American Political Science Review (2017) 111, 2, 219–236
doi:10.1017/S0003055416000782

Taking Sides in Wars of Attrition

ROBERT POWELL *University of California, Berkeley*

Third parties often have a stake in the outcome of a conflict and can affect that outcome by taking sides. This article studies the factors that affect a third party's decision to take sides in a civil or interstate war by adding a third actor to a standard continuous-time war of attrition with two-sided asymmetric information. The third actor has preferences over which of the other two actors wins and for being on the winning side conditional on having taken sides. The third party also gets a flow payoff during the fighting which can be positive when fighting is profitable or negative when fighting is costly. The article makes four main contributions: First, it provides a formal framework for analyzing the effects of endogenous intervention on the duration and outcome of the conflict. Second, it identifies a "boomerang" effect that tends to make alignment decisions unpredictable and coalitions dynamically unstable. Third, it yields several clear comparative-static results. Finally, the formal analysis has implications for empirical efforts to estimate the effects of intervention, showing that there may be significant selection and identification issues.

INTRODUCTION

Third parties often have a stake in the outcome of a civil or interstate war and can affect that outcome by taking sides, i.e., by supporting one side or another. Recent civil-war examples include the Sunni awakening of 2007 when Sunni tribal leaders in Iraq sided with the United States against Al-Qaeda (Biddle, Friedman, and Shapiro 2012). Afghan tribal leaders have faced a choice at various times between supporting the government or the Taliban (Berman, Shapiro, Felter 2011; Malkasian 2013). In March 2011, NATO—principally Britain and France with logistic support from the United States—intervened in Libya in support of the opposition trying to oust Colonel Muammar Gaddafi. Iran and then Russia have sided with President Bashar al-Assad in the Syrian civil war while Turkey, the Gulf States, and others have supported opposing factions. Saudi Arabia is supporting the government in the Yemeni civil war. (See Regan (2000) for a survey of foreign intervention in civil wars.) The history of interstate war is also full of examples of third-party interventions. At least one state intervened in an ongoing war in 27 percent of the wars between 1816 and 2007 (Kyle, Ghosn, and Bayer 2013).

What factors affect a third party's decision to take sides? How does intervention affect the duration of the conflict and the chances that one side or the other ultimately prevails? This article studies these questions by adding a third actor to a standard continuous-time war of attrition with two-sided asymmetric information. The third actor has preferences over which of the other two actors wins. It also prefers to be on the winning side if it decides to take sides. Finally, the third party gets a flow payoff for as long as the fighting lasts. This payoff can be negative if fighting entails a net cost, or it can be positive if fighting is profitable. The latter may be the case when ongoing fighting permits the third party to exploit lootable resources, engage in smuggling, or drug trafficking.

This article makes four main contributions. First, it provides a formal framework for analyzing the effects of endogenous intervention on the duration and outcome of the conflict.[1] Existing formal work on duration generally focuses on two-actor models and is therefore silent on issues related to third-party intervention (e.g., Fearon 2004; 2007; Leventoglu and Slantchev 2007; Powell 2012, 2013; Krainin 2014a). By contrast, models of intervention or more generally alliance formation center on the effect that taking sides has on the ultimate outcome. These analyses typically lack a time dimension and consequently cannot address the effect that taking sides has on how long the fighting lasts and the related effect that this has on the cost of fighting and the incentive to intervene (e.g., Gent 2008; Krainin 2014b; Morrow 1991; Powell 1999; Smith 1995; Werner 2000). As a result of these theoretical limitations, empirical work on the effects of intervention on duration and outcome must rely on testing hypotheses that, while theoretically plausible, cannot be derived from an underlying theoretical framework.[2] The first contribution of the present analysis is to begin to develop just such a framework.

Second, the analysis identifies equilibrium pressures that tend to make alignment decisions unpredictable whenever the third party's preference for one side over the other is not too large and fighting is costly. These pressures also tend to make coalitions dynamically unstable with the third party joining one side at a given time and then switching with positive probability to

Robert Powell is Professor, Travers Department of Political Science, University of California, Berkeley, 210 Barrows Hall, UC Berkeley, Berkeley, CA 94708-1150 (RPowell@Berkeley.edu).

I am grateful to Ernesto Dal Bó, James Fearon, John Morgan, Santiago Oliveros, Gerard Padro, Jas Sekhon, Leo Simon, and Noam Yuchtman for helpful comments, criticisms, and discussion. I gratefully acknowledge the support of the National Science Foundation (SES 1456516).

Received: January 26, 2016; revised: December 9, 2016; accepted: December 12, 2016. First published online: February 14, 2017.

1 In keeping with much of the empirical work on intervention, "outcome" refers to which side prevails even though the length of the conflict might also be considered to be part of the outcome.

2 A sampling of important empirical work on the effects of intervention includes Balch-Lindsay, Enterline, and Joyce 2008; Collier, Hoeffler, and Soderbom 2004; Cunningham 2010; Gleditsch 2007; Regan 2000; 2002; Salehyan, Gleditsch, and Cunningham 2011.

the other side at a later time. These results are in keeping with recent work on alliance formation during civil wars that emphasizes the propensity of many groups to change sides during the conflict (Christia 2012; Seymour 2014). Christia, for example, describes recent Afghan history "as replete with examples of warring leaders choosing to switch sides" (2012, 3).

More specifically, the analysis highlights a kind of equilibrium "boomerang effect" on alignment decisions in which the expectation that the third party will support a given side makes it less likely that the third party will actually do so. Suppose, for example, that the two main actors, 1 and 2, expect the third party to support 1 at time T. This expectation creates incentives that induce 1 and 2 to take actions that make the third party less likely to support 1. Similarly, the expectation that the third party will support 2 at T leads 1 and 2 to act in ways that make the third party less likely to support 2. The boomerang effect can be swamped by a strong preference for a given side. If, for example, the third party strongly prefers 1 to prevail in the contest against 2, then the third party is sure to support 1. But if the third party's preference for one side over the other is not too large and if fighting is costly for the third party, then the boomerang effect dominates.

Formally, the boomerang effect leads the third party to play a mixed equilibrium strategy in which it joins each side with a positive probability. Mixed equilibria often lack substantive appeal, but the mechanism underlying the boomerang effect is quite intuitive. Suppose again that 1 and 2 expect the third party to support 1 at time T. This support increases 1's chances of prevailing if the war of attrition continues on after T. Anticipating the higher chance of prevailing, 1 is more likely to fight on until T and 2 is less likely. More precisely, some marginal player types of 1 who would have dropped out prior to T are now willing to fight on. Conversely, some marginal types of 2 who would have fought until T drop out before T because they expect their prospects of prevailing to go down when the third party joins 1. The net effect is that there is a large set of relatively less resolute types of 1 that are still active at T only because they believe that the third party will support 1 at that time. If the third party supports 2 instead, these less resolute types of 1 drop out immediately, thereby shortening the subsequent war of attrition. This in turn creates an incentive for the third party to deviate from supporting 1 when fighting is costly and shorter wars are therefore better.[3] Although described in the context of a war of attrition, this boomerang effect and the mechanism underlying it seem much more general and are likely to be present in any conflict in which uncertainty about how long the other side is willing to fight is a key factor.

The third contribution is to derive several clear comparative-static results which provide formal support for some common claims about conflict. Ross, for example, suggests in his analysis of the resource curse that "resource wealth could lengthen a conflict if it provides funding to the weaker side, helping it equalize the balance of forces" (2006, 282). The more equal the distribution of power in the model, the longer the expected duration of the fighting.

Finally, the formal analysis has implications for empirical efforts to estimate the effects of intervention. These studies often attempt to assess the effect of intervention in support of one side on that side's probability of prevailing by estimating the effects of intervention on the hazard rate of the conflict, i.e., on the probability that that side will prevail in the next instant given that the conflict has not yet ended (e.g., Balch-Lindsay and Enterline 2000; Balach-Lindsay, Enterline, and Joyce 2008; Gent 2008; Regan 2002). This approach is based on the presumption that there is a direct link between the probability that a side will prevail and the hazard rate. The present analysis shows that this is not the case. The chances that a faction prevails depend solely on the ratio of its cost of fighting relative to the other side's while the expected duration and hazard rate depend on both the cost ratio and on the sum of the two actors' costs of fighting. As a result, one cannot infer the effects of intervention on the likelihood that one side will prevail from changes in the hazard rate. In addition to these identification issues, the formal analysis also highlights potential selection effects.

The next section relates the taking-sides model to existing work. The following two sections describe a baseline two-actor, continuous-time war of attrition and then add a third party to this baseline model. The third party chooses between the two alternatives of supporting one side or the other at an exogenously given time $T \geq 0$. Forcing the third party to choose between these two options and not allowing it to decide to stay out by not supporting either side simplifies the analysis and makes it easier to see the key forces at work. Subsequent sections describe the equilibria of the game and the boomerang effect, analyze the comparative statics, and discuss the formal analysis' implications for empirical efforts to estimate the effects of intervention. The last two sections examine extensions. One extension allows the third party to stay out as well as take sides. The other extension adds a second decision time at which the third party can decide whether to switch sides after having previously taken sides.

RELATED WORK

Wars of attrition provide workhorse models for analyzing many different kinds of conflict. Applications include the study of exit in a declining industry (Fudenberg and Tirole 1986; Ghemawat and Nalebuff 1985), strikes (Kennan and Wilson 1989), economic stabilization (Alesina and Drazan 1991; Cassella and Eichengreen 1996), the provision of public goods (Bliss and Nalebuff 1984), lobbying and vote buying (Dekel, Jackson, and Wolinsky 2008; 2009), and international bargaining (Fearon 1998).

Adding a third party to a war of attrition seems like a very natural point of departure for studying intervention. But third-party intervention in a war of

[3] More concisely, the third party's strategy of supporting 1 for sure at T is not sequentially rational.

attrition appears not to have been formally modelled before. The two most relevant studies are Bulow and Klemper (1999) and Cassella and Eichengreen (1996). The former examines two versions of a generalized war of attrition in which $p+m$ players compete for p prizes. In the case of a natural oligopoly of p firms, a firm pays a cost until it exits the market. By contrast, each firm continues to pay a cost for as long as the contest lasts when the firms are competing over setting industry standards since those standards are not set until all of the firms stop fighting. The latter is similar to the present model in that the third party cannot exit the game and its ultimate payoff depends on when the other players quit. A key difference is that the third party's decision affects the distribution of power between the players that are fighting which is not the case in the standards-setting model.

Cassella and Eichengreen (1996) study the effect of aid on stabilization. The aid, which arrives at an exogenously specified time, reduces the stakes by narrowing the gap between winning and losing. Whether or not aid accelerates or delays stabilization depends on when it arrives. Early arrival accelerates stabilization whereas late arrival delays it. The fundamental difference between that model and the present analysis is that what happens at the exogenous date is a strategic decision in the model studied here in that the third party decides what to do.

The present analysis is also related to two other areas: formal work on the causes and conduct of war and on alignment, intervention, and duration. As noted above, existing formal work on alignment usually lacks a dimension of time whereas work on duration generally centers on two-actor models. The present analysis begins to bridge this gap.

Nevertheless, two features of the standard war of attrition probably make it a more natural model of civil rather than interstate intervention. The first is the absence of bargaining. Most recent formal work on the causes and conduct of war frames the issue in terms of bargaining and a related inefficiency puzzle (Fearon 1995; Powell 2006). Why does bargaining break down in costly conflict which is *ex post* inefficient? An important argument in the context of civil wars is that negotiated settlements, especially those calling on one of the factions to disarm, often entail very severe commitment problems (Fearon and Laitin 2008; Walter 2002; 2009). As a result, civil-war outcomes often have the winner-take-all quality of wars of attrition.[4]

Second, a state or faction can fight as long as it chooses to pay the cost. Neither can defeat the other. There are no battles which, if lost, result in the elimination of a state or faction. Nor is it possible for one belligerent to fight so long that the other runs out of resources and the ability to continue.[5]

[4] Abreu and Gul (2000) also show that bargaining itself can resemble a war of attrition when there is a small probability that each bargainer faces a nonstrategic obstinate type. See Acharya and Grillo (2015) for a related model of war.

[5] A goal of future work is to add just such a resource constraint.

Although the model developed here is likely a better fit for civil war intervention, it still may provide useful insights into intervention into interstate conflict. Models often make strong, simplifying assumptions. They leave some important considerations out in order to focus on others. For example, interstate wars occur in the shadow of third parties that have a stake in the outcome, can intervene, and choose to do so in more than a quarter of the cases (Kyle, Ghosn, and Bayer 2013). Yet the canonical models of bargaining and war (e.g., Fearon 1995; Powell 1999) set intervention aside in order to focus on bargaining with two-actor bargaining models. The present approach sets bargaining aside and makes other strong assumptions in order to try to get some traction on the problem of intervention.

THE TWO-ACTOR BASELINE WAR OF ATTRITION

This section reviews some general results about wars of attrition which are needed to analyze the equilibria of the taking-sides game. The section then defines a baseline model by imposing four additional assumptions on a standard model. The first guarantees that the baseline model has a unique equilibrium. The second makes it possible to derive explicit, closed-form expressions for the equilibrium strategies. Having a unique, explicit equilibrium in the two-actor model simplifies the analysis when the third actor is introduced. The third and fourth assumptions are more substantively oriented. They ensure that one of the two main actors is stronger than the other. Taking one actor to be stronger than the other helps connect the present analysis to existing work in two ways. Governments are usually stronger than rebel groups, and some analyses rely on this (e.g., Gent 2008). Still other work frames alignment decisions in terms of whether a third party balances by joining the weaker side or bandwagons by supporting the stronger side (e.g., Christia 2012; Waltz 1979).

In a standard war of attrition, two actors, say 1 and 2, compete for a prize by deciding how long they will fight for it. The longer the fight lasts, the higher each faction's cost. The side that quits first loses. We can think of these actors as states or factions in the present context. More formally, faction 1 pays a marginal cost of c_1 and a total cost of $c_1 t$ if the fight ends at time t. Winning brings a payoff of w_1 while the losing brings zero. If the factions stop at the same time, each wins with probability 1/2.[6] Taking $t_j \geq 0$ to be the time at which j will stop fighting if the game has not already ended by this time and letting $t_j = \infty$ mean that j would fight forever, 1's payoff is

$$U_1(t_1, t_2) = \begin{cases} -c_1 t_1 & \text{if } t_1 < t_2 \\ w_1/2 - c_1 t_1 & \text{if } t_1 = t_2 \\ w_1 - c_1 t_1 & \text{if } t_1 > t_2 \end{cases} \quad \textbf{(1)}$$

[6] The tie breaking is largely irrelevant in that ties occur with probability zero in equilibrium.

and similarly for $U_2(t_1, t_2)$. It proves convenient to reparametrize the model in terms of the cost ratio $\rho_0 \equiv c_1/c_2$ and the total (marginal) cost $k_0 = c_1 + c_2$ via $c_1 = \rho_0 k_0/(1+\rho_0)$ and $c_2 = k_0/(1+\rho_0)$.

Each faction is unsure of the other's payoff to winning. Specifically, 1 believes 2's payoff w_2 is distributed over $(\underline{w}_2, \infty)$ with $\underline{w}_2 \geq 0$ according to the cumulative distribution $G_2(w_2)$. Similarly, 2 believes that w_1 is distributed over $(\underline{w}_1, \infty)$ with $\underline{w}_1 \geq 0$ according to $G_1(w_1)$.[7] A strategy for j is a function $\sigma_j(w_j)$ that specifies the time at which each type $w_j \in (\underline{w}_j, \infty)$ stops.

The key to determining the equilibrium strategies is that a player type stops when the marginal gain from fighting a bit longer is just offset by the marginal cost. The marginal gain for type w_1 is its payoff to winning, namely w_1, times 2's hazard rate, i.e., the probability that 2 will quit in the next instant given that it has not already quit. The marginal cost is c_1. Equating the marginal gain with the marginal cost for w_1 and w_2 yields two differential equations that the equilibrium strategies $\sigma_1^*(w_1)$ and $\sigma_2^*(w_2)$ must satisfy. (See the Appendix for the derivation of these strategies.)

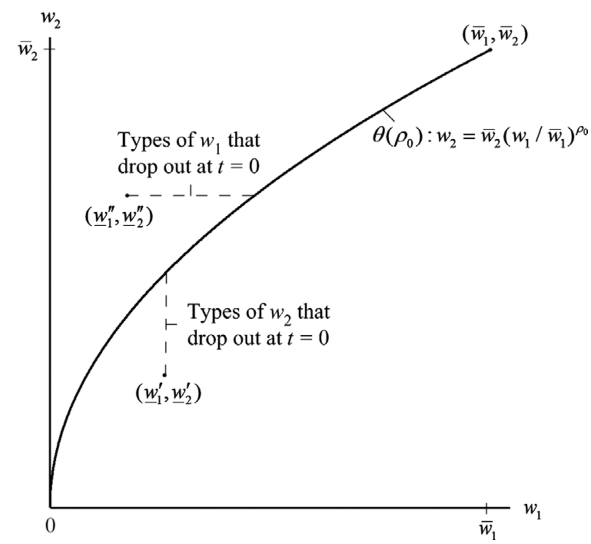
The main equilibrium result is that the higher a state's or a faction's payoff to winning, i.e., the larger w_j, the longer the faction fights. More formally, $\sigma_1^*(w_1)$ and $\sigma_2^*(w_2)$ are continuous and strictly increasing at any $\sigma_j^*(w_j) > 0$. There are however infinitely many equilibria, and it is generally impossible to derive explicit closed-form expressions for the equilibrium strategies.[8]

We make two assumptions to simplify the baseline model as much as possible before adding the third party. Assume first that there is an arbitrarily small probability that each faction is nonstrategic and fights forever and, second, that each faction believes the other faction's payoffs are exponentially distributed. The first assumption ensures that the two-actor baseline game has a unique equilibrium; the second makes it possible to derive explicit expressions for the factions' strategies.[9] Formally, assume that w_j is distributed exponentially with $G_j(w_j) = 1 - e^{\underline{w}_j - w_j}$ and that there exists a $\overline{w}_j$ such that all $w_j > \overline{w}_j$ are nonstrategic types that fight forever.

It follows that the factions' unique equilibrium stop times when $\underline{w}_1 = \underline{w}_2 = 0$ are

$$\sigma_1^*(w_1) = \frac{\overline{w}_1 \overline{w}_2}{k_0} \left(\frac{w_1}{\overline{w}_1} \right)^{1+\rho_0},$$

FIGURE 1. Equilibrium Play in the Baseline Two-actor War of Attrition



$$\sigma_2^*(w_2) = \frac{\overline{w}_1 \overline{w}_2}{k_0} \left(\frac{w_2}{\overline{w}_2} \right)^{1+1/\rho_0}. \qquad \textbf{(2)}$$

(See Equation (A3) for the more general expressions when $\underline{w}_1 \geq 0$ and $\underline{w}_2 \geq 0$.)

The types of 1 and 2 that stop at the same time play an important role in the analysis. Solving $\sigma_1^*(w_1) = \sigma_2^*(w_2)$ from Equations (2) or (A3) shows that w_1 and $w_2 = \overline{w}_2 (w_1/\overline{w}_1)^{\rho_0}$ stop at the same time. The curve $\theta(\rho_0)$ in Figure 1 traces out these points for $0 \leq w_1 \leq \overline{w}_1$.

Figure 1 also illustrates the equilibrium dynamics. To see how the game unfolds over time, suppose that the pair of lowest-payoff types $(\underline{w}_1, \underline{w}_2)$ is below or to the right of $\theta(\rho_0)$, e.g., at $(\underline{w}_1', \underline{w}_2')$. Then the first thing that happens in equilibrium is that all types w_2 between $\underline{w}_2'$ and $\overline{w}_2 (\underline{w}_1'/\overline{w}_1)^{\rho_0}$ drop out at $t = 0$. Graphically, play moves vertically up from $(\underline{w}_1', \underline{w}_2')$ to $\theta(\rho_0)$. Equilibrium play thereafter moves along $\theta(\rho_0)$ with w_1 and $w_2 = \overline{w}_2 (w_1/\overline{w}_1)^{\rho_0}$ dropping out at the same time until $\overline{w}_1$ and $\overline{w}_2$ simultaneously quit. If, by contrast, the lowest-payoff types are above or to the left of $\theta(\rho_0)$ at, say, $(\underline{w}_1'', \underline{w}_2'')$, then play moves horizontally from $(\underline{w}_1'', \underline{w}_2'')$ to $\theta(\rho_0)$ with types w_1 between $\underline{w}_1''$ and $\overline{w}_1 (\underline{w}_2''/\overline{w}_2)^{1/\rho_0}$ dropping out at $t = 0$. Subsequent play moves along $\theta(\rho_0)$.

Which player types drop out at the start of the game ($t = 0$) and why they do turns out to be crucial for understanding the equilibria of the three-actor game. A rough intuition begins with the observation that the lowest-payoff type of 2, $\underline{w}_2$, is relatively small compared to the lowest-payoff type of 1, $\underline{w}_1$, at points below or to the right of $\theta(\rho_0)$. As a result, low-payoff types of 2 are so pessimistic about their chances of prevailing when $(\underline{w}_1, \underline{w}_2)$ is below $\theta(\rho_0)$ that they drop out immediately. The converse holds when the lowest-payoff types are

[7] Not including $\underline{w}_j$ in the type space and using the interval $(\underline{w}_j, \infty)$ instead of $[\underline{w}_j, \infty)$ evades a technicality discussed in the derivation of the equilibrium of the baseline model in the Appendix and, especially, footnote 31.

[8] For an analysis of wars of attrition, see Nalebuff and Riley (1985) and Fudenberg and Tirole (1986).

[9] Fudenberg and Tirole (1986) postulate a small probability of these nonstrategic types in their model of exit from a declining industry. The formal import of this assumption is that it provides an additional boundary condition on the system of differential equations defined by the first-order conditions. This condition, which is that the highest-payoff strategic types must stop at the same time, pins down a unique solution. Ponasati and Sakovics (1995) show that types willing to fight forever arise quite naturally when players discount the future. However, allowing for explicit discounting when there are three actors makes the analysis much more complicated.

above θ: the low-payoff types of 1 are so pessimistic about their chances that they quit at $t = 0$. Finally, if $(\underline{w}_1, \underline{w}_2)$ is on $\theta(\rho_0)$, the only types dropping out at $t = 0$ are $\underline{w}_1$ and $\underline{w}_2$, and the probability that either faction drops out is zero.[10]

It follows that the probability that a faction prevails in a war of attrition depends on the cost ratio but not on the total cost.[11] For example, the probability that 1 prevails is the probability that it stops after 2, i.e., that $\sigma_1^*(w_1) > \sigma_2^*(w_2)$. This in turn is the probability (w_1, w_2) lies below $\theta(\rho_0)$. Since $\theta(\rho_0)$ depends on ρ_0 but not k_0, the probability that 1 wins depends on ρ_0 but not on k_0. Both factions quit sooner when the total cost is higher as is clear from Equation (2) for the case when $\underline{w}_1 = \underline{w}_2 = 0$ or Equation (A3) for the more general case. But a higher total cost does not affect the chances of prevailing as long as the cost ratio remains the same.

As noted above, the government is often stronger than the rebel group and assumed to be so in some work (e.g., Balach-Lindsay, Enterline, and Joyce 2008; Gent 2008; Regan 2002). Other work frames third-party alignment decisions in both civil and interstate conflict in terms of a tradeoff between balancing or bandwagoning (e.g., Aydin and Regan 2008; Christia 2012; Powell 1999; Waltz 1979). A third party "balances power" when it joins, aligns with, or supports the weaker side. It "bandwagons" when it supports the stronger side. A third party faces a tradeoff between balancing and bandwagoning when it prefers the weaker side to win. Bandwagoning maximizes its chances of being on the winning side and obtaining the benefits that come with that. Balancing by contrast maximizes a third party's payoff conditional on being on the winning side.

In light of this work, it is useful to make one faction stronger than the other in the baseline model. Most simply, one faction is stronger than the other when it is more likely to prevail in the war of attrition. To formalize this, let $\Pi_j(z_1, z_2, \rho)$ denote the probability that j prevails in the war of attrition with cost ratio ρ given that at least one faction is strategic and that z_1 and z_2 are the lowest-payoff types still active, i.e., all $w_1 \leq z_1$ and $w_2 \leq z_2$ have already dropped out.[12] Then j is stronger than i at (z_1, z_2) when $\Pi_j(z_1, z_2, \rho_0) > \Pi_i(z_1, z_2, \rho_0)$ or equivalently when $\Pi_j(z_1, z_2, \rho_0) > 1/2$.

As time passes and equilibrium play moves along θ, the probability that 1 prevails given that neither side has quit varies. Two assumptions ensure that 1 is always stronger than 2, i.e., everywhere along θ. Assume first that the cost ratio favors 1: $c_1 < c_2$ or $\rho_0 < 1$. This implies that 1 is stronger than 2 in an otherwise symmetric war of attrition where $\underline{w}_1 = \underline{w}_2$ and $\overline{w}_1 = \overline{w}_2$. It implies in the asymmetric case that 1's probability of prevailing decreases as time goes on and play moves along θ. Indeed, the probability that 1 prevails decreases to $[1 + \overline{w}_1/(\rho_0\overline{w}_2)]^{-1}$ in the limit as all strategic types drop out and play approaches $(\overline{w}_1, \overline{w}_2)$ along θ. When this limit is less than 1/2, faction 1 may be stronger than 2 in the early phase of the war of attrition and then become weaker in the latter phase. The assumption that $\rho_0 > \overline{w}_1/\overline{w}_2$ prevents this by guaranteeing that this limit is larger than 1/2 and therefore that faction 1 remains stronger than 2 throughout the conflict. Lemma 1 summarizes these results (see the Online Appendix for the proof).

Lemma 1 *If $\overline{w}_1/\overline{w}_2 < \rho_0 < 1$, then 1 is more powerful than 2 as time passes and play moves along θ: 1's chances of prevailing decline throughout the war of attrition and converge to $[1 + \overline{w}_1/(\rho_0\overline{w}_2)]^{-1} > 1/2$.*

Finally, it is important to note how changes in the cost ratio and in the total cost affect the outcome and duration of the conflict as these effects play a major role in determining which side the third party supports. The more the cost ratio favors one side in a war of attrition, the more likely that side is to prevail. To see why, note that as the cost ratio shifts in 1's favor (i.e., as $\rho_0 = c_1/c_2$ declines), $\theta(\rho_0)$ bows up, the area under it increases, and the probability that 1 prevails goes up. Similarly, the probability that 2 prevails increases as its relative cost of fighting goes down, i.e., Π_2 is increasing in ρ_0. And, as noted above, $\theta(\rho_0)$ is independent of k_0, so shifts in the total cost have no effect on either faction's chances of winning as long as the cost ratio remains constant.

Turning to duration, the more symmetric the two factions' costs, i.e., the closer ρ_0 is to one, the longer the fight. The model thus provides formal support for the familiar idea that "making the strength of the two parties more nearly equal" tends to make for longer conflicts (Deutsch 1964). This idea, along with the presumption that the government is generally stronger than the opposition, underlies several empirical studies and leads to the hypothesis that intervention on the government's side tends to shorten conflicts whereas intervention on the rebels' side tends to make them longer.

The total cost also affects the duration. As noted above, a higher cost k_0 induces all types to stop sooner. This makes for shorter fights. Lemma 2 summarizes the results.

Lemma 2 (Baseline Effects of Changes in the Relative and Total Costs of Fighting): *The lower a faction's relative cost of fighting, the greater its chances of prevailing ($\partial\Pi_1/\partial\rho < 0$ and $\partial\Pi_2/\partial\rho > 0$ when $\rho < 1$). The more symmetric the cost ratio, i.e., the closer the cost ratio is to 1, the longer the expected duration. Changes in the total cost do not affect either faction's chances of prevailing, but the expected duration is decreasing in the total cost.*[13]

[10] More formally, the probability that j drops out at $t = 0$ is the probability that its payoff is $\underline{w}_j$ or less which is $G_j(\underline{w}_j)$ and equal to zero when $\underline{w}_1$ and $\underline{w}_2$ are on $\theta(\rho_0)$.

[11] That the probability of winning is determined solely by the cost ratio is a general property of wars of attrition and does not depend on assuming an exponential distribution of types. See Equation (9) in Nalebuff and Riley (1985).

[12] Conditioning on at least one type being strategic means that one faction is sure to prevail, i.e., $\Pi_1(z_1, z_2, \rho) + \Pi_2(z_1, z_2, \rho) = 1$.

[13] See the Online Appendix for some technical qualifications and the proof.

A MODEL OF TAKING SIDES

This section describes a game in which a third party, M, has preferences over which side prevails in the war of attrition between 1 and 2 and can affect the outcome by taking sides. More specifically, M can decide at an exogenously specified time $T \geq 0$ to support 1 or 2. Support for j shifts the distribution of power in j's favor and makes j more likely to win. More precisely, M's support for j shifts the cost ratio in j's favor in the continuation game between 1 and 2 that follows M's decision. As will be seen, this continuation game is itself a war of attrition, and Lemma 2 implies that the favorable shift in j's cost ratio increases its chances of prevailing.

To specify the effects of M's taking sides, let $c_{j|n}$ denote j's cost of fighting if M supports n at T. For example, 1's cost of fighting after T is $c_{1|2}$ if M aligned with 2 and $c_{1|1}$ if M aligned with 1. Take $\rho_n \equiv c_{1|n}/c_{2|n}$ to be the cost ratio between 1 and 2 if M joins n, and let $k_n = c_{1|n} + c_{2|n}$ be the total (marginal) cost of fighting for 1 and 2. Then make the following assumption:

Assumption 1 (The effects of Taking Sides): *Assume (i)* $\rho_1 < \rho_0 < \rho_2$ *and (ii)* $\overline{w}_1/\overline{w}_2 < \rho_1$ *and* $\rho_2 < 1$.

Part (i) formalizes the idea that M's support for j shifts the cost ratio in j's favor. If M aligns with 1, the cost ratio drops from ρ_0 to ρ_1 which makes 1 more likely to prevail (Lemma 2). Joining 2 increases the cost ratio from ρ_0 to ρ_2 and makes 2 more likely to prevail. This is the only part of the assumption needed to characterize the equilibria.

Part (ii) ensures that 1 is stronger than 2 regardless of M's actions. Part (ii) thus excludes the possibility that the third party can shift the distribution of power in 2's favor. Allowing for this would not change the equilibrium analysis but would add many more cases to the comparative-static analysis. Note that no assumption is being made about how M's support affects the total cost of fighting, i.e., whether k_j is larger or smaller than k_0.

Faction 1's payoffs in the three-actor game are still given by Equation (1) if the game ends prior to T. To specify its payoff if the game ends at or after T, suppose for example that M joins 1 at T and that the game subsequently ends at $t \geq T$ when 2 quits. Then 1's payoff is $w_1 - c_1 T - c_{1|1}(t - T)$, and 2's is $-c_2 T - c_{2|1}(t - T)$. More concisely, suppose the game ends at $t \geq T$, the two factions' stop times are t_1 and t_2, and M supports 1. The 1's payoff if M joins n at T is

$$U_{1|n}(t_1, t_2) = \begin{cases} -c_1 T - c_{1|n}(t_1 - T) & \text{if } t_1 < t_2 \\ w_1/2 - c_1 T - c_{1|n}(t_1 - T) & \text{if } t_1 = t_2 \\ w_1 - c_1 T - c_{1|n}(t_2 - T) & \text{if } t_1 > t_2 \end{cases} \quad \textbf{(3)}$$

with analogous payoffs for 2. As in the baseline war of attrition, each faction is uncertain of the other's payoff to winning and believes it to be exponentially distributed.

Turning to M's payoffs, M gets v_1 if 1 wins and v_2 if 2 prevails. In addition to caring about who wins, M may also prefer to be on the winning side if it takes sides. That is, M's payoff if it aligns with j at T is $v_j + \gamma$ if j subsequently prevails and $v_i - \gamma$ if $i \neq j$ eventually wins where $\gamma \geq 0$ is the premium for being on the winning side. Note that if M prefers the weaker side 2 to prevail ($v_2 > v_1$), then M faces a tradeoff between bandwagoning by joining the stronger faction and balancing by aligning with the weaker side.

That longer conflicts are more costly for the protagonists is inherent in the notion of a war of attrition. Third parties, however, may have very different incentives as Deutsch (1964), Balch-Lindsay and Enterline (2000), and Cunningham (2010) emphasize. Sometimes the third parties profit from conflict and are better off the longer it lasts. In still other cases, fighting imposes costs on the third party as well as the two protagonists. Formally, let f_0 denote M's flow payoff prior to taking sides, and take f_j to be M's net flow payoff after supporting side j. M receives this flow for as long as the conflict lasts. Possible benefits accruing during the fighting include the gains from being able to exploit lootable resources, engage in smuggling or drug trafficking, or the direct side payments from j in return for M's support.[14] Possible costs include coming under more intense attack from the other faction, being sanctioned by other third parties, the cost of supplying arms or material to j, etc. The key distinction between f_j and v_j is that the former is contingent on the continuation of the fighting. M gets f_j for as long as the conflict lasts. This flow stops when a faction wins at which point M gets a final payoff of v_1 or v_2.

When $f_j < 0$, supporting j comes at a net cost which grows as the contest continues. In these circumstances, M prefers shorter conflicts to longer ones. By contrast, supporting j brings a positive flow of benefits when $f_j > 0$ and M prefers longer fights.[15] For example, the Central Intelligence Agency's history of the Yugoslav conflict reports that

> it was in the interest of the Serb republic that its two enemies, the Croats and the Muslims, continue to fight each other as long as possible. The Bosnia Serbs therefore consciously set out to provide military support for both sides, depending on the military balance in a given sector, ...most often to the side that was in the weaker positions, thus prolonging the fighting and increasing the cost to both sides" (2002, 179–80).

In sum, M's payoff if j prevails at $t < T$ is $v_j + f_0 t$. M's payoff if it supports j at T and j subsequently wins at $t \geq T$ is $v_j + \gamma + f_0 T + f_j(t - T)$. If the other faction wins at $t \geq T$, M gets $v_i - \gamma + f_0 T + f_j(t - T)$.

M has no private information about the factions. Its prior belief is that each faction's payoff to prevailing is

[14] See for example Fearon 2004; Findley and Marineau 2014; Humphreys 2005; Lujala, Gleditsch and Gilmore 2005; Ross 2004; 2006; Seymour 2014.

[15] Balach-Lindsay and Enterline (2000), Balch-Lindsay, Enterline and Joyce (2008), and Cunningham (2010) emphasize that international actors may pursue more general goals that effectively prolong the fighting. The analysis underlying their empirical studies centers on the greater difficulty of achieving negotiated settlements. A limitation of the war-of-attrition model is that there are no negotiated settlements.

distributed according to $G_j(w_j)$. M also believes that $w_j > \overline{w}_j$ are nonstrategic and fight forever.[16]

THE EQUILIBRIA

This section provides a less technical overview of the perfect Bayesian equilibria (PBEs) of the game. (The formal derivation is in the Appendix.) An equilibrium of the taking-sides model turns out to have a very simple structure composed of three phases. The two factions fight a war of attrition with cost ratio ρ_0 and total cost k_0 during the first phase. There follows an interval of pure fighting during which neither side ever quits. These first two phases last until T at which point M decides whether to support 1 or 2. The third phase follows and is again a war of attrition but now with the cost ratio and total cost determined by M's actions. If for example M supports 1 at T, then 1 and 2 fight a war of attrition with cost ratio ρ_1 and total cost k_1 during the last stage of the game.

To describe these phases more precisely, let (z_1, z_2) denote the pair of lowest-payoff types still active at T, i.e., all $w_1 < z_1$ and $w_2 < z_2$ will have dropped out during the first two phases before M decides what to do at T. Since M cannot intervene until t, all types that drop out prior to T are effectively playing a war of attrition with each other akin to the baseline war of attrition but with z_1 and z_2 rather than $\overline{w}_1$ and $\overline{w}_2$ being the highest-payoff strategic types.[17] Substituting z_1 and z_2 for $\overline{w}_1$ and $\overline{w}_2$ in Equation (2) gives the stop times for $w_1 < z_1$ and $w_2 < z_2$ during this phase:

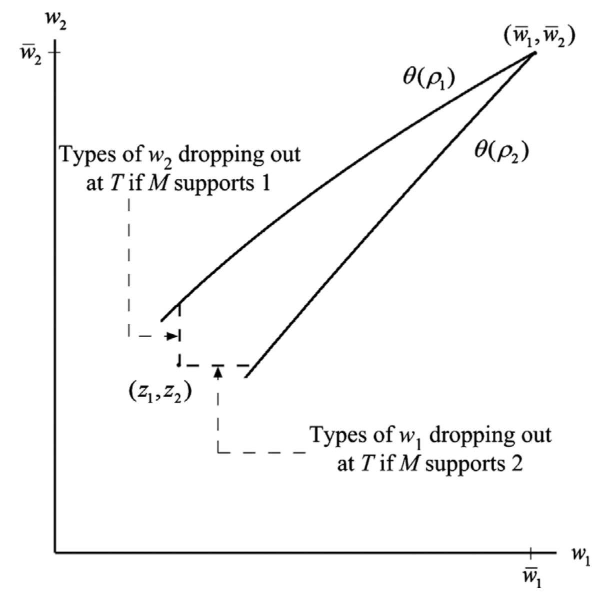
$$\sigma_1^*(w_1) = \frac{z_1 z_2}{k_0}\left(\frac{w_1}{z_1}\right)^{1+\rho_0} \quad \text{and}$$

$$\sigma_2^*(w_2) = \frac{z_1 z_2}{k_0}\left(\frac{w_2}{z_2}\right)^{1+1/\rho_0}, \qquad \textbf{(4)}$$

where we continue to assume $\underline{w}_1$ and $\underline{w}_2$ are zero. The first phase lasts until all of the $w_j < z_j$ drop out or until time $z_1 z_2 / k_0$.[18]

The second phase is an interval of pure fighting of length $\lambda \geq 0$ lasting from $T - \lambda$ to T during which 1 and 2 never quit, i.e., no types drop out. An interval of pure fighting never occurs in a standard two-actor war of attrition. But it arises naturally in the three-actor game whenever M mixes, i.e., M supports 1 and 2 with positive probabilities.[19] The length of this interval is a function of z_1 and z_2 and will often be written as $\lambda(z_1, z_2)$.[20]

FIGURE 2. The Continuation Game when *M* Takes Sides



The third phase follows M's decision and is a war of attrition with the cost ratio and total cost determined by M's action. If for instance M supports 1, then the continuation game between 1 and 2 following M's decision is a war of attrition with cost ratio ρ_1 and total cost k_1. Types z_1 and z_2 are the lowest-payoff types during this phase and play a role analogous to $\underline{w}_1$ and $\underline{w}_2$ in the baseline two-actor war of attrition. The factions stop times in this phase are given by the more general expressions for σ_1^* and σ_2^* in the Appendix for the case in which the lowest-payoff types are not zero (see Equation (A3)).

Characterizing the equilibrium strategies of the taking-sides model thus amounts to finding z_1 and z_2 as well as the probability that M supports 1 at (z_1, z_2) which will be denoted by α_1. The pair (z_1, z_2) must satisfy two conditions, an incentive constraint for M and a timing constraint.

The incentive constraint follows from reasoning backwards from the end of the game, i.e., from the war of attrition played during the third phase following M's decision. The dynamics of this stage are illustrated in Figure 2. Recall that payoff types w_1 and $w_2 = \overline{w}_2(w_1/\overline{w}_1)^{\rho_0}$ quit at the same time in the baseline two-actor war of attrition when the cost ratio is ρ_0. The curves $\theta(\rho_1)$ and $\theta(\rho_2)$ depict the pair types (w_1, w_2) that stop at the same time in the two-actor baseline

[16] These beliefs pose a technical issue. The nonstrategic types were introduced in the baseline two-actor war of attrition so that the game would have a unique equilibrium. This however means that M believes in the three-actor game that the fight will last forever with positive probability, namely, the probability that both 1 and 2 are nonstrategic. As a result, M's payoffs are unbounded. We finese this issue by assuming that M believes that at least one of the factions it is facing is strategic (see the Appendix for details).

[17] Formally, z_1 and z_2, like $\overline{w}_1$ and $\overline{w}_2$, have to quit at the same time, and this provides a boundary condition pinning down a unique solution to the system of two differential equations defining the equilibrium strategies.

[18] Taking the limit of w_j's stop time as w_j approaches z_j from below gives $\lim_{w_j \to z_j^-} \sigma_j^*(w_j) = z_1 z_2 / k_0$.

[19] See the discussion of Result 1 below.

[20] Roughly, types z_1 and z_2 must be indifferent between stopping at $T - \lambda$ and just after T. Otherwise they would not be the weakest types still active at T. An expression for λ follows this indifference (see the Appendix).

model when the cost ratios are ρ_1 and ρ_2 respectively. That is, $w_2 = \overline{w}_2(w_1/\overline{w}_1)^{\rho_j}$ along $\theta(\rho_j)$.

To see how the third phase plays out, assume that (z_1, z_2) is between $\theta(\rho_1)$ and $\theta(\rho_2)$ as in Figure 2 and that M supports 1 at (z_1, z_2). (Lemma 4 in the Appendix shows that (z_1, z_2) must lie on or between $\theta(\rho_1)$ and $\theta(\rho_2)$ in any equilibrium.) Since (z_1, z_2) is below $\theta(\rho_1)$, the low-payoff types of 2 are so pessimistic about their chances of prevailing when the cost ratio is ρ_1 that they drop out as soon as M joins 1. Play moves vertically up from (z_1, z_2) to $\theta(\rho_1)$ with $w_2 \in [z_2, \overline{w}(z_1/\overline{w}_1)^{\rho_1}]$ quitting immediately after M joins 1. Subsequent play then moves along $\theta(\rho_1)$.

Suppose instead that M joins 2 at (z_1, z_2) and the cost ratio becomes ρ_2. The pair (z_1, z_2) is now above $\theta(\rho_2)$ and the low-payoff types of 1 are so pessimistic in light of this cost ratio that they drop out immediately after M aligns with 2. Play moves horizontally from (z_1, z_2) to $\theta(\rho_2)$ at T and then along $\theta(\rho_2)$ for the rest of the game.

Given that the continuation game following M's decision is a war of attrition, we can determine the factions' strategies in that war and then use them to calculate M's payoffs to supporting 1 or 2 at (z_1, z_2). Let $S_1(z_1, z_2)$ and $S_2(z_1, z_2)$ respectively denote M's payoffs to supporting 1 or 2 and define $\Delta_{12}(z_1, z_2) \equiv S_1(z_1, z_2) - S_2(z_1, z_2)$ to be the difference between these payoffs.[21] Then M weakly prefers supporting 1 to aligning with 2 at (z_1, z_2) when the payoff to supporting 1 is at least as large as the payoff to supporting 2 or $\Delta_{12}(z_1, z_2) \geq 0$. M weakly prefers 2 when $\Delta_{12}(z_1, z_2) \leq 0$ and is indifferent between supporting 1 or 2 when $\Delta_{12}(z_1, z_2) = 0$.

M's incentive constraint now follows. Result 1 (stated as Lemma 4 in the Appendix) establishes that (z_1, z_2) must be on or between $\theta(\rho_1)$ and $\theta(\rho_2)$. Moreover, M must support 1 for sure when (z_1, z_2) is on $\theta(\rho_1)$, support 2 for sure when (z_1, z_2) is on $\theta(\rho_2)$, and support both factions with positive probabilities $(0 < \alpha_1 < 1)$ when (z_1, z_2) is between $\theta(\rho_1)$ and $\theta(\rho_2)$. In order for these actions to be consistent with equilibrium play, M must at least weakly prefer supporting 1 when (z_1, z_2) is on $\theta(\rho_1)$, i.e., $\Delta_{12}(z_1, z_2) \geq 0$ when (z_1, z_2) is on $\theta(\rho_1)$. M must also weakly prefer 2 if (z_1, z_2) is on $\theta(\rho_2)$, and M must be indifferent between 1 and 2 when (z_1, z_2) is between $\theta(\rho_1)$ and $\theta(\rho_2)$. Collectively these conditions on $\Delta_{12}(z_1, z_2)$ describe M's incentive constraint.

Result 1 *(M's incentive constraint)*: *Let (z_1, z_2) denote the weakest types still active at T in a PBE. Then (z_1, z_2) must be on or between $\theta(\rho_1)$ and $\theta(\rho_2)$ and*

i. *M must support 1 for sure $(\alpha_1 = 1)$ and, necessarily, $\Delta_{12}(z_1, z_2) \geq 0$ when (z_1, z_2) is on $\theta(\rho_1)$;*
ii. *M must support 2 for sure $(\alpha_1 = 0)$ and $\Delta_{12}(z_1, z_2) \leq 0$ when (z_1, z_2) is on $\theta(\rho_2)$;*
iii. *M must support 1 and 2 with positive probability $(0 < \alpha_1 < 1)$ and $\Delta_{12}(z_1, z_2) = 0$ when (z_1, z_2) is between $\theta(\rho_1)$ and $\theta(\rho_2)$.*

Result 1 is the key to characterizing the equilibria, and it is worth sketching the argument underlying part (iii). Claims (i) and (ii) follow from very similar arguments.

Proof (sketch): We argue by contradiction. That is, we first assume that M is sure to support 1, i.e., $\alpha_1 = 1$, and then show that this leads to a contraction and hence to the conclusion that α_1 must be less than 1.

The first step in reaching a contradiction is to demonstrate that there is an interval $(T - \lambda, T)$ for a $\lambda > 0$ during which 1 never quits, i.e., no player type w_1 quits. We again argue by contradiction by assuming that there is no such interval and then demonstrate that this assumption this leads to a contradiction. More specifically, suppose that there are types w_1 that do quit arbitrarily close to T. This leads to the contradiction that some of these types could profitably deviate from their equilibrium strategy of quitting before T by waiting until T to quit.

Suppose in particular that $\widehat{w}_1$'s equilibrium strategy is to quit at $T - \varepsilon$ where $\varepsilon > 0$ is arbitrarily small. Now compare $\widehat{w}_1$'s payoff to quitting at $T - \varepsilon$ to its payoff to waiting until T to quit. The net payoff to quitting at $T - \varepsilon$ is zero: there are no further loses but there is also no chance of securing future gains when the other quits. If by contrast $\widehat{w}_1$ waits to quit until T, then M by assumption is sure to support 1 and as a result all w_2 between z_2 and $\theta(\rho_1)$ quit. This implies that $\widehat{w}_1$'s payoff to waiting until T to stop is at least $\widehat{w}_1\alpha_1[e^{\underline{w}_2 - z_2} - e^{\underline{w}_2 - \overline{w}_2(z_1/\overline{w}_1)^{\rho_1}}] - c_1\varepsilon$ where $\alpha_1 = 1$ by assumption and the expression in brackets is the prior probability that the w_2 between z_2 and $\theta(\rho_1)$ drop out.[22] The expression in brackets is sure to be positive because (z_1, z_2) is between $\theta(\rho_1)$ and $\theta(\rho_2)$ and, consequently, the payoff to this deviation is sure to be positive for ε small enough. This contradiction ensures that no w_1 quits in $(T - \lambda, T)$ for a $\lambda > 0$.

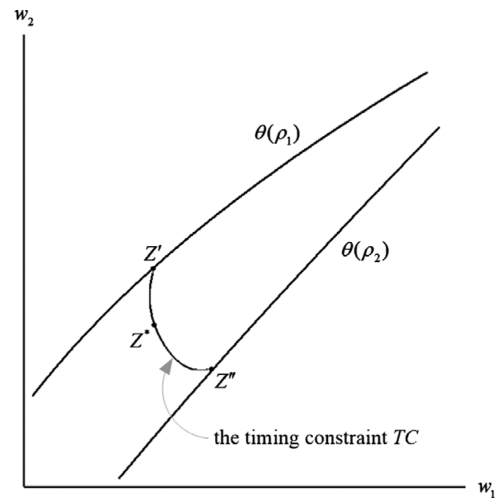
Given that no w_1 quits in $(T - \lambda, T)$ and $\alpha_1 = 1$, the w_2 between z_2 and $\theta(\rho_1)$ that quit at T have no chance of winning after time $T - \lambda$. As a result, they would have done strictly better by quitting at $T - \lambda$ than they do by fighting on until T. Having a profitable deviation from an equilibrium strategy is a contradiction, and it follows from this contradiction that α_1 must be less than one. An analogous argument yields $\alpha_1 > 0$. Putting positive probably on supporting both 1 and 2 implies that M must be indifferent between them and $\Delta_{12}(z_1, z_2) = 0$. □

The pair (z_1, z_2) must also satisfy a timing constraint. The length of the war of attrition during the first phase plus the length of the interval of pure fighting must sum to T. As noted above, the war of attrition between types $w_1 < z_1$ and $w_2 < z_2$ ends at time

[21] The expression for $S_j(z_1, z_2)$ is given in the Appendix.

[22] This lower bound may understate $\widehat{w}_1$'s payoff to quitting at T in two ways. First, it assumes that no w_2 drop out between $T - \varepsilon$ and T. Second, it uses the prior probability that w_2 is between z_2 and $\theta(\rho_1)$ rather than the conditional probability that w_2 is between z_2 and $\theta(\rho_1)$ given that play has lasted until $T - \varepsilon$.

FIGURE 3. The Timing Constraint



$z_1 z_2/k_0$. Hence, (z_1, z_2) must satisfy the timing constraint $\sigma_1^*(z_1) + \lambda(z_1, z_2) = T$ or $z_1 z_2/k_0 + \lambda(z_1, z_2) = T$. The points that do are illustrated in Figure 3.

At least one pair (z_1, z_2) must satisfy both constraints. To verify this, suppose that Z' in Figure 3 does not satisfy both constraints. Then M's incentive constraint must fail to hold. This implies that $\Delta_{12}(Z') < 0$ or, more substantively, that M strictly prefers supporting 2. Suppose further that Z'' does not satisfy both constraints and thus that M's incentive constraint fails to hold here too. This means that $\Delta_{12}(Z'') > 0$. But if $\Delta_{12}(Z'') > 0$ and $\Delta_{12}(Z') < 0$, then continuity ensures that $\Delta_{12}(z_1, z_2)$ must be zero and hence that M's incentive constraint must be satisfied somewhere along the timing constraint between Z' and Z''.

Proposition 1 (which is formally stated in the Appendix) shows that a unique PBE is associated with each (z_1, z_2) that satisfies the two constraints.[23] Suppose for example that $Z^* = (z_1^*, z_2^*)$ in Figure 3 satisfies both constraints. Then $w_1 < z_1^*$ and $w_2 < z_2^*$ play a war of attrition with cost ratio ρ_0 and total cost k_0. This first phase ends at $z_1^* z_2^*/k_0$. There follows an interval of pure fighting lasting until T at which point M supports 1 with probability α_1. (See Proposition 1 for the expression for α_1.) If M supports j at T, then $w_1 > z_1^*$ and $w_2 > z_2^*$ fight a war of attrition with cost ratio ρ_j and total cost k_j during the final phase.

A numerical example helps fix ideas. Suppose that the initial cost ratio is $\rho_0 = 1/2$, decreases to $\rho_1 = 3/7$ if M joins 1, and increases to $\rho_2 = 9/11$ if M joins 2. To keep things simple, the total cost is assumed to remain the same with $k_0 = k_1 = k_2 = 1$. Taking $\underline{w}_1 = \underline{w}_2 = 0$, $\overline{w}_1 = 5$, and $\overline{w}_2 = 20$ satisfies Assumption 1 with $\overline{w}_1/\overline{w}_2 < \rho_1 < \rho_0 < \rho_2 < 1$. M's payoff if 1 prevails is $v_1 = -5$ and $v_2 = 5$ if 2 prevails. The payoff to being on the winning side conditional on having taken sides is $\gamma = 1$. M's flow payoff prior to taking sides is $f_0 = -20$ which means fighting is costly and shorter fights are better. The cost is even higher after M takes sides with $f_1 = f_2 = -40$. By way of calibrating the model, a marginal cost of 40 means that if M paid this while 1 and 2 fought the two-actor war of attrition, then M's expected cost would be 2.6 or 26 percent of the stakes $|v_1 - v_2|$. M decides what to do at $T = 25$. (Figures 1, 2, 3, and 5 are actually plots based on this example.)

In the unique equilibrium, the lowest-payoff types still active at T are $z_1^* \approx 1.9$ and $z_2^* \approx 12.2$. Types $w_1 \in (0, 1.9)$ and $w_2 \in (0, 12.2)$ fight a war of attrition with cost ratio $\rho_0 = 1/2$ during the first phase which ends at $z_1^* z_2^*/k_0 \approx 21.1$. There follows an interval of pure fighting lasting until $T = 25$ when M joins 1 with probability $\alpha_1 = 0.7$ at T.

THE BOOMERANG EFFECT

The equilibrium strategies exhibit a boomerang effect when the stakes for M are not too large and fighting is costly, i.e., when $|v_1 - v_2|$ is not too large and f_1 and f_2 are negative. The expectation that M is sure to support a given faction leads both factions to take actions that make it less likely that M will actually support that faction when the time comes. The formal effect of these forces is that M typically plays a mixed equilibrium strategy as is the case in the numerical example. The substantive import of the boomerang effect is that alignment patterns will often be unpredictable and coalitions dynamically unstable.[24] Although identified in the specific context of a war of attrition, the boomerang effect seems likely to be present in any conflict in which there is uncertainty about each side's willingness to fight, more resolute types fight longer than less resolute types, and a third party's support advantages one side.

To develop some intuition for the boomerang effect and the forces inclining M to mix, suppose that both 1 and 2 expect M to join one of the factions, say 1, for sure. M's support for 1 at T will increase that faction's subsequent chances of prevailing. These better prospects will in turn induce some w_1 that would have otherwise dropped out prior to T to fight on until at least T. By contrast, some w_2 that would have fought until T drop out prior to T. The net effect is that there is a large set of types of w_1 that have fought until T only because they expected M to support 1 at T. Realizing that they have been had, these types will immediately drop out if M supports 2 instead, thereby shortening the war of attrition fought in the third phase. The prospect of a shorter war in turn creates an incentive for M to support 2 when fighting is costly, and the larger incentive to support 2 tends to undermine the original expectation that M is sure to support 1. Formally, M's strategy of supporting 1 at T is not sequentially rational: M prefers to deviate by supporting 2 once play gets to T.

[23] I have been unable to prove a unique pair simultaneously satisfies the two constraints and hence that there is always a unique equilibrium. But this has been the case in every numerical example.

[24] Stability will be examined below when the model is extended to allow M to decide what to do at two different times.

Figure 3 illustrates the boomerang effect graphically. The pair (z_1, z_2) must lie on $\theta(\rho_1)$ if M is expected to support 1 at T (Result 1). If, however, M actually supports 2, then all w_1 between z_1' and the point on $\theta(\rho_2)$ horizontally across from Z' instantly drop out. This tends to shorten the war and creates an incentive for M to support 2 when fighting is costly.

To emphasize that it is the way 1 and 2 react to the *expectation* that M will support a given faction that undermines this expectation, assume that M has to decide what to do at the very start of the game ($T = 0$). Since M effectively moves before 1 and 2, neither faction can react in anticipation of M's action. Absent these anticipatory reactions, M is almost sure to strictly prefer supporting one side or the other. M will virtually never be indifferent and play a mixed strategy.[25]

Equilibrium forces play out differently when the stakes are high or M profits from the fighting (and we are back to the case in which $T > 0$). When M has a strong preference for one side, it supports that side and does not mix as will be seen below in the discussion of comparative statics. When fighting is profitable, M wants to prolong the conflict rather than shorten it. Now the fact that M's deviating from what it was expected to do tends to shorten the war makes M less rather than more inclined to deviate.

At the risk of pushing this result too hard, it suggests that coalition formation will be more predictable and coalitions will be more stable when fighting is profitable.

WHAT MAKES *M* MORE LIKELY TO SUPPORT A GIVEN FACTION?

The effects of changes in M's payoffs on the chances that M supports j turn out to be quite straightforward and intuitive. Anything that increases M's payoff to supporting a given faction makes M more likely to support that faction: The higher M's payoff if 1 prevails or the flow payoff it derives from supporting 1, the more likely M is to support 1. The higher M's payoff if 2 prevails or the higher M's flow payoff from supporting 2, the less likely M is to support 1. The effects of changes in the total cost that 1 and 2 pay if M supports a given faction, i.e., the effects of changes in k_1 and k_2, are also straightforward. The higher the total cost of fighting k_j, the shorter the war of attrition following M's decision to support j. Thus a higher cost k_j (and therefore a shorter third-phase war) makes supporting j more attractive when fighting is costly ($f_j < 0$) and less attractive when fighting is profitable ($f_j > 0$). As a result, an increase in k_j makes M is more likely to support j when fighting is costly ($f_j < 0$) and less likely when fighting is profitable ($f_j > 0$). The effects of changes in the payoff to being on the winning side are ambiguous.

These results are stated formally in Proposition 2 in the Online Appendix. The remainder of this section sketches the derivation, and readers less interested in the technical results may omit the rest of this section. The comparative statics follow from three observations. First, $\theta(\rho_1)$, $\theta(\rho_2)$, and the timing constraint $T = z_1 z_2/k_0 + \lambda(z_1, z_2, \rho_0, k_0)$ are independent of M's payoffs v_1, v_2, γ, f_0, f_1, and f_2 as well as the total costs k_1 and k_2.[26] It follows that changes in these parameters have no effect on these curves. Second, Result 1 implies that $\alpha_1 = 0$ at Z'' in Figure 3 where $\theta(\rho_2)$ and the timing constraint intersect. Result 1 also implies that $\alpha_1 = 1$ at the other end of the timing constraint Z'. More generally, the farther the equilibrium pair (z_1, z_2) is from Z'' along the timing constraint, the higher α_1. This means that we can determine the effects of, say, an increase in v_1 on α_1 by seeing how (z_1, z_2) moves along the timing constraint.

A qualification is in order before making the third observation. One might hope that the effects of a change in an underlying parameter would have the same directional effect in all of the equilibria when there are multiple equilibria. In a model of war, for example, one might hope that a change in the cost of fighting would make war less likely in all equilibria. Or, the higher the payoff if 1 wins in the present game, the more likely M is to support 1 in all equilibria.

Hope as one might, it is often the case that the direction of the effects of a parametric change vary across equilibria. Typically, a change in x makes y more likely in some equilibria and less likely in others. A less ambitious hope is that a parametric change has the same effect on, say, the probability that M supports 1 in the equilibrium in which M is *least* likely to support 1 as well as the equilibrium in which M is *most* likely to support 1. This is the approach take here. (See Ashworth and Bueno de Mesquita (2006) for a discussion.)

Turning to the third observation, let μ be any parameter, e.g., 1's payoff to winning v_1. Then the probability that M supports 1 in the equilibria in which M is most and least likely to support 1 is increasing in μ if Δ_{12} is increasing in μ (i.e., if $\partial\Delta_{12}(z_1, z_2, \mu)/\partial\mu > 0$). Conversely, if Δ_{12} is decreasing in μ, then M is less likely to support 1 (and hence more likely to support 2) in the equilibria in which M is most and least likely to support 1. To see why, consider Figure 4 which shows how $\Delta_{12}(z_1, z_2, \mu)$ varies as we move along the timing constraint from Z'' where the timing constraint and $\theta(\rho_2)$ intersect (and $\alpha_1 = 0$ by Result 1) to Z' where the timing constraint and $\theta(\rho_1)$ intersect and $\alpha_1 = 1$.[27] By assumption, there are no pure strategy equilibria in the illustration. That is, M strictly prefers supporting 1 at Z'' ($\Delta_{12}(Z'') > 0$) whereas equilibrium play would require it to support 2 (see Result 1). Analogously, M strictly prefers 2 at Z' ($\Delta_{12}(Z') < 0$) whereas equilibrium play would have it support 1.

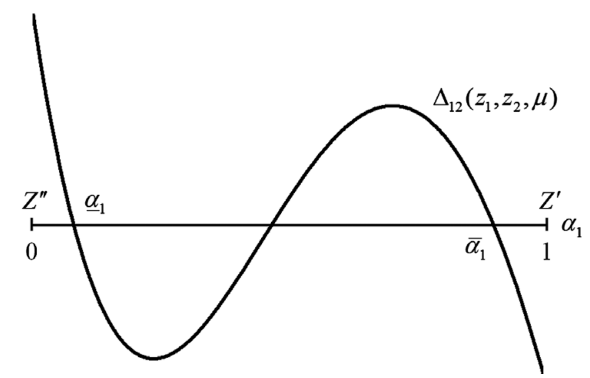
Rather than supporting either faction for sure, M mixes in the three equilibria depicted in Figure 4. More

[25] That is, $\Delta_{12}(z_1, z_2) \neq 0$ except possibly at a set of parameters with measure zero.

[26] The timing constraint is also independent in the more general case when $\underline{w}_1 \geq 0$ or $\underline{w}_2 \geq 0$ and the factions' stop times are given by Equation (A3).

[27] The figure is an illustration and not an actual plot. As noted above, all numerical examples have had a unique equilibrium.

FIGURE 4. Δ_{12} along the Timing Constraint



specifically, the incentive and timing constraints intersect wherever $\Delta_{12} = 0$ along the timing constraint. According to Proposition 1, a PBE corresponds to each of these intersections. Let $\underline{\alpha}_1$ and $\overline{\alpha}_1$ respectively denote the equilibria in which M is least and most likely to support 1. Clearly, an increase in μ to $\mu' > \mu$ that induces an upward shift in Δ_{12} leads to an increase in both $\underline{\alpha}_1$ and $\overline{\alpha}_1$.

Given that α_1 is increasing in a parameter if Δ_{12} is, it remains to be determined how Δ_{12} varies with changes in the parameters. Recall that $\Pi_j(z_1, z_2, \rho)$ is the probability that j prevails in the two-actor war of attrition given that z_j is the lowest-payoff type still active, the cost ratio is ρ, and given that at least one faction is strategic. Take $D(z_1, z_2, \rho, k)$ to be the expected duration of this conflict. Then algebra and the Appendix show

$$\begin{aligned}\Delta_{12}(z_1, z_2) = {} & (v_1 - v_2)[\Pi_1(z_1, z_2, \rho_1) - \Pi_1(z_1, z_2, \rho_2)] \\ & + 2\gamma[\Pi_1(z_1, z_2, \rho_1) - \Pi_2(z_1, z_2, \rho_2)] \\ & + f_1 D(z_1, z_2, \rho_1, k_1) - f_2 D(z_1, z_2, \rho_2, k_2),\end{aligned}$$

where we use the fact that $\Pi_1(z_1, z_2, \rho_1) + \Pi_2(z_1, z_2, \rho_1) = \Pi_1(z_1, z_2, \rho_2) + \Pi_2(z_1, z_2, \rho_2) = 1$.

The effects of changes in v_j, f_j, and the total cost k_j for $j \in \{1, 2\}$ are straightforward and unambiguous. The probability that 1 prevails is decreasing in the cost ratio, so $\Pi_1(z_1, z_2, \rho_1) > \Pi_1(z_1, z_2, \rho_2)$. Δ_{12} is therefore increasing in v_1 and decreasing in v_2. Thus the higher v_1 and the lower v_2, the more likely M is to support 1. Since durations are always nonnegative, a higher flow payoff to supporting 1 or a lower flow payoff to supporting 2 makes M more to support 1. Because duration is also decreasing in the total cost, a higher k_1 makes M less likely to support 1 when fighting is profitable ($f_1 > 0$) and more likely when it is costly ($f_1 < 0$).

The effects of changes in the payoff to being on the winning side are ambiguous and depend on the sign of $\Pi_1(z_1, z_2, \rho_1) - \Pi_2(z_1, z_2, \rho_2)$. The first term is the probability that 1 wins with M's support or, equivalently, the probability that M will be on the winning side if M supports 1. Analogously, $\Pi_2(z_1, z_2, \rho_2)$ is the probability that M will be on the winning side if it supports 2. It follows that an increase in γ makes M more likely to join the side that is more likely to win when it has M's support. In symbols, an increase in γ makes M more likely to join 1 when $\Pi_1(z_1, z_2, \rho_1) > \Pi_2(z_1, z_2, \rho_2)$ and more likely to join 2 when $\Pi_1(z_1, z_2, \rho_1) < \Pi_2(z_1, z_2, \rho_2)$.

EMPIRICAL CHALLENGES

The theoretical model developed here and the resulting comparative statics have implications for empirical efforts to assess the effects of intervention. Many studies attempt to evaluate claims about the effects of intervention by estimating a duration model. Gent (2008), for example, uses a competing-risk model to test the hypothesis that third party support for the rebels increases their chances of prevailing. Balch-Lindsay and Enterline (2000) and Balach-Lindsay, Enterline, and Joyce (2008) also use this approach to test their hypotheses about intervention. Regan (2002) estimates a Weibull duration model.

The present analysis shows that using duration models to evaluate the effects of intervention faces at least three formidable challenges. The most basic is that the likelihood that a faction prevails depends on the cost ratio but not on the total cost whereas the expected duration and hazard rate depend on both. Thus trying to infer changes in the likelihood of prevailing (i.e., in the cost ratio) from changes in the duration or hazard rate is problematic. A second challenge is that the hazard rate of equilibrium behavior may be very complicated even in a relatively simple model like this one. This casts doubt on the proportionality assumption of the Cox approach. Finally, there may be significant selection effects.

To illustrate these challenges, suppose we were trying to use a competing-risk proportional hazards model to test the hypothesis that third party support for the "rebel opposition" (i.e., for the weaker faction 2) makes that faction more likely to prevail. For each observation, the (hypothetical) data indicates whether the third party supported 1, supported 2, or stayed out; whether 1 or 2 ultimately prevailed; and the duration of the conflict (see below for the formal the analysis of the game when M has three options). Let $X_j = 1$ if M supports faction j and $X_j = 0$ otherwise for $j \in \{1, 2\}$. M stays out in cases where $X_1 = X_2 = 0$. Take $H_{2|k}(t)$ to be 2's hazard rate of victory, i.e., the probability that 2 prevails in the next instant given that the conflict has lasted until t and that M supported $k \in \{0, 1, 2\}$ where supporting "0" means staying out.

The Cox model assumes that the hazard rates of the different possible outcomes are proportional. In symbols, $H_{2|k}(t) = H_{2|0}(t)e^{\beta_1 X_1 + \beta_2 X_2}$ where $H_{2|0}$ is the baseline hazard rate, i.e., the chances that 2 prevails if M stayed out. The parameter β_j measures the effect on $H_{2|0}$ if M supports j. If, for example, M supports 2 in a given observation, then $X_1 = 0$, $X_2 = 1$, and the previous equation reduces to $H_{2|2}(t) = H_{2|0}(t)e^{\beta_2}$. This leaves $H_{2|2}(t)/H_{2|0}(t) = e^{\beta_2}$. The econometric challenge is to estimate β_1 and β_2. A positive estimate of

β_j is generally taken to mean that M's support for j makes j more likely to win.

The model makes it possible to derive the hazard rates and this reveals other challenges. To derive the expression for $H_{2|2}(t)/H_{2|0}(t)$, note that the probability that the fighting lasts until at least time $t > T$ is the probability that both factions stop at t or later. The type w_1 that stops at t given that M supported 2 is the inverse $w_1 = \sigma_{1|2}^{-1}(t)$, so the probability that 1 has not yet quit at t is $1 - G(\sigma_{1|2}^{-1}(t)) = e^{-\sigma_{1|2}^{-1}(t)}$. Similarly, the probability that 2 has not yet quit at t is $1 - G(\sigma_{2|2}^{-1}(t)) = e^{-\sigma_{2|2}^{-1}(t)}$, and the probability neither faction has stopped by t is $e^{-\sigma_{1|2}^{-1}(t)-\sigma_{2|2}^{-1}(t)}$. The chances that 2 prevails in the next instant is the probability that 1 quits in the next instant or $d(1 - e^{-\sigma_{1|2}^{-1}(t)})/dt$. This yields a hazard rate of a rebel victory if M supports 2 of

$$H_{2|2}(t) = \frac{\Pr\{1 \text{ quits in the next instant}\}}{\Pr\{\text{ neither faction quits before } t\}}$$
$$= \frac{d(1 - e^{-\sigma_{1|2}^{-1}(t)})/dt}{e^{-\sigma_{1|2}^{-1}(t)-\sigma_{2|2}^{-1}(t)}} = e^{\sigma_{2|2}^{-1}(t)}\frac{d\sigma_{1|2}^{-1}(t)}{dt}.$$

Using the analogous expression for $H_{2|0}(t)$,

$$\frac{H_{2|2}(t)}{H_{2|0}(t)} = \frac{k_2}{1+\rho_2}\left[\frac{k_2\widehat{t}}{\overline{w}_1\overline{w}_2} + \left(\frac{z_{1|2}}{\overline{w}_1}\right)^{1+\rho_2}\right]^{\frac{-\rho_2}{1+\rho_2}}$$
$$\times e^{\overline{w}_2\left[\frac{k_2\widehat{t}}{\overline{w}_1\overline{w}_2} + \left(\frac{z_{1|2}}{\overline{w}_1}\right)^{1+\rho_2}\right]^{\frac{\rho_2}{1+\rho_2}}}$$
$$\times \left(\frac{1+\rho_0}{k_0}\right)\left[\frac{k_0\widehat{t}}{\overline{w}_1\overline{w}_2} + \left(\frac{z_{1|0}}{\overline{w}_1}\right)^{1+\rho_0}\right]^{\frac{\rho_0}{1+\rho_0}}$$
$$\times e^{-\overline{w}_2\left[\frac{k_0\widehat{t}}{\overline{w}_1\overline{w}_2} + \left(\frac{z_{1|0}}{\overline{w}_1}\right)^{1+\rho_0}\right]^{\frac{\rho_0}{1+\rho_0}}} \quad \textbf{(5)}$$

where $\widehat{t} = t - T$ is the time since intervention and $z_{1|k}$ is the lowest payoff type of 1 still active at T given M supported k in equilibrium.

Three observations follow from this expression. First, the proportionality assumption is clearly violated. Indeed, the hazard-rate ratio varies in a complicated way over time. Second, there are likely to be selection effects. Consider an equilibrium in which M supports 2 for sure. Then the pair of lowest-payoff types active at T, $(z_{1|2}, z_{2|2})$, lies at the intersection of the timing constraint and $\theta(\rho_2)$. Both this pair and the corresponding hazard rate $H_{2|2}$ are well defined. But the pair of lowest-payoff types still active at T had M supported 1 or stayed out, that is, $(z_{1|1}, z_{2|1})$ and $(z_{1|0}, z_{2|0})$, are not defined. This in turn implies that the corresponding hazard rates $H_{2|1}$ and $H_{2|0}$ are not well defined. The numerical example illustrates a similar point. As shown below, M mixes over supporting 1 and 2 and never stays out when the numerical example is extended to allow M to choose not to take sides. Hence $(z_{1|1}, z_{2|1}) = (z_{1|2}, z_{2|2})$, and $H_{2|1}$ and $H_{2|2}$ are well defined and can be calculated. But what would have happened had M stayed out, that is, $(z_{1|0}, z_{2|0})$ and $H_{2|0}$, cannot.

Finally, suppose that despite all of these issues one could establish empirically that $H_{2|2}(t)/H_{2|0}(t) > 1$ at all $t > T$. That is, 2 at any $t > T$ is always more likely to prevail in the next instant if M supported it than if M had stayed out. Could we infer from this that M's support makes 2 more likely to win? Is $\rho_2/\rho_0 > 1$ identified?

The answer is no. To establish this, we construct a counterexample in which M's support, by assumption, has *no* effect on either faction's chances of prevailing. Nevertheless, M's support increases the hazard rate of 2's prevailing, i.e., $H_{2|2}(t)/H_{2|0}(t) > 1$. Formally, suppose M's support has no effect on the cost ratio with $\overline{w}_1/\overline{w}_2 < \rho_0 = \rho_1 = \rho_2 < 1$ where the inequalities ensure that 1 is always stronger than 2 regardless of what M does (see Assumption 1). Suppose further that M's support for the weaker side 2 (the rebels) increases the total marginal cost of fighting: $k_2 > k_0$. The Online Appendix shows that $H_{2|2}(t)/H_{2|0}(t) > 1$ at all $t > T$.

The counterexample illustrates the basic identification problem. The increased cost of fighting following M's decision to support 2 makes it more likely that the fighting will end in the next instant. But both sides are more likely to stop, and there is no change in the factions' relative chances of prevailing since the cost ratio remains unchanged.

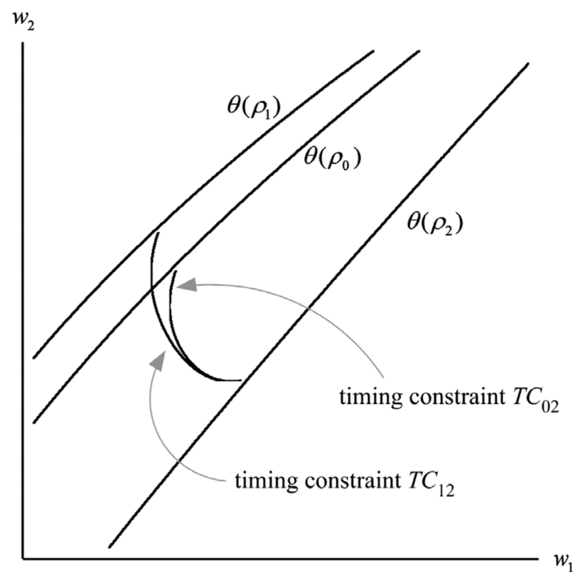
NOT TAKING SIDES

Until now M has had to take sides by supporting either 1 or 2. M could not decide to stay out. This section extends the model by giving M the three options of supporting 1 or 2 or staying out. We show that the equilibrium and comparative statics of the numerical example do not change when the option of staying out is added to the model. The basic reason is that M strictly prefers supporting 1 or 2 to staying out. Hence adding the third option of staying out with its lower payoff has no effect on equilibrium play. More generally, M will typically mix over two options when it has three alternatives as it does in the numerical example and the presence of the third option has no effect. But there may be cases in which M mixes over all three alternatives depending on the parameter values.

The analysis of the three-option game begins with an intermediate step in which M chooses between staying out or supporting 2. The analysis is fundamentally the same if M had the options of staying out or joining 1. Indeed, the main point of the intermediate step is that the equilibrium analysis is essentially the same when M only has two options regardless of what those options are. More importantly, the analysis shows that the equilibrium and comparative statics of the model when M only has two options is often an equilibrium of the model when M has three options.

To specify the model when M can stay out or support 2, assume that the cost ratio, total cost, and flow payoff

FIGURE 5. Taking Sides or Staying Out



remain ρ_0, k_0 and f_0 if M stays out (where the subscript "0" will be used for the option of staying out). They change to ρ_2, k_2, and f_2 if M supports 2. Let α_0 be the probability that M stays out. Take (z_1, z_2) to be the lowest-payoff types active at T and $S_0(z_1, z_2)$ to be M's payoff to staying out at (z_1, z_2).

Figure 5 illustrates the geometric similarly of the equilibrium analysis when M can stay out or support 2, or when when M can support 1 or 2. The lowest-payoff types (z_1, z_2) when M can support 0 or 2 must lie on or between $\theta(\rho_0)$ and $\theta(\rho_2)$ and satisfy an incentive constraint for M as well as a timing constraint. To specify the latter, note that the first phase of equilibrium play ends when z_1 and z_2 quit at time $z_1 z_2/k_0$ (assuming $\underline{w}_1 = \underline{w}_2 = 0$). Let $\lambda_{02}(z_1, z_2)$ be the length of the interval of pure fighting that follows given that the cost ratio will be ρ_0 or ρ_2 after M decides what to do.[28] This yields the timing constraint $T = z_1 z_2/k_0 + \lambda_{02}(z_1, z_2, \rho_0, \rho_2)$. Let TC_{02} denote the set of points satisfying this constrain. The points in TC_{02} are shown in Figure 5 as are the points satisfying the timing constraint TC_{12} when M could only support 1 or 2.

M's incentive constraint follows from an analogue of Result 1: If (z_1, z_2) is on the upper edge $\theta(\rho_0)$, M must stay out ($\alpha_0 = 1$) which in turn requires that M weakly prefer supporting 0, i.e., $\Delta_{02}(z_1, z_2) \equiv S_0(z_1, z_2) - S_2(z_1, z_2) \geq 0$. (See the Appendix for the expression for Δ_{02}.) If (z_1, z_2) is on $\theta(\rho_2)$, M joins 2 for sure and must weakly prefer supporting 2 ($\Delta_{02}(z_1, z_2) \leq 0$). Finally if (z_1, z_2) is strictly between $\theta(\rho_0)$ and $\theta(\rho_2)$, then M must be indifferent between supporting 0 or 2 so that M can mix in equilibrium ($\Delta_{02}(z_1, z_2) = 0$). Let MIC_{02} be the set of points satisfying this incentive constraint.

It is straightforward to show that the analogue of Proposition 1 holds. At least one (z_1, z_2) satisfies TC_{02} and MIC_{02}, and a virtually unique PBE is associated with any (z_1, z_2) satisfying both. Types $w_1 \in (\underline{w}_1, z_1]$ and $w_2 \in (\underline{w}_2, z_2]$ fight a war of attrition until time $T - \lambda_{02}(z_1, z_2, \rho_0, k_0)$ followed by a pure-fighting interval of length λ_{02}. M then decides what to do, and the war of attrition goes on with the cost ratio and total cost determined by M's action.

The comparative-static analysis parallels the argument in the case when M could support support 1 or 2. If $\Delta_{02}(z_1, z_2)$ is increasing in a parameter μ, then so are the chances that M will stay out in the equilibria in which M is most and least likely to stay out. It follows that the probability that M stays out are increasing in v_1 and f_0 and decreasing in v_2 and f_2. A higher total k_2 decreases the expected duration of fighting if M supports 2.[29] This in turn makes M more likely to support 2 when fighting is costly ($f_2 < 0$) and less likely when it is profitable. The effects of changes in the payoff to being on the winning are however unambiguous in this case. M is less likely to be on the winning side if it supports 2, so an increase in γ makes M more likely to stay out.

Finally, what happens when M has the three options of staying out or supporting 1 or 2? Let TC_{12}, TC_{10}, and TC_{02} be the timing constraints if M has to choose between supporting 1 or 2, 1 or "0" , or "0" or 2, respectively. Take MIC_{ij} to be M's incentive constraints when M has to choose between joining i or j.

In the baseline numerical example where M must support 1 or 2, M mixes at the unique (z_1^*, z_2^*) satisfying M's incentive and timing constraints with $(z_1^*, z_2^*) \approx (1.9, 12.2)$. Moreover, M's payoff to supporting 1 or 2 at (z_1^*, z_2^*) is strictly larger than its payoff to staying out. It follows that the numerical example described above when M can either support 1 or 2 is also an equilibrium when M has three options. The comparative static results for the equilibrium when M can only support 1 or 2 also hold for the equilibrium in which M can support 1 or 2 or stay out.

More generally suppose (z_1, z_2) satisfies the timing and incentive constraints when M's options are restricted to supporting i or j. The equilibrium of the two-option game will be an equilibrium of the three-option game as long as M weakly prefers supporting i or j to supporting the third option, say, n. In symbols, (z_1, z_2) is associated with an equilibrium in the three-actor game as long as $\min\{S_i(z_1, z_2), S_j(z_1, z_2)\} \geq S_n(z_1, z_2)$. The comparative statics of the equilibrium when M has three options are also the same as when M has two alternatives as long as the previous inequality is strict.

Of course, it may be the case that M always prefers to deviate to the third option n whenever (z_1, z_2) satisfies the incentive and timing constraints with respect to the other two options i and j. When this occurs, M must mix over all three options.

[28] To obtain an expression for λ_{02}, substitute ρ_0 for ρ_1 in the expression for λ_{12} given in Equation (A4).

[29] The total cost k_0 affects both Δ_{02} and TC_{02}, making it impossible to determine the effects of changes in k_0 solely through its affects on Δ_{02}.

https://doi.org/10.1017/S0003055416000782 Published online by Cambridge University Press

TAKING AND SWITCHING SIDES

What happens when M can take sides at time T' and revisit the decision at a later time T''? The boomerang effect that tends to induce mixing and make alignment decisions unpredictable when M has one decision time tends to make M's decisions unpredictable at both times. As a result, M often mixes at both T' and T'' when fighting is costly and the stakes are not too high. An outside observer looking at a set of cases would see instances in which M joins 1 at T' and subsequently sticks with 1 at T'', M joins 1 at T' and then switches to 2 at T'', M joins 2 at T' and sticks with 2 at T'', or M joins 2 at T' and switches to 1 at T''. In brief, almost anything can happen. This at least qualitatively is in keeping with recent empirical work on alliances in civil wars which finds that side switching is quite common (Christia 2012; Seymour 2014).

To specify the extended model, add an additional decision time of $T' = 15$ to the baseline numerical example where M could only support 1 or 2 at $T'' = 25$. In effect, M can now decide whether to support 1 or 2 at T' and then stick with that faction at T'' or switch sides. The Appendix sketches the derivation of the equilibrium. The key to finding the equilibrium is using the fact that the equilibrium of the one-decision-time model is the equilibrium of the continuation game when M has two decisions.

The equilibrium probability that M joins 1 at T' in the two-decision-time model is 0.52. Conditional on having joined 1 at T', the probability that M continues to support 1 at T'' is 0.74. Conditional on supporting 2 at T', the probability that M switches to supporting 1 is 0.84. The overall probability that M supports a side and then sticks with that side is 0.49. The probability that M switches sides is 0.51.

CONCLUSION

Third party intervention plays an important role in many conflicts. This study analyzes intervention by adding a third party to a standard two-actor war of attrition. The third party cares about the outcome and can influence the outcome by taking sides. The analysis yields four main results. First, it provides a conceptual framework for examining the interdependent effects of endogenous intervention on duration and outcome. Second, the analysis identifies a boomerang effect that tends to make alignment decisions unpredictable and coalitions dynamically unstable when the third-party's stakes are not too large and fighting is costly. These theoretical results resonate with recent empirical work on coalition formation in civil wars (Christia 2012; Seymour 2014). Third, the model provides many clear comparative-static results. Finally, the explicit derivation of hazard rates reveals previously unappreciated challenges for empirical efforts to assess the effects of intervention. Many of these efforts assume that intervention in support of a faction makes that faction more likely prevail and that this has a direct effect on the duration and hazard rates of the conflict. This is not the case in the model studied here.

The analysis also points toward two areas for future work. The first is to allow the third party to choose when it intervenes. The same forces that tend to undermine pure-strategy equilibria when the third party intervenes at an exogenously specified time seem likely to be at work when the third party can endogenously choose whether and when to intervene. But this conjecture remains to be established.[30]

The second area for future work is to consider the case of limited resources. In the model studied here, neither side can defeat the other militarily and each side can fight as long as it wants. The present formulation thus abstracts away from the problem of finding the means to fight. As such, the model may be a somewhat better fit for intervention into civil rather than interstate war. Future work might assume that a faction can fight only as long as it has resources and that fighting burns through these resources.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0003055416000782

APPENDIX

Derivation of the equilibrium stop times in the baseline model: Let $\tau_j(t) \equiv \sigma_j^{-1}(w_j)$ be the type that stops at t. (Fudenberg and Tirole (1986) ensure that this inverse is well defined for $t > 0$.) Then 1's payoff to stopping at t is

$$V_1(t) = -c_1 t[1 - G(\tau_2(t))] + \int_0^t (w_1 - cs) G'(\tau_2(s))\tau_2'(s)ds.$$

Setting the derivative equal to zero yields the first-order condition $0 = -c_1 + w_1\tau_2'(t)$. Using $w_1 = \tau_1(t)$ gives $c_1 = \tau_1(t)\tau_2'(t)$. Faction 2's analogous first-order condition is $c_2 = \tau_2(t)\tau_1'$. Differentiating the former gives $\tau_1' = -c_1\tau_2''/(\tau_2')^2$ and substituting this into the latter yields $-\tau_2\tau_2''/(\tau_2')^2 = c_2/c_1$. Integrating by parts gives a simple first-order differential equation that can readily be solved to yield $\tau_2(t) = m_2\left[(c_1 + c_2)t + n_2\right]^{c_1/(c_1+c_2)}$ where m_2 and n_2 are constants of integration. Similar reasoning leads to

$$\tau_1(t) = m_1\left[(c_1 + c_2)t + n_1\right]^{c_2/(c_1+c_2)}, \quad \textbf{(A1)}$$

where m_1 and n_1 are constants of integration.

Differentiating the expression for τ_2 and substituting this into the first order condition for 1 leads to

$$1 = (m_2 m_1)^{\frac{c_1+c_2}{c_1}} \left[\frac{(c_1 + c_2)t + n_2}{(c_1 + c_2)t + n_1}\right].$$

An analogous argument for 2 gives

$$1 = (m_2 m_1)^{\frac{c_1+c_2}{c_2}} \left[\frac{(c_1 + c_2)t + n_1}{(c_1 + c_2)t + n_2}\right].$$

[30] An intermediate step which may also have substantive appeal would be for the exogenous decision time to arrive stochastically rather than be set to a given time T.

Multiplying these expressions gives $1=(m_2m_1)^{c_1+c_2}$ which in turn implies $m_2=1/m_1$ and hence that $n_1=n_2$. Using these results to rewrite the expression for τ_2 gives

$$\tau_2(t)=(1/m_1)\left[(c_1+c_2)t+n_1\right]^{c_1/(c_1+c_2)}. \qquad \textbf{(A2)}$$

Two boundary conditions pin down m_1 and n_1. The first is that the highest-payoff strategic types $\overline{w}_1$ and $\overline{w}_2$ must stop at the same time. Substituting these types into A1 and A2, solving each equation for the expression in brackets, and equating the results leads to an expression that can be solved for m_1 to obtain $m_1=\overline{w}_1^{c_1/(c_1+c_2)}\overline{w}_2^{c_2/(c_1+c_2)}$.

A second boundary condition is at most an atom of one faction can quit at $t=0$. Otherwise a positive measure of types could profitably deviate by waiting an instant longer to quit. Given that $\tau_j(t)$ is strictly increasing for $t>0$ (Fudenberg and Tirole 1986), the fact that at most one faction can quit at $t=0$ implies $\min\{\lim_{t\to0^+}\tau_1(t)-\underline{w}_1,\lim_{t\to0^+}\tau_2(t)-\underline{w}_2\}=0$. This leads to

$$\tau_1(t)=\overline{w}_1\left[\frac{k_0t}{\overline{w}_1\overline{w}_2}+N\right]^{\frac{1}{1+\rho_0}} \text{ and } \tau_2(t)=\overline{w}_2\left[\frac{k_0t}{\overline{w}_1\overline{w}_2}+N\right]^{\frac{\rho_0}{1+\rho_0}},$$

where $N=(\underline{w}_1/\overline{w}_1)^{1+\rho_0}$ when $w_2\le\overline{w}_2\,(w_1/\overline{w}_1)^{\rho_0}$ and $N=(\underline{w}_1/\overline{w}_1)^{1+\rho_0}$ when $w_2\ge\overline{w}_2\,(w_1/\overline{w}_1)^{\rho_0}$. Solving for the equilibrium stop times now yields

$$\begin{aligned}&\sigma_1^*(w_1,\underline{w}_1,\underline{w}_2,\rho_0,k_0,\overline{w}_1,\overline{w}_2)\\&\quad=\max\left\{0,(\overline{w}_1\overline{w}_2/k_0)\left[(w_1/\overline{w}_1)^{1+\rho_0}-N\right]\right\}\\&\sigma_2^*(w_2,\underline{w}_1,\underline{w}_2,\rho_0,k_0,\overline{w}_1,\overline{w}_2)\\&\quad=\max\left\{0,(\overline{w}_1\overline{w}_2/k_0)\left[(w_2/\overline{w}_2)^{1+1/\rho_0}-N\right]\right\}.^{31} \qquad \textbf{(A3)}\end{aligned}$$

A finesse of the technical issue of unbounded payoffs for M. M payoffs are always unbounded when there is a positive probability that the fighting will last forever. To deal with this issue, we assume that M believes that it is facing at least one strategic type when play begins. More specifically, assume the game starts at $t=-1$ with 1 and 2 learning their types. If neither type is strategic (i.e., if $w_1>\overline{w}_1$and $w_2>\overline{w}_2$), the game ends with factions 1 and 2 along with M getting a payoff of zero. If at least one type is strategic, play continues to $t=0$ where the game described above begins. At the start of this game, a strategic w_j believes that that the other faction's payoff w_i is distributed according to $G_i(w_i)$ which is w_j's prior belief conditional on (i) at least one type being strategic and (ii) w_j is strategic. By contrast, M believes that the type pair (w_1,w_2) is distributed according to the density $G_1'(w_1)G_2'(w_2)/(1-e^{-\overline{w}_1-\overline{w}_2})=e^{\underline{w}_1+\underline{w}_2-w_1-w_2}/(1-e^{\underline{w}_1+\underline{w}_2-\overline{w}_1-\overline{w}_2})$ over the region in which at least one type is strategic. This density is M's prior belief conditional on facing at least one strategic type.

It will also be useful to let $R(x,y)\equiv e^{\underline{w}_1+\underline{w}_2-x-y}-e^{\underline{w}_1+\underline{w}_2-\overline{w}_1-\overline{w}_2}$ be the probability that $w_1\ge x$, $w_2\ge y$, and that at least one type is strategic.

Characterization of the Perfect Bayesian Equilibria. The characterization of the PBEs is done in three steps. The first is to formally describe the actors' strategies. The next is to define the lowest payoff type z_j still active at T precisely and show that it is a cutoff. The third step is to identify M's incentive constraint. The derivation of the timing constraint follows, and this leads directly to the characterization of the equilibria in Proposition 1.

A strategy for M is simply the probability α_j that M aligns with j at T where $\alpha_1+\alpha_2=1$. A pure strategy for faction j is a pair of measurable functions $\{\sigma_{j|1}(w_j),\sigma_{j|2}(w_j)\}$ where $\sigma_{j|n}(w_j)$ specifies w_j's stop time if M supports n at T. Stop times must also be nonnegative and satisfy $\sigma_{j|1}(w_j)=\sigma_{j|2}(w_j)$ whenever $\sigma_{j|n}(w_j)<T$. The later restriction means that w_j cannot condition when it quits on M's actions if w_j stops before T. Note however that w_j can condition its action at time T on what M does at T. If, for example, $\sigma_{1|1}(w')=t_1>T$, and $\sigma_{1|2}(w')=T$, then w' stops at t_1 if M supports 1 and w' quits at T if M aligns with 2. The substantive interpretation of this is that the factions get to see what M does at T before deciding what to do at T even if M is mixing.

Given a pair of strategies for 1 and 2, we can formally define the lowest-payoff type still active at T. Type w_j might stop at different times depending on what M does at T. Nevertheless, w_j is sure to be active until $\min\{\sigma_{j|1}(w_j),\sigma_{j|2}(w_j)\}$. This means that w_j is sure to be active until T if $\min\{\sigma_{j|1}(w_j),\sigma_{j|2}(w_j)\}\ge T$. Define the lowest-payoff type still active at T to be $z_j\equiv\inf\{w_j:\min\{\sigma_{j|1}(w_j),\sigma_{j|2}(w_j)\}\ge T\}$.

Type z_j is a cut point between the types of j stopping before T and those stopping at T or later. The definition of z_j implies that all $w_j<z_j$ stop before T. Lemma 3 shows all $w_j>z_j$ stop at T or later. It follows that faction j believes at (z_1,z_2) that i's payoffs are distributed according to the truncated distribution $G_i(w_i)/[1-G_i(z_i)]=1-e^{z_i-w_i}$. This along with the specification of the payoffs in Equation (3) mean that the continuation game following M's decision is a simple two-actor war of attrition with the cost ratio and total cost determined by M's action.

Lemma 3 *If z_j is the lowest-payoff type still active at T and $w_j>z_j$, then w_j is still active at T.*

Proof: Arguing by contradiction, assume $w>z_j$ and that w quits at $\tau<T$. We show that w can profitably deviate by mimicking z_j's strategy. Let $U_j(x,\tau)$ be x's payoff to following w's strategy of fighting until τ and quitting. We can write $U_j(x,\tau)=x\pi_\tau-\eta_\tau$ where π_τ is the probability that j wins with this strategy and η_τ is the expected cost of fighting. Similarly, $U_j(x,T)=x\pi_T-\eta_T$ is x's payoff to following z_j's strategy.

Since z_j cannot profitably deviate from its strategy, $U_j(z_j,T)\ge U_j(z_j,\tau)$ or $z_j(\pi_T-\pi_\tau)\ge\eta_T-\eta_\tau$. Nonstrategic types never quit, so there is a positive probability of

[31] A technical problem arises if the type space includes $\underline{w}_j$, i.e., G_j is distributed over $[\underline{w}_j,\infty)$ and $\underline{w}_j>0$. Suppose 2 quits with probability ψ at $t=0$. Then, $\underline{w}_1$ gets $\underline{w}_1(\psi/2)$. If $\underline{w}_1$ fights for an instant before quitting, it gets $\underline{w}_1\psi$ at an arbitrarily small cost. Hence, a type $\underline{w}_1$would have an incentive to deviate and no equilibrium would exist. This issue can be resolved by excluding the $\underline{w}_j$'s from the type space or focusing on generic PBEs, i.e., PBEs in which the set of types with a profitable deviation has measure zero. The former approach is taken here.

fighting between τ and T. Hence, $\eta_T - \eta_\tau > 0$ which yields $\pi_T - \pi_\tau > 0$. It follows that $w(\pi_T - \pi_\tau) > z_j(\pi_T - \pi_\tau)$ and $U_j(w, T) > U_j(w, \tau)$. □

We can now specify M's payoff in the continuation game following its decision at T. Recall that $\Pi_j(z_1, z_2, \rho)$ is the probability that j prevails in the two-actor war of attrition given that z_j is the lowest-payoff type of j still active, the cost ratio is ρ, and at least one type is strategic. $D(z_1, z_2, \rho, k)$ is the expected duration of the conflict. It follows that M's payoff to supporting 1 at a (z_1, z_2) which is on or between $\theta(\rho_1)$ and $\theta(\rho_2)$ is

$$\begin{aligned} S_1(z_1, z_2) &= (v_1 + \gamma)\Pi_1(z_1, z_2, \rho_1) + (v_2 - \gamma)\Pi_2(z_1, z_2, \rho_1) \\ &\quad + f_1 D(z_1, z_2, \rho_1, k_1) \\ &= (v_1 + \gamma)\left[\frac{e^{\underline{w}_1 - z_1}(e^{\underline{w}_2 - z_2} - e^{\underline{w}_2 - \overline{w}_2(z_1/\overline{w}_1)^{\rho_1}})}{R(z_1, z_2)}\right. \\ &\quad \left. + \int_{\overline{w}_2(z_1/\overline{w}_1)^{\rho_1}}^{\overline{w}_2} \frac{e^{\underline{w}_1 + \underline{w}_2 - w_2 - \overline{w}_1(w_2/\overline{w}_2)^{1/\rho_1}}}{R(z_1, z_2)} dw_2\right] \\ &\quad + (v_2 - \gamma)\int_{z_1}^{\overline{w}_1} \frac{e^{\underline{w}_1 + \underline{w}_2 - w_1 - \overline{w}_2(w_1/\overline{w}_1)^{\rho_1}}}{R(z_1, z_2)} dw_1 \\ &\quad + f_1 \int_{z_1}^{\overline{w}_1} \left[\sigma_1(w_1, z_1, z_2, \rho_1, k_1, \overline{w}_1, \overline{w}_2)\right. \\ &\quad \times \left(1 + \left(\frac{\rho_1 \overline{w}_2}{\overline{w}_1}\right)\left(\frac{w_1}{\overline{w}_1}\right)^{\rho_1 - 1}\right) \\ &\quad \left. \times \frac{e^{\underline{w}_1 + \underline{w}_2 - w_1 - \overline{w}_2(w_1/\overline{w}_1)^{\rho_1}}}{R(z_1, z_2)}\right] dw_1. \end{aligned}$$

M's payoff to joining 2, $S_2(z_1, z_2)$, is defined analogously.

M's incentive constraint follows from $\Delta_{12}(z_1, z_2)$. For notational convenience, let F_{12} (for "feasible") be the set of points weakly between $\theta(\rho_1)$ and $\theta(\rho_2)$. Define int F to be the interior of F_{12}. Then Lemma 4 or Result 1 as stated in the text holds.

Proof of Lemma 4 (Result 1): Part (iii) is proved in the text. Parts (i) and (ii) follow from similar arguments. We show here that (z_1, z_2) must be in int F_{12}. Arguing by contradiction, assume $(z_1, z_2) \notin F_{12}$ with, say, $z_2 < \overline{w}_2(z_1/\overline{w}_1)^{\rho_2}$. Then an atom of types of 2 quit at T regardless of whether M supports 1 or 2. The argument in the proof of 4(iii) then implies that there must be a $\lambda > 0$ such that no w_1 quit during the interval $(T - \lambda, T)$ regardless of what M does.

It follows that any w_2 in a neighborhood $[z_2, z_2 + \xi)$ can profitably deviate by quitting slightly earlier than T. Take $\varepsilon < \lambda$ and ξ sufficiently small that all $w_2 \in (z_2, z_2 + \xi)$ quit at T regardless of what M does. Conditional on being at $T - \varepsilon$, $w_2 \in [z_2, z_2 + \xi)$ gets zero if it quits at that time. If it fights on until T and then quits, its payoff is $-c_2\varepsilon$. Hence any $w_2 \in [z_2, z_2 + \xi)$ does strictly better by quitting at $T - \varepsilon$ rather than at T. The definition of z_2 also implies that $[z_2, z_2 + \xi)$ is nonempty, so some types can profitably deviate. □

Let MIC_{12} denote the points $(x_1 x_2)$ satisfying M's incentive constraint, i.e., the set of points such that $(x_1 x_2)$ is weakly between $\theta(\rho_1)$ and $\theta(\rho_2)$ and (i) $\Delta_{12}(x_1 x_2) \geq 0$ if $(x_1 x_2)$ is on $\theta(\rho_1)$, (ii) $\Delta_{12}(x_1 x_2) \leq 0$ if $(x_1 x_2)$ is on $\theta(\rho_2)$, or (iii) $\Delta_{12}(x_1 x_2) = 0$ if $(x_1 x_2)$ is strictly between $\theta(\rho_1)$ and $\theta(\rho_2)$.

The lowest-payoff types still active at T must also satisfy a timing constraint. Suppose (z_1, z_2) is strictly between $\theta(\rho_1)$ and $\theta(\rho_2)$. The proof of Lemma 4(iii) shows that there must be an interval of pure fighting just before T during which no types of either faction quit. Let λ be the length of this interval. Then z_1 and z_2 must be indifferent between quitting at $T - \lambda$ and $T + \varepsilon$ for an arbitrarily small $\varepsilon > 0$. (As will be seen in the formal construction of the equilibria, there is a discontinuous jump in the payoff to quitting at T and slightly later. This jump is due to the tie-breaking rule of winning with probability 1/2 in the event of a tie.) To see why this indifference must hold, suppose z_j strictly prefers to wait until time T. Then so would some w_j slightly less z_j, and consequently z_j would not be the lowest-payoff type active at T. If z_j strictly preferred stopping at $T - \lambda$, some w_j slightly above z_j would also strictly prefer to stop at $T - \lambda$. Again, z_j would not be the lowest-payoff type active at T.

Type z_1's indifference implies $0 = z_1\alpha_1[1 - e^{z_2 - \overline{w}_2(z_1/\overline{w}_1)^{\rho_1}}] - c_1\lambda$. In words, when z_1 is deciding what to do at $T - \lambda$, its net benefit of quitting is zero. If it fights on until T, its expected gain is the payoff to winning times the probability that M supports 1 weighted by the probability that 2 drops out at T. (If M supports 2, z_1 gets zero.) Type z_1's cost of fighting to T is $c_1\lambda$. Indifference requires that the net benefit of quitting at $T - \lambda$ equal the net benefit of quitting at T. Similarly, z_2's indifference gives $0 = z_2(1 - \alpha_1)[1 - e^{z_1 - \overline{w}_1(z_2/\overline{w}_2)^{1/\rho_2}}] - c_2\lambda$. Solving for α_1 and λ yields

$$\begin{aligned} &\alpha_1(z_1, z_2, \rho_0) \\ &\quad = \frac{\rho_0 z_2(1 - e^{z_1 - \overline{w}_1(z_2/\overline{w}_2)^{1/\rho_2}})}{\rho_0 z_2(1 - e^{z_1 - \overline{w}_1(z_2/\overline{w}_2)^{1/\rho_2}}) + z_1(1 - e^{z_2 - \overline{w}_2(z_1/\overline{w}_1)^{\rho_1}})}, \\ &\lambda(z_1, z_2, \rho_0, k_0) \\ &\quad = \frac{z_1 z_2(1 + \rho_0)(1 - e^{z_2 - \overline{w}_2(z_1/\overline{w}_1)^{\rho_1}})(1 - e^{z_1 - \overline{w}_1(z_2/\overline{w}_2)^{1/\rho_2}})}{k_0[\rho_0 z_2(1 - e^{z_1 - \overline{w}_1(z_2/\overline{w}_2)^{1/\rho_2}}) + z_1(1 - e^{z_2 - \overline{w}_2(z_1/\overline{w}_1)^{\rho_1}})]}. \end{aligned} \quad \textbf{(A4)}$$

The timing constraint on (z_1, z_2) now follows. The length of the war of attrition during the first phase plus the length of the pure-fighting interval must sum to T: $\sigma_1^*(z_1, \underline{w}_1, \underline{w}_2, \rho_0, k_0, z_1, z_2) + \lambda(z_1, z_2) = T$ where σ_1^* is given by Equation (A3).

Proposition 1 demonstrates that a unique PBE corresponds to each (x_1, x_2) that satisfies M's incentive constraint and the timing constraint. In equilibrium, types $w_1 \in (\underline{w}_1, x_1]$ and $w_2 \in (\underline{w}_1, x_2]$ fight a war of attrition with cost ratio ρ_0 and cost k_0 lasting until time $T - \lambda(x_1, x_2)$. An interval of pure fighting then ensues lasting from $T - \lambda$ to T. M joins 1 with probability α_1 at T. Types $w_1 \in (x_1, \overline{w}_1]$ and $w_2 \in (x_2, \overline{w}_2]$ then fight a war of attrition with cost ratio ρ_1 and total cost k_1 for the rest of the game. M supports 2 with probability $1 - \alpha_1$, and $w_1 \in (x_1, \overline{w}_1]$ and $w_2 \in (x_2, \overline{w}_2]$ then fight a war of attrition with cost ratio ρ_2 and total cost k_2.

Proposition 1 *At least one* (x_1, x_2) *satisfies both M's incentive constraint and the timing constraint, and a unique PBE corresponds to each* (x_1, x_2) *satisfying these constraints. The* *equilibrium strategies are given by*

$$\sigma^*_{j|n}(w_j) = \begin{cases} \sigma^*_j(w_j, \underline{w}_1, \underline{w}_2, \rho_0, k_0, x_1, x_2) & \text{if} \quad w_j \in (\underline{w}_j, x_j] \\ T + \sigma^*_j(w_j, x_1, x_2, \rho_n, k_n, \overline{w}_1, \overline{w}_2) & \text{if} \quad w_j \in (x_j, \overline{w}_j] \end{cases},$$

$$\alpha_1(x_1, x_2, \rho_0) = \frac{\rho_0 x_2(1 - e^{x_1 - \overline{w}_1 (x_2/\overline{w}_2)^{1/\rho_2}})}{\rho_0 x_2(1 - e^{x_1 - \overline{w}_1 (x_2/\overline{w}_2)^{1/\rho_2}}) + x_1(1 - e^{x_2 - \overline{w}_2 (x_1/\overline{w}_1)^{\rho_1}})}$$

for $j, n = 1, 2$ *where the stop times* $\sigma^*_j(w_j, \underline{w}_1, \underline{w}_2, \rho_0, k_0, x_1, x_2)$ *are given by Equation (A3). Beliefs follow directly from Bayes' law.*

Proof of Proposition 1: It is straightforward to verify that the assessment stated in the proposition is a PBE save for one technicality, namely to show that z_j's strict best response is to stop at $T - \lambda$. Suppose (z_1, z_2) is in the interior of F_{12} and the interval of pure fighting is λ. The equilibrium of the two-actor war of attrition implies that z_1 prefers stopping at at $T - \lambda$ to quitting at any earlier time during the first phase. By construction, z_1 is indifferent between stopping at $T - \lambda$ and T if the probability of winning at T is $\alpha_1[1 - e^{z_2 - \overline{w}_2(z_1/\overline{w}_1)^{\rho_1}}]$. But z_1 will be stopping at the same time as the atom of types of 2 that quit at T, so z_1's probability of winning is actually $\alpha_1[1 - e^{z_2 - \overline{w}_2(z_1/\overline{w}_1)^{\rho_1}}]/2$. As a result, z_1 strictly prefers stopping at $T - \lambda$ to stopping at T. Finally, the equilibrium strategies of the war of attrition in phase three imply that z_1's payoff is decreasing at all $t > T$ since it would be the lowest-payoff type fighting until T. Moreover the limit of z_1's payoff to stopping at $t > T$ as t decreases to T is $z_1\alpha_1[1 - e^{z_2 - \overline{w}_2(z_1/\overline{w}_1)^{\rho_1}}] - \lambda c_1$ which is z_1's payoff to stopping at $T - \lambda$. It follows that z_1's strict best response is to stop at $T - \lambda$.

As for beliefs, note that even if M plays a pure strategy, say $\alpha_1 = 1$, then j's beliefs about i following the out-of-equilibrium action of M's supporting 2 must be the same as j's beliefs about i following M's support of 1 (see Fudenberg and Tirole (1991, 332), though they limit their discussion to games with finitely many types). Thus Bayes' rule pins down the factions' beliefs on and off the equilibrium path.

Uniqueness follows from the fact that the continuation games following M's decision are wars of attrition that have a unique equilibrium. Similarly, $w_j \in (\underline{w}_j, z_j]$ are effectively playing a war of attrition that has a unique equilibrium. □

REFERENCES

Abreu, Dilip, and Faruk Gul. 2000. "Bargaining and Reputation." *Econometrica* 68: 85–117.

Acemoglu, Daron, Andrea Vindigni, and Davide Ticchi. 2010. "Persistence of Civil Wars." *Journal of the European Economic Association* 8 (2-3): 664–76.

Acharya, Avidit, and Edoardo Grillo. 2015. "War with Crazy Types." *Political Science Research and Methods* 3 (2): 281–307.

Alesina, Alberto, and Allan Drazen. 1991. "Why are Stabilizations Delayed?" *American Economic Review* 81 (5): 1170–88.

Ashworth, Scott, and Ethan Bueno de Mesquita. 2006. "Monotone Comparative Statics for Models of Politics." *American Journal of Political Science* 50 (1): 214–31.

Balch-Lindsay, Dylan, and Andrew Enterline. 2000. "Killing Time: The World Politics of Civil War Duration 1820–1992." *International Studies Quarterly* 44 (4): 615–42.

Balch-Lindsay, Dylan, Andrew J. Enterline, and Kyle A. Joyce. 2008. "Third-Party Intervention and the Civil War Process." *Journal of Peace Research* 45 (3): 345–63.

Berman, Eli, Jacob Shapiro, and Joseph Felter. 2011. "Can Hearts and Minds Be Bought?" *Journal of Political Economy* 119 (4): 766–819.

Biddle, Stephen, Jeffrey A. Friedman, and Jacob N. Shapiro. 2012. "Testing the Surge: Why Did the Violence in Iraq Decline in 2007?" *International Security* 37 (1): 7–40.

Bliss, Christropher, and Barry Nalebuff. 1984. "Dragon-Slaying and Ballroom Dancing: The Private Supply of a Public Good." *Journal of Public Economics* 25 (1–2): 1–12.

Bulow, Jeremy, and Paul Klemperer. 1999. "The Generalized War of Attrition." *American Economic Review* 89 (1): 175–89.

Cassella, Alessandra, and Barry Eichengreen. 1996. "Can Foreign Aid Accelerate Stabilisation?" *The Economic Journal* 106 (436): 605–19.

Christia, Fotini. 2012. *Alliance Formation in Civil Wars*. New York: Cambridge University Press.

Central Intelligence Agency. 2002. *Balkan Battlegrounds: A Military History of the Yugoslav Conflict, 1990–1995*, vol. 1. Washington, DC: CIA Office of Russian and European Analysis.

Collier, Paul, Anke Hoeffler, and Mans Soderbom. 2004. "On the Duration of Civil War." *Journal of Peace Research* 41 (3): 253–73.

Cunningham, David E. 2010. "Blocking Resolution: How External States Can Prolong Civil Wars." *Journal of Peace Research* 47 (2): 115–27.

Dekel, Eddie, Matthew Jackson, and Asher Wolinsky. 2008. "Vote Buying: General Elections." *Journal of Political Economy* 116 (2): 351–80.

Dekel, Eddie, Matthew Jackson, and Asher Wolinsky. 2009. "Vote Buying: Legislatures and Lobbying." *Quarterly Journal of Political Science* 4 (2): 103–28.

Deutsch, Karl. 1964. "External Involvement in Internal War." In *Internal War: Problems and Approaches*, ed. Eckstein Harry. New York: Free Press, 100–10.

Fearon, James D. 1995. "Rationalist Explanations of War." *International Organization* 49(3): 379–414.

Fearon, James. 1998. "Bargaining, Enforcement, and International Cooperation." *International Organization* 52 (Spring): 269–305.

Fearon, James D. 2004. "Why Do Some Civil Wars Last So Long?" *Journal of Peace Research* 41 (3): 275–302.

Fearon, James D. 2007. "Fighting Rather than Bargaining." Manuscript, Department of Political Science, Stanford University. Available at http://www.stanford.edu/~jfearon/papers/fearon%20ciar0607.pdf.

Fearon, James, and David Laitin. 2008. "Civil War Termination." Manuscript, Department of Political Science, Stanford University. Available at http://www.stanford.edu/jfearon/papers/termination.pdf.

Findley, Michael G., and Josiah F. Marineau. 2014. "Lootable Resources and Third-Party Intervention into Civil Wars." *Conflict Management and Peace Science* 32 (5): 465–486.

Fudenberg, Drew, and Jean Tirole. 1986. "A Theory of Exit in Duopoly." *Econometrica* 54 (4): 943–60.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge, MA: MIT Press.

Gent, Stephen. 2008. "Going in When it Counts: Military Intervention and the Outcome of Civil Conflicts." *International Studies Quarterly* 52 (4): 713–35.

Ghemawat, Pankaj, and Barry Nalebuff. 1985. "Exit." *RAND Journal of Economics* 16 (2): 184–94.

Gleditsch, Kristian Skrede. 2007. "Transnational Dimensions of Civil War." *Journal of Peace Research* 44 (3): 293–309.

Humphreys, Macarten. 2005. "Natural Resources, Conflict and Conflict Resolution." *Journal of Conflict Resolution* 49 (4): 508–37.

Joyce, Kyle, Fatan Ghosn, and Resat Bayer. 2013. "When and Whom to Join: The Expansion of Ongoing Violent Interstate Conflicts." *British Journal of Political Science* 44 (1): 205–38.

Kennan, John, and Robert Wilson. 1989. "Strategic Bargaining Models and Interpretation of Strike Data." *Journal of Applied Econometrics* 4 (S1): S87–S130.

Krainin, Colin. 2014a. "Multilateral War Bargaining." *Quarterly Journal of Political Science* 9 (4): 407–39.

Krainin, Colin. 2014b. "Preventive War as a Result of Long Term Shifts in Power." Working paper.

Leventoğlu, Bahar, and Branislav Slantchev. 2007. "The Armed Peace: A Punctuated Equilibrium Theory of War." *American Journal of Political Science* 51 (4): 755–71.

Lujala, Paivi, Nils Petter Gleditsch, and Elisabeth Gilmore. 2005. "A Diamond Curse: Civil War and a Lootable Resource." *Journal of Conflict Resolution* 49 (4): 538–62.

Malkasian, Carter. 2013. *War Comes to Garmser*. London: Hurst.

Morrow, James D. 1991. "Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances." *American Journal of Political Science* 35 (4): 904–33.

Nalebuff, Barry, and John Riley. 1985. "Asymmetric Equilibria in the War of Attrition." *Journal of Theoretical Biology* 113 (3): 517–27.

Ponasati, Clara, and Jozsef Sakovics. 1995. "The War of Attrition with Incomplete Information." *Mathematical Social Sciences* 29 (3): 239–54.

Powell, Robert. 1999. *In the Shadow of Power*. Princeton, NJ: Princeton University Press.

Powell, Robert. 2006. "War as a Commitment Problem." *International Organization* 60 (1): 169–204.

Powell, Robert. 2012. "Persistent Fighting and Shifting Power." *American Journal of Political Science* 56 (3): 620–37.

Powell, Robert. 2013. "Monopolizing Violence and Consolidating Power." *Quarterly Journal of Economics* 128 (2): 807–59.

Regan, Patrick. 2000. *Civil Wars and Foreign Powers: Outside Intervention in Intrastate Conflict*. Ann Arbor, MI: University of Michigan Press.

Regan, Patrick M. 2002. "Third-Party Interventions and the Duration of Intrastate Conflicts." *Journal of Conflict Resolution* 46 (1): 55–73.

Regan, Patrick, and Aysegul Aydin. 2008. "Diplomacy and Other Forms of Intervention in Civil Wars." *Journal of Conflict Resolution* 50 (5): 736–56.

Ross, Michael. 2004. "How Do Natural Resources Influence Civil War? Evidence from Thirteen Cases." *International Organization* 58 (1): 35–67.

Ross, Michael. 2006. "A Closer Look at Oil, Diamonds, and Civil War." *Annual Review of Political Science* 9: 265–300.

Salehyan, Idean, Kristian Skrede Gleditsch, and David E. Cunningham. 2011. "Explaining External Support for Insurgent Groups." *International Organization* 65 (4): 709–44.

Seymour, Lee. 2014. "Why Factions Switch Sides in Civil Wars." *International Security* 39 (2): 92–131.

Smith, Alastair. 1995. "Aliance Formation and War." *International Studies Quarterly* 39 (4): 405–25.

Walter, Barbara. 2002. *Committing to Peace*. Princeton, NJ: Princeton University Press.

Walter, Barbara F. 2009. "Bargaining Failures and Civil War." *Annual Review of Political Science* 12: 243–61.

Waltz, Kenneth. 1979. *Theory of International Politics*. Long Grove, IL: Waveland Press.

Werner, Suzanne. 2000. "Deterring Intervention: The Stakes of War and Third-Party Involvement." *American Journal of Political Science* 44 (4): 720–32.