, P fires after Q or at the same time.

Finally, Q will be willing to share Q's revenue stream, rather than try to eliminate P, iff condition (2) holds, exactly as in Model I. But then it will not be in P's interest to refrain from firing at Q.

Thus, Figure 1 applies to Model II as well as Model I: At every point (q, p) in the unit square, either (1), (2), or both fail. It follows that, whatever the values of the parameters, preemption is optimal in every period in which both players survive.

4. Model III: Multiple-Period Play with Discounting of, and Damage to, Assets

In Model III, Model II is modified to include a damage factor. In the event that a player fires at but does not eliminate its opponent in any period, the total assets of that period are reduced by damage factor *s*, where $0 \le s \le 1$. The reduced assets are shared in the ratio $a:1-a$ if nobody is eliminated, regardless of which player or players fired a shot.[11]

In other words, there is a cost to the players of shooting, which, of course, is most harmful if a player is eliminated (it gets 0) but still hurts the player or players that survive. Besides the damage factor, the players' assets continue to be reduced by the discount factor *r* from period to period. Does the damage factor change the preemption behavior that we found in Models I and II?

As in Model II, we search for stationary (or Markov) equilibrium behavior and begin by analyzing the situation at time 0. Suppose P believes that Q will not fire first at time 0. If P does not fire and both players survive, P's expected current value of its stream of revenues is

$$a + ar + ar^2 + ar^3 + \ldots = \frac{a}{1-r}. \tag{6}$$

On the other hand, if P tries to eliminate Q in every period, and if Q responds by trying to eliminate P, then the expected current value of P's revenue stream will be

$$EP_1^{\text{III}} = p[1 + r + r^2 \ldots] + q[0] + (1-p)(1-q)[a + rsEP_1^{\text{III}}],$$

---

[11] To be sure, P and Q will seek to destroy their opponents' assets rather than their own. But the fact, as we will show once again, that if shooting by one player is rational, it will lead to *mutual* preemption means that the assets of both sides will be damaged, whether or not the shooting eliminates an opponent. Thus, the assumption of Model III that shooting will be damaging to both sides seems reasonable, because both players will always shoot.

which, after applying stationarity, can be solved to obtain

$$EP_1^{\mathrm{III}} = \frac{p + (1-p)(1-q)(1-r)a}{(1-r)[1-(1-p)(1-q)rs]}. \tag{7}$$

Note that EP_1^{III} is analogous to EP_1^{II} in Model II.

Subtracting (7) from (6), P will be at least as well off not firing at Q as firing at it, and thereby sharing in the income stream, iff

$$\frac{a}{1-r} - \frac{p + (1-p)(1-q)(1-r)a}{(1-r)[1-(1-p)(1-q)rs]} = \frac{a - p - a(1-p)(1-q)[1-r(1-s)]}{(1-r)[1-(1-p)(1-q)rs]} \geq 0. \tag{8}$$

As previously, because the denominator on the right side of equation (8) must be positive, the inequality will be true iff the numerator on the right is non-negative:

Define $R = r(1-s)$. This parameter represents the current value of the reduction in the next period's income if the players attempt to eliminate each other during the first time period, 0. We call R the *discounted damage*. Note that $0 \leq R \leq 1$. The parameters, r and s, affect the stability of power sharing only through the discounted damage.

The condition for P to accept power sharing—that the numerator of the right side of (8) be non-negative—is

$$a - p - a(1-p)(1-q)(1-R) \geq 0,$$

which is equivalent to

$$p \leq \frac{aq + a(1-q)R}{1 - a(1-q)(1-R)}. \tag{9}$$

If $R = 0$, the fraction on the right side of (9) is identical to the fraction on the right side of (1). Thus, if the discounted damage is 0, the condition for power sharing is identical to that in the previous models. But if $R > 0$, condition (9) is a new constraint on power sharing.

It is easy to verify that the fraction on the right side of (9) is a strictly increasing function of q. The value of this function at $q = 0$ is

$$p_0 = \frac{aR}{1 - a + aR}.$$

Note that $0 < p_0 < a$ provided that $R > 0$. The graph of this fraction is the curved line passing from $(0, p_0)$ to $(1, a)$ in Figure 2; this figure is drawn under the assumption that $a = 0.6$, the same value as in Figure 1, and $R = 0.2$.

Figure 2 about here

An analogous calculation for Q shows that it will be willing to accept power sharing iff

$$q \leq \frac{(1-a)p + (1-a)(1-p)R}{1-(1-a)(1-p)(1-R)},$$

which is equivalent to

$$p \geq \frac{aq - (1-a)(1-q)R}{(1-a)(1-q)(1-R)}. \tag{10}$$

It is easy to verify that the fraction on the right side of (10) is a strictly increasing function of q. It equals 0 when $q = q_0$, where

$$q_0 = \frac{(1-a)R}{a + (1-a)R},$$

and it equals 1 when $q = 1 - a$. The graph of this fraction is the curved line passing from $(q_0, 0)$ to $(1-a, 1)$ in Figure 2.

As shown in Figure 2, p_0 and q_0 help to define the region (shaded) in the (q, p) unit square, where both P and Q are willing to accept power sharing in the ratio $a : 1 - a$.

Considered as functions of $R = r(1-s)$, p_0 and q_0 are strictly increasing and satisfy $p_0 \to 0^+$ and $q_0 \to 0^+$ as $R \to 0^+$, and $p_0 \to a^-$ and $q_0 \to (1-a)^-$ as $R \to 1^-$.

A sufficient condition for (q, p) to fall in the stable power-sharing region is $q \le q_0$ and $p \le p_0$. But it is not necessary; a portion of the shaded region in Figure 2 lies above and to the right of the point (q_0, p_0).

To analyze this region further and, in particular, to find a necessary condition for stable power sharing, we study the coordinates of the point (q_1, p_1), shown in Figure 2. Formally, (q_1, p_1) is defined as the value of (q, p) achieving equality in both (9) and (10). Equating the right sides of (9) and (10) and solving for *R* gives

$$R = \frac{aq^2}{1 - a - q + aq^2}. \tag{11}$$

To invert (11) to obtain a solution for q_1 in terms of *R*, note first that (11) is equivalent to the quadratic equation

$$a(1-R)q^2 + Rq - (1-a)R = 0.$$

Because the coefficient of q^2 is positive and the left side of the equation is negative at $q = 0$, it follows that $q = q_1$ is the unique positive root of this equation, which is

$$q_1 = \frac{-R + \sqrt{R^2 + 4aR(1-a)(1-R)}}{2a(1-R)}. \tag{12}$$

An analogous calculation shows that

$$p_1 = \frac{-R + \sqrt{R^2 + 4aR(1-a)(1-R)}}{2(1-a)(1-R)}. \tag{13}$$

As illustrated in Figure 2, a necessary condition for power sharing to be stable is

$q \le q_1$ and $p \le p_1$. Note that $p_1 : q_1 = a : 1 - ap_1$, so the probabilities at the upper right-hand corner of the stable power-sharing region are in the same ratio as the power shares.

Moreover, it can be shown that as $R \to 0^+, p_1 \to 0^+$, and $q_1 \to 0^+$, whereas as $R \to 1, p_1 \to a^-$ and $q_1 \to (1 - a^-)$. Finally, for any *R* satisfying $0 < R < 1$, it can be verified that $0 < p_0 < p_1 < a$ and $0 < q_0 < q_1 < 1 - a$. Thus, when the discounted damage, *R*, is close to 0, the stable power-sharing region is very close to the point $(q, p) = (0,0)$.

We interpret these conditions as follows. When *R* is close to 0—that is, the discounted damage is low—the stable power-sharing region is small because the assets retain much of their value over time even when there is shooting. On the other hand, when *R* is close to 1—that is, the discounted damage is high—the stable power-sharing region almost fills the rectangle defined by $0 \le q \le 1 - a$ and $0 \le p \le a$. Because much can be lost by shooting in this case, the players have good reason to refrain, especially if *p* and *q*, respectively, fall below the players' shares of power, *a* and $1 - a$. A diminution of *p* and *q* could well be facilitated by an arms-control agreement.

The situation when $a = 1/2$, and power is shared equally, is noteworthy. First, equal power-sharing maximizes the area of the power-sharing region, which might help to persuade the players to renegotiate an unequal sharing agreement, especially should they be uncertain about the values of *p* and *q*.

Second, all points, lines, and regions are symmetric with the respect to the 45^0-line joining $(0,0)$ and $(1,1)$. The endpoints of the stable region extending away from the origin are given by

$$p_0 = q_0 = \frac{R}{1-R}; \quad p_1 = q_1 = \frac{-R + \sqrt{R}}{1-R}.$$

In this situation, it is not difficult to show that if $0 < R < 1$, then (i) $0 < p_0 < p_1 < 1/2$, (ii)

$p_1 \to 0$ as $R \to 0$, and (iii) $p_0 \to 1/2$ as $R \to 1$. Thus, when R is close to 1 and shooting at an opponent quickly damages or destroys assets, the stable power-sharing region approaches the square with corners $(0,0)$, $(0,1/2)$, $(1/2,1/2)$, $(1/2,0)$. Then the players should never shoot, provided neither can eliminate its opponent with probability greater than 1/2.

Of course, the dark side of Model III is that if the players are outside the stable power-sharing region, they have exactly the same motivation to shoot as they do throughout the unit square in Models I and II. At least one player will find it rational to fire at its opponent, propelling both players into a race to preempt. Arresting this race in Model III requires that the players' assets deteriorate rapidly over time and the damage caused by shooting be significant, especially if the marksmanship of the players is relatively good.

5. Conclusions

While our results may be viewed as profoundly pessimistic, we believe there is some basis for optimism. The pessimism stems from our first two models, which say that no matter what the values of any of the parameters are, rational players will be impelled to try to eliminate a partner in a power-sharing agreement.

This incentive to shoot is not mitigated by a player's having a low probability of eliminating an opponent (p or q) or getting the lion's share of the power in the beginning (a or $1 - a$). At least one player, and sometimes both, will have an immediate incentive to try to aggrandize all the power by firing at its opponent in both Models I and II.

True, one player (say, P) may be deterred under some conditions if Q is cooperative, but Q never has this incentive at the same time that P does. Because their deterrence regions never overlap—a necessary condition for power sharing—the players will race to preempt, independent of the parameter values in the model.

Surprisingly, this conclusion is not altered by Model II, in which the Model 1 game is repeated, but with assets reduced by discount factor r in each period of play. While repeated play has worked to inhibit noncooperative behavior in games like Prisoners' Dilemma (Axelrod, 1984), it does not work in the game underlying Model II. Exactly the same incentives to preempt exist as in Model I, except now they apply in each period of play.

One possible way to foster more cooperation would be to impress on the two players that at least one, and possibly both, will end up with nothing if they fire at each other in every period of play. If they are maximin players who wish to avoid their worst outcomes—or if they are risk-averse—it would behoove them to sign a binding and enforceable agreement to desist from shooting at each other at any time. The problem with such an agreement, of course, is finding a way to enforce it when it is always in the interest of the players to break it, perhaps ambiguously or surreptitiously in order to evade detection.

Model III, wherein shooting reduces the assets through damage factor s, provides some grounds for optimism. If the players know they will suffer not only a discounting of assets with the passage of time (introduced in Model II) but also additional damage or destruction of these (Model III), they have a more compelling reason to refrain from shooting.

But they will not always refrain. The discounted damage R, which reflects both the discount factor, r, and the reduction in assets, $1-s$, must be sufficiently high. Only then may the passage of time, and the damage inflicted by shooting, render attempts to eliminate an opponent less appealing, especially when the marksmanship of the players is relatively poor. Note that as long as the shares of the players are equal $(a = 1 - a = 1/2)$, any elimination probability greater than 1/2 (for either player) will induce rational players to preempt.

How can we explain that as the accuracy of ballistic and, later, cruise missiles increased from the 1960s on, the superpowers reached more and more arms-control agreements? One reason was the growing realization that the damage that a nuclear exchange could cause tremendous damage, possibly resulting in a "nuclear winter," which echoes the damage factor in Model III.[12]

But at least as important were improvements in satellite reconnaissance and other verification techniques, which facilitated the detection of treaty violations and eased the enforcement problem. In addition, both sides developed a nearly invulnerable second-strike capability, primarily though submarine-launched missiles, which meant that a first strike could not wipe out an opponent's ability to respond. Devastating as the damage might be, the enemy could strike back with almost the same fury.

Memories of earlier devastation caused by war also matter. The American civil war, the bloodiest in U.S. history, was an event so searing that no serious challenge to the unity of the country has been mounted since. By contrast, when a faction of a divided country believes it can strike a lethal blow against another faction that might challenge it, war becomes more likely. This happened in both Lebanon and the former Yugoslavia in the 1980s and 1990s, with gruesome results for each country.

Other countries have suffered greatly from civil wars in recent times. Often it takes a generation or more of conflict before a situation reaches a point in which leaders say "no more," or outside parties intervene. Unfortunately, this can be a slow and costly way to learn, as has recently been demonstrated in, among other places, Angola, Burundi, the Democratic Republic of the Congo, Liberia, Rwanda, Sierra Leone, Somalia, Sri Lanka, and Sudan, wherein various factions have attempted, through armed conflict, to usurp power.

We see no easy way to speed up the process. Sometimes the damage must be inflicted to sink into collective memories and bring people to their senses, whether the conflict is interpersonal, international, or something in between.

To be sure, not all conflicts are unwarranted. Some marriages turn bad, and divorce

---

[12] In Brams and Kilgour (2007), we show that if payoffs are not accumulated in repeated play but occur only when play terminates—and are diminished if there has been shooting that does not eliminate a player, as in Model III—then the players are more likely to be deterred from shooting. In fact, they will always be deterred if power-sharing is equal $(s = 1 - s = 1/2)$ and their shooting is simultaneous.

is the best option for both spouses. Some corporate battles can be settled only by the ouster of one side if a company is to survive. And some wars, such as World War II, seem morally necessary to rid the world of an evil dictator or an inhumane movement.

But many wars, especially civil, are morally suspect if not bankrupt. They may be fueled by the personal whims of an autocrat, or by ethnic, racial, or religious differences in the populace that become excuses for ferocious fighting that, on occasion, turns into genocide.

It would be desirable if one could demonstrate, in advance, the untoward damage that such wars can cause so that the disputants, anticipating the dire consequences, can pull back in time. While Model III shows that such damage can be an important deterrent to fighting, it does not show how to make it felt either quickly or decisively.

Acknowledgments

We thank Matthew Braham, Frank Steffen, Stephen Haptonstahl, and an anonymous referee for valuable comments. We also benefited from discussions at the Conference on Power: Conceptual, Formal, and Applied Dimensions, Hamburg, Germany, 17-20 August 2006, and the Workshop on Power-Sharing and Democratic Governance in Divided Societies, Centre for the Study of Civil War (CSCW), International Peace Research Institute, Oslo (PRIO), 21-22 August 2006, and other conferences where earlier versions of the paper were presented.

References

Axelrod, Robert (1984). *The Evolution of Cooperation*. New York: Basic Books.

Brams, Steven J., and D. Marc Kilgour (2007). "Stabilizing Power-Sharing." Preprint.

Brams, Steven J., and D. Marc Kilgour (1988). *Game Theory and National Security*. New York: Basil Blackwell.

Collier, Paul, V. L. Elliott, Havard Hegre, Anke Hoeffler, Marta Reynal-Querol, and Nicholas Sambanis (2003). *Breaking the Conflict Trap: Civil War and Development Policy*. Washington, DC: World Bank and Oxford University Press.

Fearon, James D., and David D. Laitin (2003). "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97, no. 1 (February): 75-90.

Garfinkel, Michelle R., and Stergios Skaperdas (2007). "Economics of Conflict: An Overview." In Todd Sandler and Keith Hartley (eds.), *Handbook of Defense Economics*, vol. 2. Amsterdam: Elsevier.

Garfinkel, Michelle R., and Stergios Skaperdas (eds.) (1996). *The Political Economy of Conflict and Appropriation*. New York: Cambridge University Press.

Hirshleifer, Jack (2001). *The Dark Side of the Force: Economic Foundations of Conflict Theory*. Cambridge University Press.

Kadera, Kelly M. (2001). *The Power-Conflict Story*. Ann Arbor, MI: University of Michigan Press.

Karlin, Samuel (1959). *Mathematical Models and Theory in Games, Programming, and Economics*, vol. 2. Reading, MA: Addison-Wesley.

Powell, Robert (1999). *In the Shadow of Power: States and Strategies in International Politics*. Princeton, NJ: Princeton University Press.

Rothchild, Donald, and Philip G. Roeder (eds.) (2005). *Sustainable Peace: Power and Democracy after Civil War*. Ithaca, NY: Cornell University Press.

Walter, Barbara (2002). *Committing to Peace: The Successful Settlement of Civil Wars*. Princeton, NJ: Princeton University Press.

Zagare, Frank C., and D. Marc Kilgour (2000). *Perfect Deterrence*. New York: Cambridge University Press.

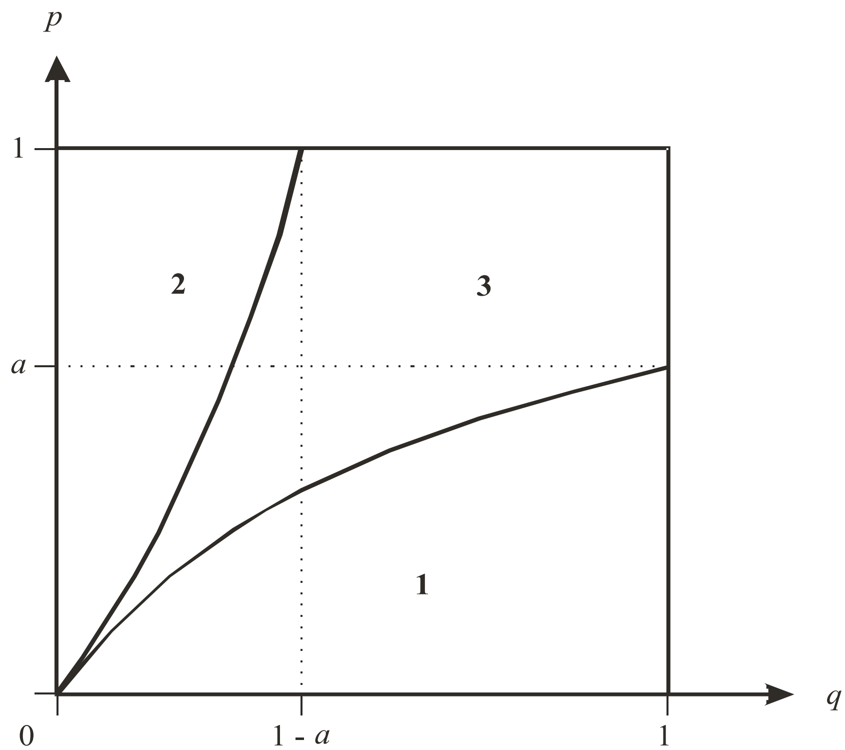

Figure 1. Models I and II

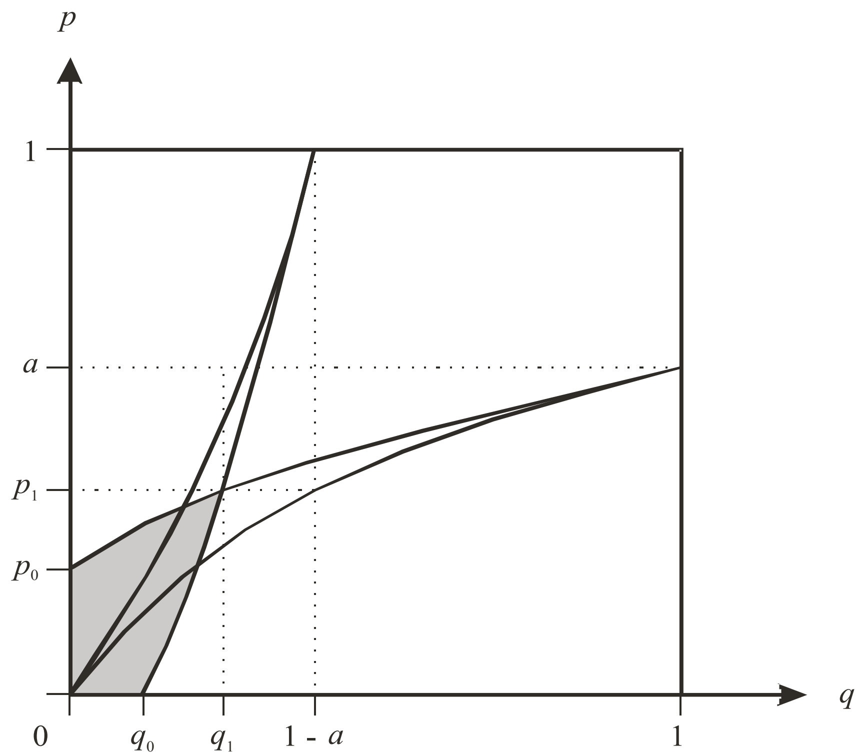

Figure 2. Model III