The Power to Hurt:

Costly Conflict with Completely Informed States

Branislav L. Slantchev
Department of Political Science
University of Rochester

February 16, 2002

Overview

- Why do wars occur? Why don't they end soon?
- Wars are costly, risky, terribly inefficient.
- Why fight instead of negotiating?
- War as bargaining failure (Fearon 1995):
  - private information, incentives to misrepresent
  - commitment problems
  - issue indivisibilities
- But are these models of conflict intuitive?

- One of central puzzles in IR is why wars occur

- Many varieties of explanations: structural, normative, domestic-politics, regime type, bureaucratic, bounded rationality, state centric, rationalist

- All try to answer why inefficient behavior occurs

- Inefficient: since fighting costly, there should be bargain ex ante accepted instead of fighting and then settling

- One of most influential: rationalist (summarized by Fearon)

- Assumptions in his model: (i) states as unitary actors, (ii) war as lottery over exogenous outcomes, (iii) war is a game-ending move

- Assumptions (ii) and (iii) relaxed here

War as a Bargaining Process

- "War. . . always lasts long enough for influence to be exerted on the goal" — Clausewitz

- "War is a bargaining process—dirty, extortionate, and often quite reluctant" — Schelling

- But. . . prevalent models (traditional and formal)
  - treat war as a costly lottery over exogenous outcomes
  - explain conflict with private information about *costs*
  - end with the outbreak of war!

Problems with lottery over outcomes assumption:

- do not allow for outcomes to be conditional on how well states are doing in the war (perhaps except for WWII and its unconditional surrender, this is unrealistic)

- cannot distinguish between ability to bear costs and capacity to inflict them: must be lumped (implicitly) together in the war outcome (once disaggregated, we can make this distinction—to my knowledge no theory in IR does that); more on that in discussion

- I show that this difference is very important because it determines the types of negotiated outcomes available—in fact, I call the combined effect of being able to do well on both dimensions, *the power to hurt*

Problems with relying on incomplete information exclusively:

“if both sides knew how the pie would be divided after the war, both would be better off if they divided accordingly before the war” (Goemans, Fearon) Consequently, virtually all models argue that uncertainty about this division is necessary to produce war

- merely knowing how a pie would be divided does not mean it is the only way of dividing it in equilibrium — there exists a range of possible divisions, not all of them good for both players

- because there exists really bad deals obtainable in equilibrium, players might want to fight to avoid them if they believe that relenting sooner would convince the enemy of their weakness

- result can occur with complete information!

Problems with war as game-ending move:

- game-theory tells us that if a game is played repeatedly, outcomes that are not equilibria in the game may become equilibria (think of cooperation in repeated Prisoner's Dilemma)

- bargaining situation is not really a repeated game (since agreement ends it)

- similarly, war is not a repeated game since a negotiated outcome or victory ends it

- still, if we model the two situations as processes, then outcomes not available in the game-ending models might arise (I've already argued that these *are* processes)

Summary of Argument

1. treat war as a *process*, not *outcome*

2. allow outcomes to be endogenous

3. examine strategic incentives for inefficient behavior in absence of uncertainty

Some historical illustrations:

- France vs. Germany in 1916 and 1940
- NATO air strikes in Kosovo in 1999
- Japanese surrender in 1945
- Napoleon's Continental System

Therefore, today we shall look at a model that:

- Treats war as a bargaining process
- Allows demands to be conditional on fighting

In order to examine:

- incentives to fight inefficient wars
- effect and types of war costs
- the argument that uncertainty is necessary

Sketch of Central Findings

Model completely stacked against war:

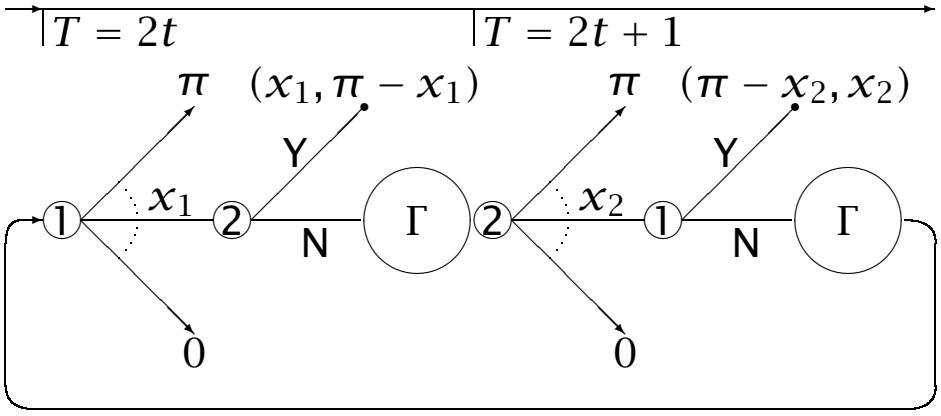
- peace can be supported in equilibrium
- peace is most preferred outcome
- fighting brings costs, no benefits
- complete information

Still we find:

- equilibria with inefficient fighting
- delay in reaching agreements
- costs of hurting and being hurt important

The Model

- Players: $i \in \{1, 2\}$
- Two-way division of benefits: $x_1 + x_2 = \pi$
- Game schematic (complete information):



- Payoffs: $(1 - \delta_i) \sum_{t=0}^{\infty} \delta_i^t z_{it}$ where

$$z_{it} = \begin{cases} r_{it} & \text{payoff in } \Gamma \text{ until agreement} \\ x_{iT} & \text{if agreement at time } T \end{cases}$$

An Example of the Conflict Game Γ

		Player 2	
		F	*NF*
Player 1	*Fight*	3 1	4 3
	Not Fight	5 2	6 4

General Assumptions

- Peace is a subgame perfect equilibrium $(s_1 + s_2 = \pi)$
  $(4 + 6 = 10)$

- Fighting brings costs only, no benefits $(r_i < s_i)$

- If any player fights, both suffer

Thus, model completely stacked against fighting:

- model wouldn't be interesting if it said that states fought when they found war profitable

- war is not profitable here by assumption

Peaceful and Efficient Equilibria

Proposition 4.1. *For each subgame perfect equilibrium of* Γ*, the bargaining game has a unique subgame perfect equilibrium of a simple form, in which agreement is reached immediately and no fighting occurs.*

Corollary 4.2. *Peace can be supported in an equilibrium of the bargaining game.*

Note that even if fighting is a SPE of Γ, Prop. 4.1 still tells us that no fighting would actually occur in equilibrium — the threat of it determines the offer immediately accepted and players never get to fight.

Peaceful Equilibria with "Bad" Outcomes

Proposition 4.7. *When players care sufficiently about the future, the bargaining game has subgame perfect equilibria, in which one player obtains payoffs strictly smaller than the status quo payoffs from Corollary 4.2. (holds for both players)*

Still peaceful, still (almost) efficient. . .

These results not surprising given my determination to make peace a viable option.

Prop. 4.7 tells us that there is an equilibrium that gives player 1 a really bad payoff, and there is another, which gives player 2 a really bad payoff

These equilibria are still efficient: agreement reached immediately or after at least one-period delay during which no fighting occurs (the latter is a technical result due to possible differences in players' patience)

Thus far. . .

Some results encouraging for peace:

- many equilibria but all peaceful
- agreement reached immediately (or almost so)
- revision of benefits without fighting

Equilibria supported by the threat to fight in Γ:

- cost of punishing the other player
- cost when other is punishing

But. . . *these are not the only equilibria!*

Inefficient Equilibria

Proposition 5.2. *When players care sufficiently about the future, there exist equilibria with history-dependent strategies, in which players make nonserious offers, reject all proposals, and fight for some time before agreeing on a split.*

These equilibria look as follows:

- $\underline{s}$ – player 1's lowest payoff (resp., 2's highest)
- $\overline{s}$ – player 1's highest payoff (resp., 2's lowest)
- for any $s \in (\underline{s}, \overline{s})$ and some time $T > 0$, there are inefficient equilibria with fighting at $t < T$ and settlement on $(s, \pi - s)$ at T
- if player deviates, switch to equilibrium with its lowest payoff (from Prop. 4.7)

- Many agreements in which *both* players better off (for example, an agreement on s prior to T), but players cannot obtain them

- Result is under complete information

- Result does not disappear with quick alternation of proposals

- Result persists in extended models that allow renegotiation of agreements

What is the Intuition Here?

- Recall equilibria with low payoffs (Prop. 4.7)
- Player who quits “too early” gets worst payoff
- Think about situation where trying to bargain “too early” signals a low-resolve (high-cost) type; interpreted as weakness

Emphasize:

- the model makes no unique prediction that ISPE will occur, it is a possibility result
- incentives to engage in behavior that is suboptimal in the short run in order to get better outcome in the long run

Costs of Fighting and Mutual Coercion

Two types of costs *(the power to hurt)* jointly determine bargaining range. Consider player 1's payoffs:

		Player 2 *f*	Player 2 *nf*
Player 1	*Fight*	3 1	4 3
	Not Fight	5 2	6 4

- 1's cost when 2 is inflicting pain on 1: *(NF,f)*

  (determines how much player 1 can give up)

- 1's cost when 1 is inflicting pain on 2: *(F,nf)*

  (determines how much player 1 can demand)

France defended Verdun in 1916 at cost of $400,000$ casualties, but in 1940 she gave up after suffering $90,000$ dead and $250,000$ wounded.

Crucial difference: in WW2, Germany destroyed organizational capacity of France and damaged its ability to inflict losses on the Germans (limits how much France can expect to get from Germany)

No sense to fight if there is no chance to obtain a better bargain

Still, because not everything was lost, France retained some residual capacity to fight, so its worst bargain was actually not 0 (i.e. she kept the fleet and precluded German use of it—a possibility that so bothered Britain that Churchill had the RAF bomb it to prevent it)

Some Implications

- States may fight as long as an early peace means bad settlement

- Diminished capacity to hurt is major reason to seek war termination

- Denial of power to hurt undermines bargaining position of opponent

- Military victory not necessary to end war

- Power to hurt can be military, economic, humanitarian, etc. but comprises ability to bear both types of costs

- "Paradoxical" outcomes when a militarily and economically stronger side loses can be explained

For example, consider NATO's air strikes in Kosovo. Some (Stam, 1999) argue that a ground invasion would have had less than 50% chance of success, which is why it was not attempted.

I find this doubtful. According to this analysis, even if such an operation would have had 100% chance of success, it would have been costly (mountainous terrain unlike desert in the Persian Gulf), giving some Milosevic opportunity to bargain.

Aerial strikes virtually eliminated his capacity to get a better bargain because there was no way to influence the alliance.

Maybe this is why Milosevic was brought to The Hague at the end of June this year, while Hussein is still in Baghdad.

Japan Surrenders in 1945

Rare instance of capitulation:

- Morale high despite incendiary/nuclear attacks
- Major islands still in Japanese hands
- Army that could defend the home territory

Why did Japan surrender?

Peace overtures began in May by Togo (FM) through Moscow (Stalin didn't even tell the US), Emperor states on July 12 his desire to end war. Army-Navy fearful of Soviet attack ever since Stalin repudiated the Neutrality Pact.

On June 6 (Suzuki is PM), "The Fundamental Policy to Be Followed Henceforth in the Conduct of War" calls for continuation of war on the homeland

Few believed that Japan could win (General Miyazaki said "beyond all expectation"), the point was to obtain a better settlement through operational success (making invasion terribly costly to the US)

Lt-Gen. Arisue: "inflict tremendous losses [and thus] bring about the termination of hostilities on comparatively favorable terms"

Soviets declare war on August 9, undermining the hope that invasion could be made sufficiently costly

Militarists, who had refused to convene the Supreme Council after Hiroshima, immediately met to discuss acceptance of Potsdam Declaration

After hearing the news of the debacle of the vaunted Kwantung Army in Manchuria (on which they had pinned their hopes), Suzuki remarked, “Is the KA that weak? Then the game is up.”

Japanese intransigence based on hope of extracting better terms, for which an ability to hurt the US was necessary. With Soviet involvement and collapse of KA, they realized hopes illusory, so they surrendered

Must emphasize:

- US firmly believed in success of invasion and did not believe it to be too costly, so it's not clear that Japanese plan would have succeeded even without the Soviet attack

- Standard claim that nuke was necessary undermined by timeline of events: militarists ignored first bombing and decision made prior to the second

Conclusions

- Misleading to treat war as game-ending costly lottery over exogenously fixed outcomes

- War can occur without uncertainty or credibility problems

- Power to hurt is major factor in war termination

- These results cannot be derived from traditional models

- Theory of war termination is a first step toward theory of war initiation

Both practitioners of war (Clausewitz, Fuller) and scholars (Schelling, Wagner) argue war is bargaining process. I give stylized model that incorporates this idea.

To examine incentives for inefficient behavior in the starkest setting possible, I stacked the model against war. Still, even under complete information, there exist incentives for such behavior.

Since results seem to make sense but cannot be derived from standard models, we should re-evaluate theoretical approaches that explicitly or implicitly treat war as a game-ending move and a lottery over exogenously fixed outcomes.

I made a distinction between two types of costs, jointly called *the power to hurt* and argued that a state that is weak on either one of these dimensions is unlikely to obtain a good bargain.

Note that it is possible to be strong on one dimension (e.g. one could pay to hurt the other, but cannot bear the cost of being hurt back)

Power to hurt not necessarily military: Napoleon could not get at Britain, so he tried to strangle it economically through Continental System.

Since power to hurt may not translate directly into factors commonly used to measure strength: military capabilities, geo-political configurations, economic resources, this analysis provides clues as to why asymmetric conflict can occur and end with the "stronger," in these conventional measures, party on the losing side.

More attention needs to be paid in statistical models that mechanically aggregate such factors.

A theory of war termination is a first step toward a theory of war initiation, not the other way round.

Bonus: The Other Model, 1

In Chap. 3, I explicitly model war as a probabilistic process that develops over time, and then examine negotiations superimposed on that process:

- With complete info, agreement immediate *but depends on how well states are doing on the battlefield* (war aims endogenous)

- When one side makes all the offers, it can extract all the surplus over and above what the other can guarantee itself by fighting

- When defeat is near, states may accept bargains that leave them *less than what they are getting while fighting* — because they cannot keep it if they lose completely

Bonus: The Other Model, 2

In a novel development, I model uncertainty not over costs, but over distribution of power (this matches nonformal theories much better)

Thus, states are uncertain about the probability of winning individual battles (and hence about how the war will end)

Here we find that pessimistic types conclude agreements earlier and settle for less and war is a type of screening process where battlefield outcomes reveal information about the unknown probability.

It takes time for this information to accrue, and if some war is "too short," it can end with the complete defeat of one side since the other would not have made an acceptable offer.

Bonus: Empirical Tests

Statistical analysis of hypotheses from models (Ch. 4) test three classes of dependent variables: duration of war, type of war outcome, type of political settlement

- Capacity to bear costs when being hurt: regime type (losses?)
- Capacity to inflict pain: type of war, terrain, economic resources, military capability
- Wars end when states "agree" on outcome: uncertainty results in longer wars
- Factors influencing uncertainty: third parties, parity of strength

Also, a more detailed case study of War of 1812 between Britain and US to illustrate the endogeneity of war aims.

Dissertation Summary

1. “Power to Hurt” — incentives to fight wars (Ch. 2)

2. “Stochastic Model of Wartime Negotiations” — how does interwar bargaining depend on battlefield performance? (Ch. 3)

3. Chapter 4 — statistical tests (hypotheses from Ch. 3, with insights from Ch. 2)

The Future

- Effect of public opinion (contra audience costs)
- Effect of domestic political institutions
- Effect of third-parties

A Numerical Example

Let $\delta_1 = .95$, $\delta_2 = .9$, and Γ be ($\pi = 10$):

		Player 2	
		F	*NF*
Player 1	*Fight*	3 1	4 3
	Not Fight	5 2	6 4

Then, from Proposition 4.7:

- Player 1's worst payoff is 3.379 (2's best is 6.621)
- Player 1's best payoff is 6 (2's worst is 4)

$\Rightarrow$ bargaining rage for player 1's share is $(3.379, 6)$

For this Γ, $\underline{v}_1 = 2, \underline{v}_2 = 4$, and $\overline{v}_1 = 4, \overline{v}_2 = 6$.

Player 1's worst payoff is (since $\delta_1 > \delta_2$) in Type I extremal equilibrium: $x_1^* = 2 + \frac{2(1-.9)}{1-(.9)(.95)} \approx 3.379$. Thus, player 2's best payoff is $10 - 3.379 = 6.621$

Player 1's best payoff is in Type II extremal equilibrium (nonserious offers by 2): $x_1^* = \frac{10-4+(.9)(6)}{1+.9} = 6$. Thus, player 2's worst payoff is $10 - 6 = 4$.

A Numerical Example, cont'd

Consider inefficient equilibrium: fight for 3 periods, then settle on $(4.15, 5.85)$

Following strategies from Prop. 5.2 yields payoffs:

P.1: $(1-\delta_1)\left[1+\delta_1(1)+\delta_1^2(1)+\delta_1^3\left(\frac{4.15}{1-\delta_1}\right)\right] = 3.701$

P.2: $(1-\delta_2)\left[3+\delta_2(3)+\delta_2^2(3)+\delta_2^3\left(\frac{5.85}{1-\delta_2}\right)\right] = 5.078$

Suppose player 1 proposes $10-3.701 = 6.299$ at $t=0$. By rejecting, 2 gets $(1-\delta_2)(4)+\delta_2(6.621) = 6.359$. Therefore, 2 prefers to reject. This leaves player 1 at most with $(1-\delta_1)(4)+\delta_1(3.379) = 3.410$. *Therefore, player 1 strictly prefers to fight.*

Suppose player 2 proposes $10-5.309 = 4.691$ at $t=1$. By rejecting, 1 gets at least $(1-\delta_1)(2)+\delta_1(6) = 5.8$. Therefore, 1 prefers to reject. This leaves player 2 at most with $(1-\delta_2)(6)+\delta_2(4) = 4.2$. *Therefore, player 2 strictly prefers to fight.*