Wars of Attrition with Evolving Payoffs

Germán Gieczewski*

June 2019
PRELIMINARY AND INCOMPLETE.

Abstract

I analyze a model of wars of attrition with evolving payoffs. Two players fight over a prize by paying state-dependent flow costs until one player surrenders. The state of the world is commonly observed and evolves over time. The equilibrium is unique and uses threshold strategies: each player surrenders when the state is unfavorable enough to her, while for intermediate states both players strictly prefer to fight on. Taken as a refinement of the model of wars of attrition with complete information, this model makes related but distinct predictions from the standard reputation-based refinements (Abreu and Gul, 2000). The model is versatile and can be tractably extended to study partial concessions, commitment devices and deadline effects.

1 Introduction

The canonical model of wars of attrition is applicable to a variety of phenomena in economics and political science. Two firms in a duopoly engaging in a price war; an activist group boycotting a firm; two countries, or two factions in a civil war, engaging in a protracted military conflict; an army besieging a city, or two political parties threatening to shut down the government over budgetary disagreements can all be understood as wars of attrition. Previously, such models were applied in biology to study animal conflicts (Smith, 1974). In all these settings, we are typically interested in the answer to three questions. Who is the likely winner of the war? How much delay is there before the loser concedes or, in other words, how inefficient is the outcome? How do these results change as we vary the parameters (or other details of the model)?

The simplest model dynamic model of a war of attrition assumes that the payoffs from winning, as well as the flow costs of continuing the war, are common knowledge and fixed over

*Department of Politics, Princeton University.
time. Although natural, these assumptions deliver unsatisfying results. As is known, this model has three equilibria: two where one agent never surrenders while the other surrenders immediately, and a third, totally mixed equilibrium where both players have a positive probability of winning, and are always indifferent about surrendering at any time on the equilibrium path. These equilibria give diametrically opposed predictions as to who will win and what the expected delay is, making the model agnostic on our first two questions unless we take a stand on equilibrium selection. If we require the equilibrium to answer the first question reasonably—in particular, when the game is symmetric, both players should have equal chances of winning—only the mixed equilibrium is acceptable. But, as is also known, the mixed equilibrium has dysfunctional comparative statics: when the benefits and costs are asymmetric, the weaker player (that is, the one with a higher cost/benefit ratio) is more likely to win. This occurs because, in a totally mixed equilibrium, players must be indifferent about surrendering at all times; if a player has high costs, for her to be willing to continue at all, the other player’s surrender rate must be high, and vice versa. In the limit, if one player has arbitrarily high flow costs, she would be expected to win almost surely. This is unreasonable: surely in this case it would be more sensible to simply select the equilibrium where the high-cost player loses.

A more effective approach, which has also been used in related models of bargaining (Abreu and Gul, 2000) and entry deterrence (Kreps and Wilson, 1982; Milgrom and Roberts, 1982), is the introduction of behavioral or commitment types which add reputational concerns to the game. In its simplest form, a war of attrition with reputation follows the same rules as above, except that each player $i$ has a small probability $\epsilon_i > 0$ of being a commitment type that never surrenders. A central result of this literature is that, even when the $\epsilon_i$ are small, so that in all likelihood both players are rational, the temptation to pretend to be a commitment type has a powerful effect on the equilibrium behavior of rational players. In general, the model with reputation has a unique equilibrium, in which the weaker (higher cost, lower prize) player surrenders at once with some probability, and the equilibrium path becomes totally mixed thereafter.

This type of model yields more sensible comparative statics than the totally mixed equilibrium from the benchmark model, but it has two limitations. First, its predictions concerning the winner of the war are sensitive to the values chosen for the $\epsilon_i$: the player with a lower probability of being irrational is at a disadvantage (and a player who is rational for sure loses with certainty, even if the other player’s probability of being irrational is very low and her relative cost-benefit ratio is higher). This is problematic if we take the $\epsilon_i$ as being not a descriptively true, measurable feature of the practical application in question, but rather a perturbation used for equilibrium refinement, as there is then little hope of predicting a
winner confidently. Second, the elegance of this approach wanes in more complicated games: for example, in a war of attrition in which players can make partial concessions, employ commitment devices, or exert effort to increase the other player’s costs, or in bargaining games where players can make different offers over time, devices in bargaining applications where players can make different offers over time, it’s no longer obvious what set of commitment types is appropriate, and decisions regarding this may affect behavior on the equilibrium path.

In this paper, I develop a new model of dynamic wars of attrition that addresses these concerns. In contrast with previous work, the model features an evolving state of the world which is commonly observed. Concretely, costs are influenced by a one-dimensional parameter $\theta_t$, which changes over time following a stochastic process. Intuitively, $\theta_t$ parameterizes the extent to which current conditions in the war favor the first player over the second. This device allows us to select a unique equilibrium despite the fact that there is complete information on the equilibrium path; in particular, we do not rely on reputational concerns. The selected equilibrium also has a novel structure, underlying intuition and implications: when $\theta_t$ starts at an intermediate value, both players choose to continue the war initially, and they have strict incentives to keep doing so. Eventually, when $\theta_t$ reaches an extreme enough value, the player unfavored by the current conditions surrenders. (Of course, if $\theta_t$ starts at an extreme value, the unfavored player surrenders immediately.) As a result, the model predicts an equilibrium where both players have a positive probability of winning, yet the equilibrium path is deterministic (except for the exogenous variation in $\theta_t$) rather than totally mixed, and both players obtain strictly positive payoffs from following their strategies, compared to their utility from surrendering immediately.

I argue that the payoff perturbation driving the model can be interpreted in two ways. On the one hand, it can be taken as a feature that reflects a real phenomenon present in practical examples: for example, the combatants in a war would know how the weather, the state of their provisions, etc. is changing over time and how that affects their odds; firms in a price war would know how long they can afford to continue it for, given their financial health. In this vein, a serious attempt at modeling wars of attrition including this perturbation would entail estimating the process driving the underlying parameter $\theta_t$, and including that in the model. On the other hand, as often done in reputational models, the changes in $\theta_t$ can be taken as a purely theoretical tool to obtain an equilibrium refinement. Under this interpretation, we are interested in the results of the limiting case as the rate of change in $\theta_t$ goes to zero. I perform this exercise and obtain three interesting results. First, the limiting model predicts a unique equilibrium for the unperturbed game, which—unlike the case in reputational models—is independent of the specified shape of the perturbations but
depends only on the fundamentals, i.e., the costs and benefits of each player. The predicted equilibrium when the cost/benefit ratios are different is simple: the stronger player (that with the lower ratio) wins immediately.\footnote{This is also predicted by the model with reputation in the limiting case, but only if the $\epsilon_i$ are taken to go to zero at a similar enough rate.} But the predicted equilibrium in the symmetric case is novel. The simplest way to describe it is to use tokens: essentially, it is as though the players play a game in every turn, where the winner takes a token from the loser. If one player runs out of tokens, she surrenders. This equilibrium gives both players the same chance of winning, but it’s not mixed and the players get strictly higher payoffs than in the totally mixed equilibrium. Equivalently, the expected delay is lower than in the totally mixed equilibrium, but it is guaranteed to be strictly positive.

Additionally, I study an extension of the model where one player has the option to make a partial concession, lowering her own payoff but decreasing the other player’s incentive to continue the war. For example, a city under siege might give some of its wealth to the attacking army to lower the expected value of looting the city if it is taken. This example is interesting in its own right, but also serves to illustrate how the simplicity of model can be leveraged to study more complex questions. I find that using this option can put the conceding player at a great advantage: for example, in the limit case where the state evolves very slowly, the conceding player can concede just enough to induce the opponent to immediately surrender. Finally, I study the general game where both players can make partial concessions over time.

This paper is connected to the existing literature on wars of attrition and bargaining. Indeed, many papers have studied variants of the basic war of attrition involving a perturbation. However, most of them have relied on various kinds of private information–generating reputational concerns–rather than evolving payoffs to obtain their results. For example, Krishna and Morgan (1997) and Ponsati and Sákovics (1995) consider models where the players’ benefits at the end of a war depend on their types, which are private information. In Fudenberg and Tirole (1986) and Egorov and Harstad (2017), it is the flow costs of continuing the war which depend on private types. Some of these models obtain a unique equilibrium while others still display multiplicity, depending on further assumptions. The framework we take as a benchmark for reputation-based perturbations is closest to Chatterjee and Samuelson (1987) and Abreu and Gul (2000).

A few papers have employed modeling strategies closer to the one used here. Ortner (2016) studies a game of bargaining with alternating offers where the bargaining protocol (i.e., the identity of the player making offers) is driven by a Brownian motion. Ortner (2017) considers two political parties bargaining in the shadow of an election; if an agreement is
reached, the result affects the relative popularity of the parties, which otherwise evolves as a Brownian motion and eventually determines the outcome of the election. In Ortner (2013), optimistic players bargain over a prize whose value changes over time. In Slantchev (2003), players bargain using alternating offers while simultaneously fighting; the course of the war affects the offers that players will accept. In contrast, I assume that the changing state of the world affects flow costs, and use this perturbation to obtain a unique, intuitive equilibrium in a model which otherwise displays equilibrium multiplicity—that is, the war of attrition—rather than to alternating-offers bargaining.

The present paper also contributes to the literature on supermodular games started by Topkis (1979) and Milgrom and Roberts (1990). Indeed, our modeling technique exploits the fact that the war of attrition is a supermodular game when players’ strategies are ordered in opposite ways (i.e., player 1’s “high” strategy is to continue while player 2’s is to surrender); and the fact that a perturbation produces equilibrium uniqueness is reminiscent of similar contagion-based results in global games (e.g., Burdzy, Frankel and Pauzner (2001)).

The paper proceeds as follows. Section 2 presents the benchmark model, and Section 3 characterizes its equilibrium. Section 4 shows how the results translate to a continuous time model. Section 5 takes the limit as the payoff perturbation becomes arbitrarily small, so that the model converges to the basic war of attrition. Section 6 compares the predictions of the model with those from the basic war of attrition and models with reputational concerns. Section 7 discusses the case with partial concessions, and Section 8 concludes. All proofs can be found in Appendix A.

2 The Model

There are two players, 1 and 2. For clarity, we first study a discrete-time model with infinite horizon: $t = 0, 1, \ldots$. In each period, each player can choose to continue (C) or surrender (S). There is a state of the world $\theta_t \in [-M, M]$ which is common knowledge at all times. $\theta_t$ represents how favorable the current conditions in the war are to either player: a high $\theta$ favors 2, while a low $\theta$ favors 1. The initial $\theta_0$ is given by Nature.\[^{2}\] Then it evolves according to a local Markov process, that is,

$$P(\theta_{t+1} - \theta_t \leq x | \theta_t) = H_{\theta_t}(x),$$

$^{2}\theta_0$ is a fixed parameter. Assuming it to be random does not change the results, since it is immediately revealed to both players.
where $H_\theta : \mathbb{R} \to [0, 1]$ is an absolutely continuous distribution function with corresponding density $h_\theta$. Let $\mu(\theta)$ be the expected value of a random variable with distribution $H_\theta$, and let $\sigma^2(\theta)$ be its variance. Fix $\eta > 0$. We assume that:

**A1** $h_\theta$ is continuous in $\theta$ for all $\theta \in [-M, M]$. More precisely, the mapping $\theta \mapsto h_\theta$ is continuous, taking the 1-norm in the codomain.

**A2** $H_\theta$ is weakly FOSD-increasing in $\theta$ for all $\theta \in [-M + \eta, M - \eta]$. In other words, if $\theta, \theta' \in [-M + \eta, M - \eta]$ and $\theta \geq \theta'$ then $H_\theta \leq H_{\theta'}$.

**A3** For all $\theta \in [-M, M]$, $\text{supp}(h_\theta)$ is a closed interval with nonempty interior, contained in $[-\eta, \eta] \cap [\theta - M, M - \theta]$.

Substantively, A2 says that $\theta$ tends to drift to the extremes rather than to the middle, while A3 guarantees that $\theta_{t+1}$ will stay within $[-M, M]$.

Payoffs are as follows. In any period where the war is ongoing, each player $i$ pays a flow cost $c_i(\theta)$. If either player chooses to surrender at time $t$, players do not pay flow costs on that period and the war ends. When the war ends, the winner receives an instantaneous payoff $H_i$ and the loser receives $L_i < H_i$, normalized to 0.\(^3\) If both players surrender on the same turn, both players lose.\(^4\) Both players have a common discount factor $\delta \in [0, 1]$.\(^5\)

Hence, $i$’s lifetime payoff if the war ends at time $T$ is:

$$U_i(\sigma_i, \sigma_j) = \delta^T H_i \mathbb{1}_{\{i \text{ wins}\}} - \sum_{t=0}^{T-1} \delta^t c_i(\theta_t).$$

We assume the following about the players’ payoff functions:

**B1** $c_1(\theta)$ is strictly increasing in $\theta$, and $c_2(\theta)$ is strictly decreasing in $\theta$.

**B2** $c_1(\theta)$, $c_2(\theta)$ are $C^1$.

**B3** There are $M_1$, $M_2$ such that $-M < -M_1 < 0 < M_2 < M$, $c_1(-M_1) = 0$ and $c_2(M_2) = 0$.

---

\(^3\)A player who receives $H_i$ from winning, $L_i$ from surrendering and $-c_i(\theta)$ from continuing has equivalent incentives to one who receives $H_i + \rho$ from winning, $L_i + \rho$ from losing and $-c_i(\theta) + \rho(1 - \delta)$ from continuing.

\(^4\)This assumption simplifies some formal arguments but is not essential in part: equilibrical, it ent to assis equivaafter simultaaneous surrender each player has a probability $\frac{1}{2}$ of winning. This is because,um, players will never surrender simultaneously.

\(^5\)\(\delta = 1\) does not lead to technical issues in terms of defining utility since it is never rational to continue the war forever.
The following inequalities are satisfied:

\[ c_2(M_1) \frac{M - M_1 - \eta}{\eta} > H_2 \]
\[ c_1(M_2) \frac{M - M_2 - \eta}{\eta} > H_1. \]

\[ H_i > \frac{-c_i(\theta)}{1-\delta} \] for \( i = 1, 2 \) and all \( \theta \in [-M, M] \).

Substantively, these assumptions say that player 1 is favored when \( \theta \) is low, and vice versa (B1); for favorable enough values of \( \theta \), players actively enjoy fighting (B3), and under such circumstances it will be best for the other player to surrender (B4); and players never enjoy fighting more than they do winning immediately (B5).

3 Analysis

We can now characterize the equilibrium of the game:

**Proposition 1 (Equilibrium Characterization).** There is an essentially unique subgame perfect equilibrium (SPE). This equilibrium is characterized by thresholds \( \theta_* < \theta^* \), as follows: player 1 surrenders whenever \( \theta_t > \theta^* \), player 2 surrenders whenever \( \theta_t < \theta_* \), and neither player surrenders when \( \theta_t \in (\theta_*, \theta^*) \).

The intuition behind the solution is as follows. First, if \( \theta \) is very low, then player 1 never has a myopic incentive to surrender (i.e., even if she expected to eventually lose, she would stay in the war until \( \theta \) goes above \( -M_1 \)). Then, since the gap between \( -M_1 \) and \( -M \) is large, there will be low enough values of \( \theta \) for which 2 is forced to surrender immediately, as she understands that waiting for 1 to surrender will cost more than the rewards are worth. Similarly, for very high \( \theta \), 1 must concede immediately. In other words, player 1 has two dominance regions: one with low \( \theta \)'s that force her to continue and another with high \( \theta \)'s that force her to surrender. The reverse is true for player 2.

For values of \( \theta \) between these tentative thresholds, behavior will depend on expectations about the other player’s strategy. For instance, if 2 plays a “hawkish” strategy in equilibrium (that is, she surrenders only for a small set of values of \( \theta \)), this gives 1 incentives to play a “dovish” strategy (which surrenders at a large set of values of \( \theta \)) and vice versa. Since the payoff of winning is worth waiting for a few periods, there must be a gap between the surrender regions of both players. We call this intermediate region where both players fight the disputed region. Now consider a strategy profile where 1 is hawkish and 2 is dovish. Intuitively, this is unstable because the disputed region would be an interval of high \( \theta \)'s,
where 2 has lower flow costs than 1; therefore, if the disputed region is small, 2 would be willing to expand it by reducing her surrender region; or, if the disputed region is large, 1 would want to shrink it by expanding her surrender region. The opposite problem occurs if 1 is dovish and 2 is hawkish, so the thresholds \( \theta_*, \theta^* \) delimiting the disputed region are uniquely determined.\(^6\)

We can interpret the values \( \theta_*, \theta^* \) as parameterizing two features of the equilibrium. On the one hand, the size of the disputed region, \( \theta^* - \theta_* \), reflects how willing the players are to fight to increase their odds of winning the war, which is a function of the ratio between \( H_i \) and the flow costs of fighting in the disputed region. In addition, if \( \mu(\theta) \) is increasing, this tends to shrink the disputed region, as it leads an unfavorable \( \theta \) to “snowball” and become even more unfavorable over time, prompting a quicker surrender. On the other hand, the position of the disputed region \([\theta_*, \theta^*]\) within the interval \([-M, M]\) reflects any asymmetries between the players: if \( h \) and the flow costs are symmetric then \( \theta^* + \theta_* = 0 \), while if \( \mu > 0 \) (i.e., \( \theta \) drifts to the right) or \( c_1(\theta) > c_2(-\theta) \) (i.e., 1 has higher costs) then \( \theta^* + \theta_* < 0 \) and so on. We summarize these results in the following

**Proposition 2** (Comparative Statics). The thresholds respond to changes in parameters like so:

- Increases in \( H_1 \) and decreases in \( c_1 \) raise \( \theta_* \), \( \theta^* \) and \( \theta^* - \theta_* \). Increases in \( H_2 \) and decreases in \( c_2 \) lower \( \theta_* \) and \( \theta^* \) but raise \( \theta^* - \theta_* \).

- If \( H_1, H_2 \) increase proportionally, \( \theta_* \) decreases and \( \theta^* \) increases.

- An increase in \( H_\theta \) for all \( \theta \) (in the FOSD sense) lowers \( \theta_* \) and \( \theta^* \).

### 4 Continuous Time Model

We now make an adjustment to the above game. Suppose that we split time into smaller periods; in particular, the length of each period goes from 1 to \( \kappa^2 \). The density of \( \theta' - \theta | \theta \) is now given by \( \frac{1}{\kappa} h_\theta (\frac{\kappa}{\kappa} + \mu(\theta)(1 - \kappa)) \), with mean \( \mu(\theta)\kappa^2 \) and variance \( \sigma^2\kappa^2 \); and the flow costs are now \( c_i(\theta, \kappa) = c_i(\theta)\kappa^2 \). These adjustments guarantee that the drift and variance of the stochastic process, as well as the flow costs, remain constant when measured by real time. In the limit, the stochastic process governing \( \theta \) converges to a Brownian motion with local drift \( \mu(\theta) \) and local variance \( \sigma^2 \). We then have the following

\(^6\)The formal proof also rules out cases where the surrender and disputed regions are not intervals.
Proposition 3 (Continuous Time Equilibrium). Let \( \theta_*(\kappa), \theta^*(\kappa) \) be the equilibrium thresholds from Proposition ?? as a function of \( \kappa \). Then there are \( \theta_{s0}, \theta^{*0} \) such that \( \theta_*(\kappa) \to \theta_{s0} \) and \( \theta^*(\kappa) \to \theta^{*0} \) as \( \kappa \to 0 \).

In addition, each player’s expected utility \( U_i(\theta) \) and probability of winning \( P_i(\theta) \), starting at a given \( \theta \), can be calculated in \( [\theta_{s0}, \theta^{*0}] \) as the solution to the following ODEs:

\[
U''_i(\theta) = \frac{2}{\mu(\theta)^2 + \sigma^2} (c_i(\theta) - \mu(\theta)U'_i(\theta))
\]

\[
P''_i(\theta) = -\frac{2}{\mu(\theta)^2 + \sigma^2 \mu(\theta)} P'_i(\theta)
\]

given the boundary conditions \( U_1(\theta_{s0}) = H_1; \ U_2(\theta_{s0}) = 0; \ U_1(\theta^{*0}) = 0; \ U_2(\theta^{*0}) = H_2; \ P_1(\theta_{s0}) = 1; \ P_2(\theta_{s0}) = 0; \ P_1(\theta^{*0}) = 0; \ P_2(\theta^{*0}) = 1. \)

In particular, when \( \mu \equiv 0 \), the solution in \( [\theta_{s0}, \theta^{*0}] \) reduces to

\[
U_1(\theta) = \frac{2}{\sigma^2} \int_{\theta_{s0}}^{\theta^{*0}} \left( \int_{\lambda}^{\theta^{*0}} c_1(\omega)d\omega \right) d\lambda = \frac{2}{\sigma^2} \int_{\theta_{s0}}^{\theta^{*0}} (\lambda - \theta)c_1(\lambda)d\lambda \quad P_1(\theta) = \frac{\theta^{*0} - \theta}{\theta^{*0} - \theta_{s0}}
\]

\[
U_2(\theta) = \frac{2}{\sigma^2} \int_{\theta_{s0}}^{\theta} \left( \int_{\theta_{s0}}^{\lambda} c_2(\omega)d\omega \right) d\lambda = \frac{2}{\sigma^2} \int_{\theta_{s0}}^{\theta} (\theta - \lambda)c_2(\lambda)d\lambda \quad P_2(\theta) = \frac{\theta - \theta_{s0}}{\theta^{*0} - \theta_{s0}},
\]

and the thresholds \( \theta_{s0}, \theta^{*0} \) are determined by the conditions \( U_1(\theta_{s0}) = H_1, \ U_2(\theta^{*0}) = H_2. \)

Figure 1: Equilibrium utility and winning probabilities

Figure 1 illustrates the expected utility and winning probabilities as a function of the starting \( \theta \), in two cases: when the stochastic process is symmetric (\( \mu \equiv 0 \)) and when it is asymmetric (\( \mu \equiv 0.025\kappa \)). In both examples, we take \( H = 2, \sigma = 0.5, \kappa = 0.25, \)
\[ c_1(\theta) = (\theta + 5) \frac{\kappa^2}{1000} \text{ and } c_2(\theta) = (-\theta + 5) \frac{\kappa^2}{1000}. \]

In particular, the cost functions of both players are symmetric around 0, i.e., \( c_1(\theta) = c_2(-\theta). \) Hence, as might be expected, the thresholds and utilities in Figure 1a are also symmetric around 0. Moreover, the probabilities of winning \( P_i(\theta) \) are linear as per Proposition 3. On the other hand, in Figure 1b, \( \theta \) tends to drift up over time, which structurally favors player 2, whence both thresholds are much lower. Figure 2 shows an example of a typical equilibrium path, under the same parameter assumptions as Figure 1a and starting at \( \theta_0 = 0 \): the state of the world is initially between the two thresholds, so both players continue the war until \( \theta \) reaches one of them, in this case player 1’s.

## 5 Equilibrium in the Limit

When applying the model, the above analysis is relevant if we consider the movement of \( \theta \) to be a significant feature of the application. However, the model also admits being used simply as an equilibrium selection tool when the “true” model the researcher is interested in is the basic war of attrition without perturbations. In this interpretation, we must take the limit of the solution as the movement of \( \theta \) becomes arbitrarily slow. Formally, we apply the same transformations as in the previous Section to the stochastic process: the density of \( \theta' - \theta \) is now \( \frac{1}{\nu} h_{\theta} \left( \frac{x}{\nu} + \mu(\theta)(1 - \nu) \right) \), with mean \( \mu(\theta) \nu^2 \) and variance \( \sigma^2 \nu^2 \); but the costs are left unchanged, and we take the limit as \( \nu \to 0 \). In this case the fundamentals of the game are no longer the same, as the movement of \( \theta \) has been slowed relative to the size of the flow costs.

Naturally, the limit of the game as \( \nu \to 0 \) is simply the unperturbed war of attrition, but taking the limit of the perturbed equilibria yields a uniquely selected equilibrium of the
Proposition 4 (Equilibrium Selection with Slow-Moving Processes). Suppose $H_1 = H_2 = H$.\textsuperscript{7} Let $\theta^l$ be such that $c_1(\theta^l) = c_2(\theta^l)$. Suppose $c'_1(\theta^l) = -c'_2(\theta^l)$. Let $\theta^*(\nu), \theta_*(\nu)$ be the thresholds of the equilibrium in Proposition ?? as a function of $\nu$. Then $\theta^*(\nu), \theta_*(\nu) \to \theta^l$ as $\nu \to 0$.

Hence, if $\theta_0 < \theta^l$ the perturbed equilibria converge, as $\nu \to 0$, to an equilibrium of the unperturbed game where 1 wins instantly. If $\theta_0 > \theta^l$, then 2 wins instantly in the limit.

If $\theta_0 = \theta^l$ the perturbed equilibria converge to an equilibrium of the basic war of attrition augmented with tokens. In the augmented game, players observe a payoff-irrelevant variable $\phi_t$ that follows a random walk, starting at 0 and obeying $\phi_{t+1} = \phi_t + 1, \phi_{t+1} = \phi_t - 1$ with probabilities 0.5, 0.5. The limiting equilibrium is described by a unique threshold $K$ such that 1 surrenders at any history where $\phi_t \geq K$, and 2 surrenders whenever $\phi_t \leq -K$.

In this equilibrium, both players have probability 0.5 of winning ex ante, and there is delay, but both players are strictly willing to fight, and i’s expected payoff ex ante is $H_4 > 0$.

In other words, as the perturbation vanishes, the model offers a stark prediction for what equilibrium should be selected in the unperturbed game: if one player has a higher flow cost\textsuperscript{8} than the other, then the equilibrium where she always surrenders immediately, and the other player is never expected to surrender, should be played. This is not surprising, since the perturbed game has a unique equilibrium, and any family of unique equilibria for the unperturbed game displaying the right comparative statics (i.e., higher-cost players should be more likely to lose) must make this selection. More interestingly, though, the symmetric equilibrium selected when costs are equal is not the totally mixed equilibrium of the unperturbed game, and in fact is not an equilibrium of the unperturbed game, as it requires a coordination mechanism. Indeed, the equilibrium is as if each player were initially in possession of $K$ tokens, and a stage game were played every period; both players would have equal probability of winning in each stage game, and the loser would pay one token to the winner. (In terms of the formal statement, $\phi_t + K$ is the amount of tokens held by player 2 at time $t$.) Then the selected equilibrium prescribes that a player should surrender whenever she runs out of tokens.

6 Discussion

In this Section we discuss the merits of the model’s results and compare them to existing models of wars of attrition, namely the basic unperturbed model and an alternative based

\textsuperscript{7}This serves to simplify notation.
\textsuperscript{8}In general, since the benefits from winning may differ, what matters is the cost-to-benefit ratio.
on reputational concerns. Although these alternatives are not novel, for completeness we briefly define the models and state their basic results; the details are in Appendix A. We focus on the continuous time versions of both models as the discrete time versions introduce some technical complications which are not relevant to our purposes.\footnote{Namely, the discrete time basic war of attrition has a partially mixed equilibrium in addition to the totally mixed one, where players take turns mixing vs. continuing for sure; this equilibrium converges to the totally mixed equilibrium in the continuous time limit. Similarly, the equilibrium of the war of attrition with reputation involves alternation in mixing.}

### 6.1 Basic War of Attrition

The basic, unperturbed war of attrition corresponds to a special case of our model where $\mu = \sigma = 0$, i.e., $\theta$ is constant. Alternatively, we can take $c_1(\theta) \equiv c_1$ and $c_2(\theta) \equiv c_2$ to be flat. The following Proposition summarizes its set of SPE:

**Proposition 5** (Equilibria of the Basic War of Attrition). The basic war of attrition has three SPE:

- An equilibrium where 1 surrenders at every history and 2 never surrenders.

- The opposite equilibrium where 1 wins immediately.

- A totally mixed equilibrium where both players mix at every history, and $i$ chooses to surrender at a rate $p_i = \frac{c_i}{H}$ at every $t$. Players are indifferent about continuing. Players’ expected payoffs ex ante are 0. $i$’s probability of winning is $\frac{c_i}{c_i + c_j}$.

As noted in the Introduction, this model has two main shortcomings. First, it provides no way to select an equilibrium from the options offered, which saps it of explanatory power as the set of equilibria includes extremes where either player wins for sure. Second, the mixed equilibrium seems reasonable in the symmetric case, but it clearly has dysfunctional comparative statics away from it: indeed, a player’s probability of winning increases with her own cost, and in the limit where one player has much higher costs than the other, that player wins almost surely. These results derive from the mechanics driving the equilibrium: since both players must be indifferent along the equilibrium path, if $i$ has high costs, then $j$ must surrender at a high rate to keep $i$ willing to continue with positive probability. This has little to do with the natural intuition behind the game, namely, that if $i$ knows $j$ has lower costs, she might conclude that she probably cannot convince $j$ to concede, leading to $i$’s surrender.

The perturbed model overcomes these issues: the movement of $\theta$ results in a unique equilibrium being selected. Moreover, changes in the players’ costs are parameterized simply...
by movements in $\theta$, which change the payoffs and probabilities of winning in the natural way, i.e., the player with higher costs is more likely to surrender sooner and vice versa.

In addition, the equilibrium selected by the perturbed game in the symmetric case (or, away from the slow-movement limit, in other cases where the war continues for some time) produces strictly better welfare than the totally mixed equilibrium of the basic game. As per Proposition 4, the sum of the players’ expected payoffs is $\frac{H}{2}$, whence the expected delay is $\frac{H}{4c}$, versus a total payoff of 0 and expected delay $\frac{H}{2c}$ in the totally mixed equilibrium. It should be noted, though, that the equilibrium selected by the perturbed model requires a coordination device, and if such devices are allowed then even more efficient equilibria are possible. Indeed, the most efficient equilibrium possible would involve the players flipping a coin, with the loser of the coin flip conceding immediately. This would generate a total payoff of $H$ with no delay. It is an interesting fact in its own right, though, that this equilibrium cannot be selected by the perturbed model.

6.2 War of Attrition with Reputation

The war of attrition with reputation takes the basic war of attrition outlined above, but adds for each player $i$ a probability $\epsilon_i$ of being a commitment type that never surrenders. Types are private information. The following Proposition characterizes the SPE of this game:

**Proposition 6** (Equilibrium of the War of Attrition with Reputation). The war of attrition with reputation has a unique SPE, characterized as follows. Let $t_i^*$ be the vanishing time of player $i$, characterized by $e^{-\frac{c_i}{H}t_i^*} = \epsilon_i$. If $t_i^* = t_j^*$, then we say the players are balanced. The probability that $i$ (rational or not) will continue up to time $t$ is $Q_{it} = e^{-\frac{c_i}{H}t}$ for $t \leq t_i^*$. The mass of rational $i$’s surviving up to time $t$ is $P_{it} = e^{-\frac{c_i}{H}t} - \epsilon_i$. The surrender rate for rational $i$’s is $p_{it} = e^{-\frac{c_i}{H}t} - \epsilon_i$. If $t_i^* < t_j^*$, we say that $i$ is stronger than $j$. In this case, a mass of $j$’s rational types of size $P_{j0} = 1 - \frac{\epsilon_j}{c_j}$ surrender immediately at $t = 0$; after that, the players are balanced and the equilibrium continues as above. In this case, the probability of winning for $i$’s and $j$’s rational types are

$$W_i = P_{j0} + \frac{1 - P_{j0}}{1 - \epsilon_i} \left[ \frac{c_i}{c_i + c_j} \left( 1 - e^{-\frac{c_i + c_j}{H}t_i^*} \right) - \epsilon_i \left( 1 - e^{-\frac{c_j}{H}t_i^*} \right) \right] \approx P_{j0} + (1 - P_{j0}) \frac{c_i}{c_i + c_j},$$

$$W_j = \frac{1 - P_{j0}}{1 - \epsilon_j} \frac{c_j}{c_i + c_j} \left( 1 - e^{-\frac{c_i + c_j}{H}t_j^*} \right) - \epsilon_j \left( 1 - e^{-\frac{c_j}{H}t_j^*} \right) \approx (1 - P_{j0}) \frac{c_j}{c_i + c_j}.$$

$i$’s expected payoff is $P_{j0}H$; $j$’s expected payoff is 0.
The mechanics behind the equilibrium are as follows. After \( t = 0 \), rational types must be willing to mix, as contradictions arise in all other cases. As in the unperturbed model, for \( i \) to be willing to mix, \( j \) must be surrendering continuously at a rate \( \frac{c_i}{H} \) (this is \( j \)'s overall surrender rate regardless of her rationality, so \( j \)'s rational types must be surrendering more often than this value to maintain the correct average), and vice versa. At this rate, the rational types of player \( i \) would all surrender by time \( t_i^* \). If these vanishing times were unequal, and e.g. \( t_i^* < t_j^* \), then \( j \)'s rational types would surrender at time \( t_i^* \), which would incentivize \( i \)'s rational types in some interval \([t_i^* - \epsilon, t_i^*]\) not to surrender, and so on. To prevent this, a mass of \( j \)'s rational types surrender at the beginning so that in the continuation the rational types of both players will extinguish themselves at the same time.

As in the unperturbed war of attrition, the higher-cost player fares better (in the sense of surrendering less often) after \( t = 0 \). However, if \( j \) has a higher flow cost, that also results in a higher \( t_j^* \), and therefore a higher probability of surrendering immediately. Overall the latter effect wins, so that having a higher flow cost reduces a player’s equilibrium payoff and probability of winning. However, the vanishing times are also a function of the \( \epsilon_i \), so guaranteeing that a certain relationship between flow costs translates into a certain player likely winning the war requires some assumption on these parameters. Specifically, if \( c_i < c_j \), then as \( \epsilon_i, \epsilon_j \to 0 \) we have that \( P_{j0} \to 1 \), but only if \( \frac{\epsilon_j}{\epsilon_i} \to 0 \). In other words, \( \epsilon_i \) cannot be going to 0 much faster than \( \epsilon_j \); otherwise, this effect would overrule the relationship between the flow costs and \( j \) would be favored to win the war. When applying the model, we might be willing to assume that \( \epsilon_i \approx \epsilon_j \) so that there is no issue, but if the \( \epsilon \)'s are expected to be very small this assumption will be hard to verify or disprove based on any sort of data the researcher has access to.

In this aspect, the model proposed in this paper appears more robust: if we are interested in using the perturbation as an equilibrium selection tool, the results as \( \nu \to 0 \) do not depend on the shape of the functions \( c_i(\theta) \) away from the \( \theta_0 \) we are focused on. They only depend on \( (c_1(\theta_0), c_2(\theta_0)) \) as shown by Proposition 4, so long as the functions are continuous.

Two other interesting differences should be noted, which might be used to distinguish between the models empirically. First, in the symmetric case, the reputational model again predicts the same totally mixed equilibrium as the basic war of attrition, unlike the moving-state model. Second, the density functions predicted for the length of the war \( h(t) \) differ substantially between the two models, as shown in Figure 3. To begin with, the totally mixed equilibrium results in a higher expected delay until the war ends. But, more importantly, in the case of a relatively balanced war, the moving-state model predicts that there is no chance of an immediate surrender, and the rate of surrender is hump-shaped over time: on average it will take some time until \( \theta \) moves enough to one of the extremes so as to make one
player surrender. On the other hand, the reputational model predicts a substantial chance of immediate surrender, followed by exponential decay.

7 Partial Concessions

We now extend the model to study the following variant of the war of attrition: as before, players are able to continue the war until one surrenders, but while the war is ongoing they can affect the payoffs of the eventual outcome by making partial concessions. For instance:

- An army sieges a city. The city has strong walls and cannot be taken by assault, so a siege will lead to two outcomes: either the army surrenders (meaning it leaves with nothing) or the city surrenders (meaning it opens its doors and the army plunders it). In this example, movements in $\theta$ reflect changes in the fortunes of each side: the city’s supplies may dwindle or they may be able to smuggle in fresh supplies, either side may suffer a disease outbreak, and so on. Instead of surrendering outright, the city can instead pay tribute to the army: that is, it can gather some wealth, give it to the army and invite them to leave. If this tactic is successful, it is more efficient for both sides, since the city’s gathering of its own wealth entails lower welfare losses (no buildings are burned, no civilians are killed, and so on). There is no commitment device, so the army can still stay to siege after receiving tribute. But the tactic may work because the remaining loot to be had is smaller, while the city is still eager to defend itself (since the deadweight loss from being plundered is large).
A polluting firm is boycotted by an activist group. The boycott is costly for both players, as it lowers firm sales and consumer surplus. The war ends when either the firm capitulates to the demands or the activists abandon the boycott. However, the firm may have access to a range of policies it can implement to lower its own pollution. A partial concession, in the form of a moderate level of self-regulation, may be enough to “deflate” the momentum of the boycott even if it is not what the activists demanded. Similar logic can apply to a government implementing emergency measures to appease protesters.

The common theme in these examples is that making a partial concession may benefit the party who does it: although it entails giving up some payoffs, this is often worth it if it will tilt the remaining war of attrition towards the side who partially concedes. We ask our model to answer two main questions: when are partial concessions useful? (For instance, are they only useful when they enable higher efficiency than the noncooperative outcomes, as in the siege example? Or can they be useful even when there is no efficiency gain involved?) And what is the optimal size and timing of a partial concession? We first study the case where only player 1 can make partial concessions, and then the case where both players can. In addition, we compare the results under two assumptions about timing: in one version, we take concessions to be one-shot, i.e., there is a fixed time when a concession must be made, or else the chance to do so is lost forever; in the other, multiple concessions can be made whenever the player wants, subject only to the restriction that concessions cannot be taken back.

One-Sided Concessions

The Limit Case

For simplicity and to fix ideas, we first discuss the limit case where \( \nu \approx 0 \). First, assume a one-shot concession at the beginning of the game: player 1 chooses a concession size \( x \) at \( t = 0 \). Thereafter, the game continues as usual, except that she gets \( H - \alpha x \) from winning compared to losing, while 2 gets \( H - x \) from winning. Here \( \alpha \) represents the relative efficiency of the concession technology: if \( \alpha < 1 \), a concession reduces 1’s incentive to win the war relatively less than 2’s incentive, and vice versa.

Note that \( \alpha < 1 \ (\alpha > 1) \) does not necessarily imply that the concession generates (destroys) social welfare, as payoffs may have been normalized differently for each player. For instance, in the siege example, the benefits from winning before the concession would be higher for the defender than the attacker, i.e., \( H_D > H_A \), since a losing defender suffers the cost of being plundered which is not collected by the attacker; and there would be some
flow costs $c_D$, $c_A$. A transfer $x$ from $D$ to $A$ would change D’s payoffs to $H_D - x$ for winning or 0 from losing, while A’s would change to $H_A$ or $x$, respectively, generating no net change in welfare. However, if at the outset we had normalized the problem so that benefits were equal for both sides, then we might have normalized payoffs $H = \tilde{H}_D = \alpha H_D = \tilde{H}_A = H_A$; $\tilde{c}_D = \alpha c_D$; $\tilde{c}_A = c_A$; and a transfer of $x$ would change the benefits of winning to $H - \alpha x$ for $D$ and $H - x$ for $A$, where $\alpha = \frac{H_A}{H_D}$.

**Proposition 7** (One-Shot Concession with Slow-Moving $\theta$). The equilibrium of the game is as follows:

- If $c_1 < c_2$, player 1 makes no concessions and wins immediately.

- If $c_1 \geq c_2$ and $\alpha < 1$, player 1 makes a concession $x^*$ defined by:

\[
\frac{H - \alpha x^*}{c_1} = \frac{H - x^*}{c_2}
\]

and wins immediately. This leads to a concession $x^* = H \frac{c_1 - c_2}{c_1 - \alpha c_2}$, and payoffs $U_1 = H - \alpha H \frac{c_1 - c_2}{c_1 - \alpha c_2}$, $U_2 = H \frac{c_1 - c_2}{c_1 - \alpha c_2}$.

- If $\alpha \geq 1$, concessions never benefit player 1. Hence, if $c_1 = c_2$, no concession is made ($x^* = 0$) and the equilibrium described in Proposition 4 is played. If $c_1 \geq c_2$, any $x^* \in [0, H]$ is compatible with equilibrium but player 1 loses immediately in all cases.

In other words, when $\theta$ moves very slowly, partial concessions can be useful— but only when they strengthen the conceding player’s relative cost-benefit ratio, i.e., when $\alpha < 1$. In this case they are very powerful, and indeed guarantee that player 1 always wins the game in equilibrium, even when $c_1$ is substantially higher than $c_2$. Of course, this does not necessarily mean that player 1 walks away with a large payoff, as the concession needed to win may be large. In particular, the required concession $H \frac{c_1 - c_2}{c_1 - \alpha c_2}$ is small when $\alpha << 1$ (so that player 1 only needs to take a small dent in her own payoff to substantially reduce 2’s incentive to win), or when $c_1$ and $c_2$ are close (so 1 is not very disadvantaged to begin with, so a small nudge is enough to make 2 surrender). Conversely, as $\alpha \to 1$, if $c_1 > c_2$ the required concession converges to $H$.

On the other hand, when $\alpha \geq 1$—that is, when a concession weakens 1’s relative cost-benefit ratio—making a concession can never help player 1. This seems like a natural result, although we will see that matters are subtler away from the limit case.

Finally, note that in the limit case the timing of concessions actually does not matter: although we have stated the above result in a game where 1 can only make a concession at
\[ t = 0, \] the equilibrium is trivially the same if \( \alpha \) is instead allowed to make multiple concessions at any time. Indeed, if \( \alpha < 1 \) or \( c_1 < c_2 \), so that \( 1 \) is able to win in equilibrium, she might as well make the winning concession immediately, so as to avoid any cost of waiting.

### 7.0.1 General Processes

We limit ourselves here to the case of one-shot concessions:

**Proposition 8** (Equilibrium with a One-Shot Concession). The equilibrium of the game is as follows:

- If \( \theta < \theta^* \), player 1 makes no concession and wins immediately.
- If \( \theta \geq \theta^* \), let \((\theta_*(x), \theta^*(x))\) be the thresholds resulting in the continuation game after a concession of size \( x \) is made. Let \( x_1 \) be the smallest concession for which \( \theta_*(x) \geq \theta \), i.e., player 1 wins immediately in the continuation (we take \( x_1 = \infty \) if no \( x \) satisfies this requirement). Let \( x_2 \in [0, x_1] \) be a minimal concession that maximizes 1’s surrender threshold, i.e.,

\[
x_2 = \min \left( \arg\max_{x \in [0, x_1]} \theta^*(x) \right).
\]

Then \( x_2 \) is the optimal concession. Moreover, \( \theta^*(x) \) is quasiconvex, so either \( x_2 = 0 \) or \( x_2 = x_1 \). In the first case, no concession is made; in the second, player 1 wins immediately.

In particular, keeping all other parameters fixed, there is \( \alpha^* < 1 \) such that for all \( \alpha \geq \alpha^* \) it is optimal to choose \( x^* = 0 \), while for \( \alpha < \alpha^* \) the optimal choice is \( x^* = x_2 \). In addition, player 1’s equilibrium utility is decreasing in \( \alpha \).

The intuition behind this result is as follows. By Proposition 3, as long as \( \theta \) is far from the surrender thresholds, a concession which lowers 1’s potential payoff from winning—while also changing the surrender thresholds—actually has no impact on her expected utility except by changing said thresholds. This follows immediately from the fact that \( H \) does not feature in the expression for expected utility. Moreover, this expression is clearly increasing in \( \theta^* \), whence 1 effectively wants to maximize her own surrender threshold. Two surprising implications follow.

First, the optimal concession may not be one that results in an immediate win for player 1: indeed, she may choose to make a positive concession which still leaves the outcome of the war in doubt, and produces delay.

Second, whenever \( \alpha > 1 \) concessions always hurt player 1, but this is also true for some values of \( \alpha \) that are below 1. The intuition behind this result is that there are two forces which
affect the surrender thresholds when \( x \) changes, as anticipated in Proposition 2. On the one hand, a higher \( x \) reduces the stakes of winning for both players; this reduces their willingness to accept delay, so it tends to bring both thresholds closer to each other while having no clear impact on their average. On the other hand, if \( \alpha < 1 \) (\( > 1 \)) the concession tends to increase (decrease) both thresholds and as it strengthens (weakens) player 1’s relative strength. For a positive concession to be optimal, there must be a range of concession values for which the latter effect dominates the former, which requires \( \alpha \) to be below 1 by a certain amount. Note that, in particular, this means that player 1 may decline to make a concession even when this might guarantee a win for a small cost: for example, if \( \alpha = 1 \) and \( \theta \) is initially close to \( \theta^*(0) \), a small concession might be enough to bring \( \theta^*(x) \) up to \( \theta \), clinching the win; but, since this would also decrease \( \theta^* \), it follows that the cost of the concession is higher than the gains in terms of reduced delay and increased probability of winning.

Two-Sided Concessions

We now move on to a further extension where both players are able to make concessions. This can be thought of as a form of bargaining by means of unilateral offers. To fix ideas, consider the following examples:

- Two players, a buyer and a seller, engage in a war of attrition over the price at which \( B \) might buy a good worth \( H \). As the war continues, \( B \) can make statements of the form: “I will buy for any price up to \( x \)”, while \( S \) can say, “I will sell for any prize above \( y \)”. These promises are binding, so \( B \)’s acceptable price \( x_t \) cannot decrease over time and vice versa. Whenever a player surrenders, the transaction happens at the price set by the other player’s most recent statement.

- Two armies contest a one-dimensional territory \([0, 1]\), where \( v_i(t) \) is the value assigned to point \( t \) by player \( i \), and \( H_i = \int_0^1 v_i(t)dt \). Assume \( v_1 \) is decreasing and \( v_2 \) is increasing. While the war is ongoing, the armies station troops at outposts throughout the territory to contest its control and engage in intermittent fighting, generating flow costs. Along the way, each player \( i \) can make credible, irreversible commitments to pull her troops out of a measurable region \( S \subseteq [0, 1] \). When a player surrenders, she keeps only the territory that the other player conceded, and the other player takes the rest.

As a reasonable compromise between generality and simplicity, we allow for each player \( i \) to have a concession factor \( \alpha_i \), so that if initial rewards from winning are \( H_i, H_j \) and \( i \) makes a concession \( x \), rewards become \( H_i - \alpha_i x, H_j - x \). In the bargaining example, we would have \( \alpha_1 = \alpha_2 = 1 \), since concessions take the form of transfers. In the war for territory,
the $\alpha_i$ would vary depending on the exact territory conceded, but at the margin we would have $\alpha_1\alpha_2 < 1$: indeed, if $i$ is trying to reduce $j$’s incentives to fight relative to her own, the optimal strategy is to concede territory that $j$ values most and $i$ least, i.e., 1 would make concessions of the form $[t, 1]$ and 2 would make concessions of the form $[0, t']$, which result in the conceding player’s incentive to win going down less than the recipient’s.

We will focus on the limit as $\nu \to 0$. We normalize the costs and benefits so that $c = c_1 = c_2$ to streamline the notation. For clarity, we restrict the players to the following concession protocol: at the beginning of the game, 1 chooses a partial concession, then 2 observes it and makes a partial concession of her own; then the rest of the game proceeds as usual. (It can be shown that switching the order of concessions, adding more alternating concessions, making them simultaneous or allowing concessions after $t = 0$ have no impact on the equilibrium concessions and payoffs.)

**Proposition 9** (Two-Sided Concessions with Slow-Moving Processes). The equilibrium of the game is as follows. First, if $\alpha_1 \geq 1$, $\alpha_2 \geq 1$ then concessions are not used and the equilibrium is as in Proposition 4. Similarly, if $\alpha_1 < 1$, $\alpha_2 \geq 1$ then only player 1 may use concessions and the equilibrium is as in Proposition 7. If $\alpha_1 < 1$, $\alpha_2 < 1$ then:

- If $H_1 < \alpha_1 H_2$, player 2 wins immediately without making a concession and payoffs are $(U_1, U_2) = (0, H_2)$.

- If $H_2 < \alpha_2 H_1$, player 1 wins immediately without making a concession and payoffs are $(U_1, U_2) = (H_1, 0)$.

- Otherwise, both players make partial concessions $x_1 = \frac{H_2 - \alpha_2 H_1}{1 - \alpha_1 \alpha_2}, x_2 = \frac{H_1 - \alpha_1 H_2}{1 - \alpha_1 \alpha_2}$ which jointly exhaust the prize, and payoffs are

$$U_1 = \frac{H_1 - \alpha_1 H_2}{1 - \alpha_1 \alpha_2}, \quad U_2 = \frac{H_2 - \alpha_2 H_1}{1 - \alpha_1 \alpha_2}.$$  

The intuition behind the result is as follows. As before, a player can try to use a concession to improve her relative strength and win the war. However, player 1 understands that making a concession just large enough so that $H_1 - \alpha_1 x_1 \geq H_2 - x_1$, as she would do in Proposition 7, is not enough here because player 2 will undercut her with her own concession. Hence, in equilibrium player 1 concedes enough so that player 2 cannot win except by conceding the entirety of the remaining prize. The resulting payoffs turn out to be the same as when player 2 concedes first and follows the same logic.

---

10 Technically there is a collection of equilibria giving the same outcome, as player 1 can make concessions that have no impact on payoffs.
8 Conclusions

We have analyzed a model of wars of attrition with an evolving state of the world. As intended, the perturbation used in the model yields several attractive properties which are absent from the baseline model. In particular, the game has a unique equilibrium, yielding predictions about the likely winner which reflect a natural intuition about how this game might play out in practice. Moreover, the solution exhibits well-behaved comparative statics, meaning that if a player’s benefit from winning increases or her flow cost decreases, she will be more likely to win. In the same vein, such a change will lead to the war ending sooner if the player being strengthened held an advantage to begin with, while the war will lengthen if she was an underdog at first. These properties are shared by models based on reputational perturbations, but unlike those, the present model is more robust to assumptions about how the game looks away from the perturbation, an attractive property for applied researchers who may not be able to observe enough detail about a real example to guarantee that they are modeling perturbations correctly.

In addition, the present model is highly tractable, which allows it to be used as a modeling tool in other settings or in applied problems. This is illustrated in Section 7, which shows how to account for the possibility of partial concessions during a war of attrition, and how access to this option might benefit the player making concessions. But many other applications are possible: for instance, the game can be extended to explore the use of costly commitment devices, i.e., “burning bridges”; to wars of attrition involving more than two players, as in legislative standoffs; or to cases where players have some control over the flow costs, such as price wars.

Finally, the model makes a novel prediction regarding the most natural equilibrium in balanced wars of attrition (where both players have similar chances of winning), which entails higher payoffs than a totally mixed equilibrium, as well as a different distribution for the expected length of the war. This raises the question whether this equilibrium is indeed played in real-life scenarios, or under what circumstances (e.g., does there have to be a changing, publicly-observed variable which is focal for the players?), a question which might be answered by laboratory or field experiments.
A Appendix

Definition 1. A history at time $t$, $h_t$, is given by a sequence of states of the world $(\theta)^t = (\theta_0, \ldots, \theta_t)$, and a sequence of actions $(a_{is})$ for $i = 1, 2$ and $s = 0, \ldots, t$. A history $h_t$ is non-terminal if $a_{is} = 1$ for all $i = 1, 2$ and $s \leq t$.

Henceforth we will say histories to refer to non-terminal histories for brevity, and write $h_t = (\theta)^t$ as shorthand for $h_t = ((\theta)^t, (1, \ldots, 1), (1, \ldots, 1))$.

Definition 2. A strategy $\sigma_i(h)$ for player $i$ is Markov if $\sigma_i(h) = \sigma_i(h')$ for all $h = (\theta)^t$, $h' = (\theta)^t$ such that $\theta_t(h) = \theta_t(h')$.

A strategy $\sigma_1$ for player 1 (2) is a threshold strategy with threshold $\theta^*$ if $\sigma_1(h_t) = 1$ whenever $\theta_t(h_t) < \theta^*$ ($>$) and $\sigma_1(h_t) = 0$ whenever $\theta_t(h_t) > \theta^*$ ($<$).

Lemma 1. In any SPE, player 1 never surrenders at time $t$ if $\theta_t < -M_1$, and player 2 never surrenders at time $t$ if $\theta_t > M_2$.

Proof. For player 1, surrendering when $\theta_t < -M_1$ yields a continuation payoff of 0, while continuing until the first time $s > t$ when $\theta_t \geq -M_1$ yields a strictly positive payoff. The proof for player 2 is identical. $\square$

Definition 3. For any history $h$ and strategies $\sigma_i, \sigma_j$, let $U_i(\sigma_i, \sigma_j|h)$ be $i$’s continuation utility from strategy profile $(\sigma_i, \sigma_j)$ at history $h$.

Definition 4. For any history $h_0$, any strategy $\sigma_i$ for player $i$, and any $\alpha \in [0, 1]$, let $\sigma_i^{h_0, \alpha}$ be a strategy for player $i$ given by $\sigma_i^{h_0, \alpha}(h) = \alpha$ if $h = h_0$ and $\sigma_i^{h_0, \alpha}(h) = \sigma_i(h)$ otherwise.

Definition 5. Given two strategies $\sigma_i(h), \sigma_i(h')$ for player $i$, we say $\sigma_i \geq \sigma'_i$ iff $\sigma_i(h) \geq \sigma'_i(h)$ for all histories $h$.

Lemma 2. Let $\sigma_j, \sigma'_j$ be two strategies for player $j$ such that $\sigma_j \geq \sigma'_j$. Let $\sigma_i$ be a strategy for player $i$ and let $h$ be a history. Then

$$U_i(\sigma_i^{h,1}, \sigma_j|h) - U_i(\sigma_i^{h,0}, \sigma_j|h) \leq U_i(\sigma_i^{h,1}, \sigma'_j|h) - U_i(\sigma_i^{h,0}, \sigma'_j|h).$$

Proof. Assume $h$ is a history for time $t_0$. Since $U_i(\sigma_i^{h,0}, \sigma_j|h) = U_i(\sigma_i^{h,0}, \sigma'_j|h) = 0$, we simply have to show that

$$U_i(\sigma_i^{h,1}, \sigma_j|h) \leq U_i(\sigma_i^{h,1}, \sigma'_j|h).$$

Indeed

$$U_i(\sigma_i^{h,1}, \sigma_j|h) - U_i(\sigma_i^{h,1}, \sigma'_j|h) =$$

$$\sum_{t=t_0}^{\infty} \delta^{t-t_0} \int Q_{t_0,t}(\sigma_i^{h,1}, \sigma'_j, (\theta)^t)(\sigma_j((\theta)^t) - \sigma'_j((\theta)^t)) \left( U_i(\sigma_i^{h,1}, \sigma'_j((\theta)^t)|((\theta)^t) - H_1 \right) dP((\theta)^t|((\theta)^{t_0}) \leq 0$$
where \( Q_t(\sigma_i^{h1}, \sigma_j', (\theta)^t) \) is the probability that the war continues up to time \( t \) conditional on the path of the state of the world being \((\theta)^t\) and the players using strategies \( \sigma_i^{h1}, \sigma_j' \) respectively. The last inequality follows from the fact that \( U_i\left(\sigma_i^{h1}, \sigma_j''(\theta)^t, 1\vert (\theta)^t\right) - H_1 < 0 \) by Assumption B5.

**Lemma 3.** Let \( \sigma_j, \sigma_j' \) be two strategies for player \( j \) such that \( \sigma_j \geq \sigma_j' \). Let \( \sigma_i \in BR_i(\sigma_j) \). Then there is \( \sigma_i' \in BR_i(\sigma_j') \) such that \( \sigma_i' \geq \sigma_i \).

**Proof.** In a discrete world, this should be an extension of the previous lemma plus a single deviation principle style argument.

Obs: maximum attainable utility is decreasing in the opponent’s strategy (because if the opponent lowers their strategy your utility at every history goes up weakly even without changing strategies).

How about:

1. Take the original strategy \( \sigma_j \) for player \( j \), and the best response \( \sigma_i \). There are three possible types of histories: a. Histories where the payoff of continuing and then following the optimal strategy is positive. In these histories \( h \), it must be that \( \sigma_i(h) = 1 \) and \( U(\sigma_i, \sigma_j|h) > 0 \). b. Histories where the payoff of continuing and then acting optimally is negative. In these histories \( h \) it must be that \( \sigma_i(h) = 0 \) and \( U(\sigma_i, \sigma_j|h) = 0 \). c. Histories where the payoff of continuing and then acting optimally is 0. In these histories \( h \), it must be that \( \sigma_i(h) \in [0,1] \) and \( U(\sigma_i, \sigma_j|h) = 0 \).

Now note that, by the previous claim (does this need a proof?), \( U(\sigma_i', \sigma_j'|h) \geq U(\sigma_i, \sigma_j'|h) \geq U(\sigma_i, \sigma_j|h) \) for all \( h \). It follows immediately that, in any histories of type a, \( \sigma_i'(h) = 1 \), and in any histories of type b, the payoff of continuing and then acting optimally is at least 0.

However, this does not guarantee that \( \sigma_i' \geq \sigma_i \) (even "a.e.") because it could be that, e.g., histories of type b are of positive measure and the difference between \( \sigma_j \) and \( \sigma_j' \) is "of measure zero", in which case \( \sigma_i' \) could just be a "more passive" (but still optimal) element from the common set of best responses than \( \sigma_i \).

So the next step is to argue that, if we take \( \sigma_i \) and then change the probability of continuing at any subset of b-histories, we get another best response (equivalently, expected utility at every history remains unchanged).

**Lemma 4.** There are \( M \in (-M, -M_1], \ M \in [M_2, M) \) such that, in any SPE, player 1 surrenders if \( \theta > M \) and 2 surrenders if \( \theta < M \).

**Proof.** Assume that \( \theta_t = M \). As usual, player 1 can guarantee a payoff of 0 by surrendering.

Suppose then that player 1 does not surrender immediately. Assume that player 2 plays a threshold strategy with threshold \( M_2 \).
There are two possible outcomes. Either player 1 surrenders at some time $t' > t$, or eventually $\theta_{t'} \leq M_2$ for some $t'$ and player 2 surrenders. In the first case, player 1’s continuation utility is strictly negative as she only pays flow costs and receives no prize. In the second case, player 1’s utility is

$$\delta^{t'-t} H_1 - \sum_{s=t}^{s=t'-1} \delta^{s-t} c_1(\theta_s).$$

Recall that, by assumption, $|\theta_{s+1} - \theta_s| \leq \eta$ for all $s$, so that $M - M_2 \leq |\theta_{t'} - \theta_t| \leq (t' - t)\eta$. Then

$$\sum_{s=t}^{t'-1} c_1(\theta_s) \geq (t' - t)c_1(M_2) \geq c_1(M_2) \frac{M - M_2}{\eta} > H_1,$$

whence

$$\delta^{t'-t} H_1 < \sum_{s=t}^{s=t'-1} \delta^{s-t} c_1(\theta_s) \leq \sum_{s=t}^{s=t'-1} \delta^{s-t} c_1(\theta_s).$$

Thus player 1 would strictly prefer to surrender if $\theta_t = M$. By continuity, the claim holds for all $\theta$ close enough to $M$. Note that, by Lemma 2, player 1 would also prefer to surrender if player 2 used any other strategy that does not violate Lemma 1.

The argument for player 2 is analogous. \hfill \qed

**Lemma 5.** Let $\theta_i^* \in [M_1, M_2]$. If player $i$ uses a threshold strategy with threshold $\theta_i^*$, player $j$ has an essentially unique best response, which is also a threshold strategy.

**Proof.** WLOG, assume player 2 uses a threshold strategy with threshold $\theta_*$, which we will denote by $\sigma_2^{\theta_*}$.

Let $V_i(\theta)$ be the highest continuation utility player 1 can attain conditional on the current state being $\theta$ and player 2 using strategy $\sigma_2^{\theta_*}$, i.e.,

$$V_i(\theta) = \sup_{\sigma_1} U_i(\sigma_1, \sigma_2^{\theta_*}|\theta).$$

(Note that $V_1$ clearly only depends on the current state and not on the history of states of the world, since player 2 is not conditioning on the history.)

Next, we prove several properties of $V_1(\theta)$ by a recursive argument.

**Claim 1.** $V_1(\theta)$ is weakly decreasing in $\theta$.

**Proof.** Let $V_{10}(\theta)$ be given by $V_{10}(\theta) = H_1$ if $\theta \leq \theta_*$ and $V_{10}(\theta) = 0$ otherwise. Define the operator $W : \mathbb{R}^{[-M,M]} \to \mathbb{R}^{[-M,M]}$ by

$$W(f)(\theta) = \begin{cases} H_1 & \text{if } \theta \leq \theta_* \\ \max (-c_1(\theta) + \delta E(f(\theta')|\theta), 0) & \text{if } \theta > \theta_* \end{cases} \quad (1)$$
where \( \theta' - \theta |\theta \sim H_\theta \). For each \( k \in \mathbb{N} \), define \( V_{1k} = W(V_1(k-1)) \).

We will now make several observations about \( W \). First, \( V_1 \) is a fixed point of \( W \). Second, \( W \) has at most one fixed point by the contraction mapping theorem (in fact, \( W \) is Lipschitz with constant \( \delta < 1 \) if we endow the space \( \mathbb{R}^{[-M,M]} \) with the norm \( ||\cdot||_\infty \)). \( [[[\text{add cite for CMT}]]] \)

Third, \( W \) is weakly increasing (i.e., if \( f \geq g \) everywhere, \( W(f) \geq W(g) \) everywhere) and \( V_{11} \geq V_{10} \). Hence, for each \( \theta \), the sequence \( (V_{1k}(\theta))_k \) is weakly increasing in \( k \). Since it is also bounded, it converges pointwise, and the pointwise limit is a fixed point of \( W \) (i.e., \( V_1 \)) by the monotone convergence theorem.

Fourth, \( W \) preserves decreasing-ness, i.e., if \( f \) is weakly decreasing in \( \theta \) then so is \( W(f) \). This follows from Assumptions A2, B1 and B5. Moreover, \( V_{10} \) is weakly decreasing in \( \theta \) by construction. Hence \( V_{1k} \) is weakly decreasing in \( \theta \) for all \( k \), and so is \( V_1 \). \( \Box \)

**Claim 2.** \(-c_1(\theta) + \delta E(V_1(\theta')|\theta)\) is strictly decreasing in \( \theta \).

**Proof.** This follows from the facts that \( V_1(\theta') \) is weakly decreasing in \( \theta' \); \( \theta' \) is FOSD-increasing in \( \theta \); and \( c_1(\theta) \) is strictly increasing in \( \theta \). \( \Box \)

\( [[[\text{add measurability to strategies}]]] [[[\text{check if proofs work at } M - \eta]]] \)

**Claim 3.** \( V_1(\theta) \) is continuous for \( \theta \in (\theta_*, M] \).

**Proof.** Note that \(-c_1(\theta) + \delta E(V_1(\theta')|\theta)\) is continuous in \( \theta \) because: \( c_1(\theta) \) is continuous by Assumption B2; \( V_1 \) is bounded, as \( V_1 \in \theta' \) \( \Box \)

Next, note that , as \( \delta E(V_1(\theta')|\theta) \) is weakly decreasing in \( \theta \) by the previous arguments and \(-c_1(\theta) \) is strictly decreasing by Assumption B1.

Let \( \theta^* = \inf \{ \theta : V_1(\theta) = 0 \} \). It follows that, for \( \theta \in [\theta_*, \theta^*) \), \( V_1(\theta) > 0 \) and \( V_1 \) is strictly decreasing in \( \theta \). For \( \theta > \theta^* \), \( V_1(\theta) = 0 \) but the payoff from continuing, \(-c_1(\theta) + \delta E(V_1(\theta')|\theta)\), is strictly negative, so player 1 has a strict incentive to surrender. Thus the essentially unique\(^{11}\) best response for player 1 is a threshold strategy with threshold \( \theta^* \). and \( \Box \)

**Proof of Proposition 1.** First, we prove that there is a unique equilibrium in threshold strategies. By Lemma 5, if one player is using a threshold strategy, the other player’s best response is also a threshold strategy, and the threshold is unique. Define then two functions \( T_1, T_2 : [-M, M] \rightarrow [-M, M] \) as follows: if player \( i \) uses threshold \( \theta_i \), then player \( j \)’s optimal threshold is \( T_1(\theta_i) \). An equilibrium in threshold strategies is then given by a threshold \( \theta^* \) for player 1 such that \( T_1(T_2(\theta^*)) = \theta^* \).

\(^{11}\)It is not unique in the sense that any value \( \sigma_1(\theta^*) \in [0, 1] \) is optimal.
We will now show that $T_1$ is increasing but $|T_1(x) - T_1(y)| < |x - y|$ for any $x, y$. Let $V_1^\tilde{\theta}(\theta)$ and $V_{1k}^\tilde{\theta}(\theta)$ for all $k$ be as defined in Lemma 5, conditional on player 2 using threshold $\tilde{\theta}$.

Note that, given any two values $\tilde{\theta} > \tilde{\theta}'$, $V_{10}^\tilde{\theta} \geq V_{10}^{\tilde{\theta}'}$. Moreover, $W^\tilde{\theta}(f) \geq W'^{\tilde{\theta}}(f)$ for any function $f$, and both operators are weakly increasing. Hence $V_{1k}^\tilde{\theta} \geq V_{1k}^{\tilde{\theta}'}$ for all $k$, so $V_1^\tilde{\theta} \geq V_1^{\tilde{\theta}'}$ and $T_1(\tilde{\theta}) \geq T_1(\tilde{\theta}')$, i.e., $T_1$ is weakly increasing.

Next, let $t_\Delta$ be the function $t_\Delta(\theta) = \theta - \Delta$. Take $\Delta = \tilde{\theta} - \tilde{\theta}'$. For any function $V$, denote $V = V \circ t_\Delta$. For any operator $W$, define $\overline{W}$ by $\overline{W}(f) = W(f \circ t_\Delta^{-1}) \circ t_\Delta$.

By construction, $\overline{V}^\tilde{\theta}_{1k} = \overline{V}^{\tilde{\theta}'}_{1(k-1)}$. In addition $\overline{V}^\tilde{\theta}_{10} = V^\tilde{\theta}_{10}$.

The crucial observation now is that, for any weakly decreasing function $f$, $\overline{V}^{\tilde{\theta}}(f) \geq W^{\tilde{\theta}}(f)$. Indeed,

$$\overline{V}^{\tilde{\theta}}(f)(\theta) = \begin{cases} H_1 & \text{if } \theta - \Delta \leq \tilde{\theta}' \Leftrightarrow \theta \leq \tilde{\theta} \\ \max(-c_1(\theta - \Delta) + \delta E(f(\theta_{t+1} + \Delta)|\theta_t = \theta - \Delta), 0) & \text{if } \theta > \tilde{\theta} \end{cases}$$

Note that $-c_1(\theta - \Delta) > -c_1(\theta)$ by Assumption B1, and $\theta_{t+1} + \Delta = (\theta - \Delta) + X + \Delta$ where $X$ has distribution function $H_{\theta - \Delta}$, which is weakly FOSD’d by $H_\theta$ by Assumption A2.

It follows that $\overline{V}^{\tilde{\theta}'}_{1k} \geq V^\tilde{\theta}_{1k}$ for all $k$, and hence $\overline{V}^\tilde{\theta}_{1} \geq V^\tilde{\theta}_{1}$.

Finally, from Lemma 5, we know that $-c_1(\theta) + \delta E(V^\tilde{\theta}(\theta')|\theta) = 0$ for $\theta = T_1(\tilde{\theta})$. The above argument implies that $-c_1(\theta - \Delta) + \delta E(f(\theta_{t+1} + \Delta)|\theta_t = \theta - \Delta) > 0$ for the same $\theta$, whence $T_1(\tilde{\theta}') + \Delta > T_1(\tilde{\theta})$.

The same results are true of $T_2$, and this all implies that $T_1 \circ T_2$ has at most one fixed point. Indeed, if $\theta^* \neq \theta'^*$ were both fixed points, we would have that $|T_1(T_2(\theta^*)) - T_1(T_2(\theta'^*))| < |T_2(\theta^*) - T_2(\theta'^*)| < |\theta^* - \theta'^*|$, a contradiction.

We can show $T_1 \circ T_2$ has a fixed point as follows. Take $\overline{\theta}^*_0 = M$ and $\overline{\theta}^*_n = T_1(T_2(\overline{\theta}^*_{n-1}))$. Clearly $\overline{\theta}^*_0 \geq \overline{\theta}^*_1$. Since $T_1, T_2$ are weakly increasing, it follows that $\overline{\theta}^*_1 \geq \overline{\theta}^*_2 \geq \ldots$. Since the sequence is bounded it must converge to a limit $\overline{\theta}^*$ which is a fixed point of $T_1 \circ T_2$.

Finally we rule out other equilibria that are not in threshold strategies. We use a standard argument from supermodular games similar to Milgrom and Roberts [4]. Following the notation of the previous paragraph, denote $\overline{\theta}_s = T_2(\overline{\theta}'_s)$. Also, let $\overline{\theta}_{s0} = -M$, $\overline{\theta}_{sn} = T_2(T_1(\overline{\theta}_{s(n-1)})$, $\overline{\theta}_{sn} = T_1(\overline{\theta}_{sn})$ and denote the limits by $\overline{\theta}_s, \overline{\theta}'_s$ respectively.

Since $T_1, T_2$ are weakly increasing, we have

$$\overline{\theta}_0 \geq \ldots \geq \overline{\theta}_s \geq \overline{\theta}'_s \geq \ldots \geq \overline{\theta}_0,$$

$$\overline{\theta}_{s0} \geq \ldots \geq \overline{\theta}_s \geq \overline{\theta}'_s \geq \ldots \geq \overline{\theta}_{s0}$$
By Lemma 2, whenever \( i \) plays a strategy higher than \( \sigma_i \), any best response by \( j \) must be weakly lower than \( j \)'s best response to \( \sigma_i \). Hence, any strategy played by 1 must be weakly higher than \( \sigma_1^{\theta_0^*} \) and weakly lower than \( \sigma_1^{\bar{\theta}_0^*} \); any strategy played by 2 must be weakly higher than \( \sigma_2^{\theta_0} \) and weakly lower than \( \sigma_2^{\bar{\theta}_0} \); and so on.

By induction, any strategy used by 1 must be between \( \sigma_1^{\theta^*} \) and \( \sigma_1^{\bar{\theta}^*} \). But since \( \theta^* = \bar{\theta}^* = \theta^* \), there is a (essentially) unique equilibrium strategy for player 1, and analogously for 2.

[[[add payoff assumption to guarantee one threshold strictly lower than the other]]] [[[[prove continuity]]]]
References


