The Korean Economic Review
Volume 29, Number 2, Winter 2013, 235-266.

Bargaining and War: A Review of Some Formal Models*

Sandeep Baliga** · Tomas Sjöström***

Would perfectly rational agents always negotiate peaceful outcomes at the bargaining table, or would they sometimes fight costly wars? The Coase theorem suggests that when rational agents negotiate freely, they will reach a surplus-maximizing outcome. This seems to rule out war, since war will not in general be surplus-maximizing. However, the Coase theorem is valid only under certain assumptions, such as transferable utility (no restrictions on side-payments) and complete information. Brito and Intriligator (1985) showed how incomplete information may lead to war. An aggressor who demands concessions may simply be bluffing, so it may be rational to refuse his demands. If the aggressor is not bluffing, a war may ensue. We discuss how long such a war may last, and whether other kinds of "frictions" (such as limited commitment power and limits on side-payments) may also lead to war.

JEL Classification: F51
Keywords: Bargaining, War

I. Introduction

Standard economic theory assumes the existence of a court system which enforces property rights and contracts. However, in this survey we will discuss a few models of anarchic environments where there is neither legal protection of property rights nor enforcement of contracts. In anarchic environments "might makes right": if someone is too weak to protect his possessions, others may take them from him forcibly. It may seem that this will lead to unpredictable chaos. But, in fact, economic reasoning can be applied even to such environments. An investment in military capability can be analyzed as any investment — a decision to consume less

Received: March 29, 2013. Revised: Dec. 1, 2013. Accepted: Dec. 4, 2013.
* We thank Basak Uysal for helpful comments.
** Professor, Kellogg School of Management, Northwestern University, Evanston, IL60208 USA, E-mail: baliga@kellogg.northwestern.edu
*** Corresponding Author, Professor, Department of Economics, Rutgers University, New Brunswick, NJ 08901 USA, E-mail: tsjostrom@econ.rutgers.edu

today in order to have a better outcome tomorrow. A decision to attack someone is a risky gamble that a rational person will take if and only if the expected benefits exceed the expected costs.

In an anarchic international system, *war* is one way of reallocating wealth from the weak to the strong. But war is a very inefficient method, because it destroys resources and creates human suffering. Why should war be used to reallocate wealth? Why doesn't the weaker country instead make a side-payment to the stronger country (pay a "tribute") in order to avoid a war? In fact, the *Coase theorem* states that rational agents who can bargain freely (without "transaction costs") and who can make unrestricted transfers ("side-payments") to each other, will negotiate a surplus-maximizing outcome. Since war is not surplus maximizing, if there are no transaction costs then wars should not occur.[1]

In an important paper, Brito and Intriligator (1985) asked (and answered) the question: if the leaders of all countries are perfectly rational, why would there ever be war? The answer they gave was that the leaders may have incomplete information. Incomplete information can be considered a transaction cost which prevents rational agents from negotiating a surplus-maximizing outcome (see, for example, Myerson and Satterthwaite, 1983). An aggressor who demands a tribute may simply be bluffing — he may be a weak type who will back down if the demand is rejected. Not knowing the aggressor's true type, it may be rational to reject his demand. If the aggressor is a tough type who is not bluffing, a war may ensue. To signal toughness, the aggressor may accumulate weapons. If this is a credible signal, then the tribute will be paid and there is no war. But now the issue of bluffing reappears, because a weak type (who would actually back down if rebuffed) can behave in exactly the same way.

To further develop this type reasoning requires a formal model. In Section 2, we present a version of Brito and Intriligator's model. There are two players who can be thought of as the leaders of two countries, *North* and *South*. The game has two stages. In stage one, each country allocates its productive capacity between military and civilian use, i.e., "guns or butter". This decision is publicly observed. Building military strength does not directly generate utility, but it has instrumental value. It is an investment which may pay off in stage two, when side-payments are made and war may be declared. A side-payment is simply a transfer of butter from one country to the other. If a war occurs, the country which is stronger militarily is more likely to win. Section 2.1 verifies that if the players have complete information, then there cannot be war in equilibrium (in agreement with the Coase theorem). The players will nevertheless build up military strength in stage one, because military strength

[1] This argument relies on utility being transferable, i.e., no restrictions on side-payments. In fact, if utility is not transferable then Pareto efficiency does not imply surplus maximization. War may then be a Pareto efficient outcome, even though it is not surplus-maximizing (see Section 3.3 for an example).

gives *bargaining power* and will therefore determine the stage two transfer.

Section 2.2 shows that with incomplete information, a war may occur in equilibrium. This obviously provides additional incentives to build weapons. In fact, a military build-up may be used as a signaling device. Specifically, we assume North has two possible types, weak and tough. The weak type would never start a war, but the tough type might start a war unless he is offered a sufficient side-payment. North's true type is his private information, i.e., South doesn't know North's type. If North is a *priori* likely to be weak, then there exists a semi-separating equilibrium: the tough type always accumulates weapons, the weak type accumulates weapons ("bluffs") with positive probability, the bluff is called with positive probability, and there is war with positive probability. This verifies Brito and Intriligator's insight: incomplete information may cause war.

In the simple model of Section 2 there is no commitment problem, because the game ends when a war is declared or a side-payment is made. In Section 3 the game is changed so that after a side-payment is made to North, North may return to ask for more. Of course, North might promise that he will not do so, but such a promise would not be credible if North lacks means of commitment. After the first transfer of resources, North is no weaker than before, and South is no stronger. If anything, the first transfer has made North more likely to win a war (North may convert the transfer into weapons) which should make him more aggressive.[2] Won't North then demand another, perhaps even bigger, transfer? With South anticipating this outcome, isn't it better for it to resist from the very beginning and refuse to make even the first side-payment? Might this not lead North to declare war? In Section 3, we use a version of Fearon's (1996) dynamic bargaining model to show that, under complete information, this commitment problem does not necessarily invalidate the Coase theorem. If the players are rational, they can work out the future consequences of today's actions. Based on these continuation values, if they have sufficient ability to transfer utility then there will exist a peaceful negotiated outcome that both prefer to war. This may involve a possibly very small transfer today, with the expectation that there will be more transfers in the future, making sure that at each point in time both prefer peace to war. Thus, with complete information and a sufficient ability to make side-payments, an inability to commit to future actions will not lead to war.

For lack of commitment power to lead to war under complete information,

[2] As Beviá and Corchón (2010) emphasize, transferring resources to North may in some situations serve to prevent an attack. After the transfer, North becomes more reluctant to go to war because his cost of losing is bigger (he would risk losing what he has already obtained), while his gain from winning is smaller (South has less left to take). In general, a transfer could make North more or less aggressive, depending on which effect dominates: on the one hand, the additional resources increases the probability that he will win a war, but on the other hand, it raises his cost of losing and reduces his gain from winning.

South's ability to make side-payments must be restricted. This occurs under what Fearon (1995) calls "issue indivisibility". As Acemoglu and Robinson (2001) have argued, issue indivisibility may arise if South's ability to make transfers is limited by its current output, since this output may not be large enough for appeasement to be feasible. Alternatively, as Fearon (1996) and Powell (2006) have argued, issue indivisibility may arise if even a small transfer to North today would make North dramatically more likely to win a war in the future. In this case, a distribution of utility that both prefer to war may be impossible to attain, as even a small transfer makes North so powerful that South's continuation utility falls drastically.

Another type of commitment problem occurs if a war does not necessarily end the game. In Section 4 we change the game so that after the war has started, the players may negotiate an end to the war before the war has become very costly. Recall that Section 2.2 showed that war may occur in equilibrium if North has private information about his type. But in the equilibrium discussed in Section 2.2, only the tough type declares war. Thus, by declaring war, North in effect proves he is tough, i.e., he no longer has private information about his type. The Coase theorem would then suggest that the two parties immediately negotiate a truce, and the tough type collects the side-payment that makes him prefer peace to war. But again this raises the issue of bluffing: if a very short war is not very costly to the weak type, then he may declare war as well, pretending to be tough in order to collect a large side-payment. Thus, it can no longer be the case that only the tough type declares war — the weak type must also declare war with some probability. But the *Coase conjecture* suggests that if each "bargaining period" (the time before a new offer can be made if an offer is rejected) is short, then any war must end quickly, and both weak and tough types can extract large side-payments (Gul, Sonnenschein and Wilson, 1986; Powell, 2004b, 2006; Fearon, 2007).[3] To screen North's types, and more precisely to make bluffing less tempting for the weak type, South would like to make a commitment not to terminate the war quickly. It may however be difficult for South to make such a commitment (although one can imagine that South wants to maintain a reputation for toughness and therefore has an incentive to keep fighting). Without the power to commit, the Coase conjecture logic seems to force South to end the war quickly by making a large transfer even when North is weak. But this doesn't square with the facts, since in reality not all wars end quickly.

However, the Coase conjecture is not always valid. For one thing, it holds only under one-sided incomplete information. That is, only one side, say North, can have private information. With two-sided incomplete information, Myerson and Satterthwaite (1983) show (in a buyer-seller model) that it may be impossible to reach an efficient outcome even under full commitment. Lack of commitment puts yet more constraints on implementability and typically implies even more

[3] To avoid misunderstanding, notice that the Coase conjecture is not the same as the Coase theorem.

inefficiency. The result is a war of attrition, where each side tries to prove its case for a bigger concession from the other (see Chatterjee and Samuelson, 1988, or Abreu and Gul, 2000). Even with one-sided incomplete information, for the Coase conjecture to hold, North's private information must be restricted to his cost of fighting a war. Other kinds of private information may be important, however, such as information about military strength (see Powell, 2004b and Fearon, 2007). In Section 4, we assume North has private information about how long he is able to fight, and we show that the Coase conjecture is not valid in this case. Even if South lacks commitment power, he will use a strategy which Fearon (2007) calls "screening by fighting", which forces North's tough type to fight a prolonged war. Thus, with incomplete information, wars may last a long time even though the players lack the ability to *commit* to fighting long wars.

Finally, we note that in the Brito and Intriligator (1985) model, it is always good to be militarily strong. First, if there is a war, military strength increases the probability of winning. Second, military strength may prevent the opponent from attacking, as he is unwilling to start a war that he is likely to lose. Third, if there is no war, military strength implies a better bargaining position, hence a more favorable reallocation of wealth. Fourth, a military build-up can be a *signal* which increases bargaining power even further. For these four reasons, in this model a country will always benefit by becoming militarily stronger.[4] In reality, however, a build-up of military strength may have an unintended consequence: it may instill fear in other countries, and this can make them attack in order to eliminate the threat (for a formal model, see Baliga and Sjöström, 2008). However, the models we survey here do not capture the idea that fear can be a reason for starting a war.

II. Guns or Butter

Military power can be used to take something forcibly. But it can also be used to persuade someone to give up something without a fight. As emphasized by Schelling (1966), military power is bargaining power. This is illustrated here in a simple model, based on Brito and Intriligator (1985). Related models, where agents choose between generating surplus or extracting surplus from others, include Garfinkel (1990), Hirshleifer (1988), Neary (1997), Skaperdas (1992) and Skogh and Stuart (1982).

Two risk-neutral players, N and S, represent two countries, North and

[4] Of course, this benefit must be traded off against the reduced production of consumption goods when the military expands. For the world as a whole, military build-ups unambiguously reduce welfare, even if there is no war, as they crowd out consumption.

South.[5] Player $i \in \{N, S\}$ has resources x_i which can be used to produce guns $g_i \geq 0$ and butter $b_i \geq 0$. The *budget constraint* for player $i \in \{N, S\}$ is

$$g_i + b_i = x_i. \tag{1}$$

Utility is linear in butter: each unit of butter consumed yields one unit of utility. Guns yield no direct utility.

If there is a war, the country which has more guns is more likely to win. (We assume no war can occur if $g_N = g_S = 0$.) The winner takes all available butter, $b_N + b_S$, and the loser gets nothing. Specifically, player i wins the war with probability

$$\rho(g_i, g_j) \equiv \frac{g_i}{g_i + g_j}.$$

The function ρ is the *contest success function* (CSF).[6] If there is a war, then each player i suffers a cost $c_i > 0$.

To simplify, assume South is rich and has a high cost of war, while North is poor and has a low cost of war. Specifically, $x_N < c_S$ so North does not have enough resources to make war worthwhile for South, but $x_S > c_N$ so South has enough resources to (possibly) make war worthwhile for North. Thus, South would never want to start a war, but North might. Consider then the following two-stage game.

Stage 1: Productive decisions. Each player $i \in \{N, S\}$ chooses g_i and b_i subject to the budget constraint (1). These decisions are simultaneous and are publicly observed.

Stage 2: Bargaining. South proposes to transfer an amount t of butter to North, where $0 \leq t \leq b_S$. North either accepts this proposal or declares war.

Section 2.1 considers equilibrium of this game under complete information. Section 2.2 considers incomplete information.

[5] More precisely, the players could be thought of as the leaders of their respective countries. In any case, it is a "unitary actor" model which does not consider the possibility that different agents *within* a country may have different interests. See Jackson and Morelli (2007) for a model where a leader's political bias can lead to war, and Fearon (1994) for a model where political leaders may suffer "audience costs".

[6] This specific CSF goes back to Tullock (1980). It is convenient to assume the winner takes all, but more generally, war leads to a reallocation of resources where $\rho(g_i, g_j)$ is player *i*'s expected share. Skaperdas (1996) axiomatizes several common CSFs. Corchón (2007) provides an authoritative survey of the theory of contests.

2.1 Complete Information

Assume the game has complete information. If South offers a sufficiently high side-payment (transfer) to North at stage 2, then North will not declare war. Since South knows North's preferences, South knows the minimum side-payment that will prevent a war. Since a war would be costly, South's best option is to make this minimum transfer, a policy of appeasement which prevents war. This is a version of the Coase theorem, which states that rational agents who can bargain freely will negotiate a surplus-maximizing outcome.[7]

More precisely, North will accept South's proposal if North's consumption of butter, $b_N + t$, will exceed North's expected payoff from war, which is

$$\frac{g_N}{g_S + g_N}(b_S + b_N) - c_N . \tag{2}$$

That is, the proposal is accepted if

$$b_N + t \geq \frac{g_N}{g_S + g_N}(b_S + b_N) - c_N . \tag{3}$$

The inequality (3) is the *appeasement constraint*. In the bargaining stage, South's problem is to find the smallest non-negative t that satisfies the appeasement constraint. Using the budget constraints, it can be checked that the solution to this problem is to set

$$t = \frac{g_N x_S - g_S x_N}{g_S + g_N} - c_N \tag{4}$$

if this expression is non-negative, and $t = 0$ otherwise. (If (4) is negative, then North does not declare war even if South offers $t = 0$, so of course, this is what South will propose.) Substituting from the budget constraints into (4) reveals that $t < b_S$. Thus, the constraint $t \leq b_S$ does not prevent appeasement.

In any equilibrium, North will set $g_N > 0$ and get a positive transfer $t > 0$. For if $g_N = 0$, then South would have no motive to produce guns either, so $g_S = 0$. But then North could produce an arbitrarily small amount of guns, get a transfer $x_S - c_N > 0$ by (4), and be better off. Thus, $g_N > 0$ must hold. But then

[7] The equilibrium outcome is not fully efficient because guns are produced. However, the countries are assumed to be unable to negotiate arms-control agreements. The Coase theorem therefore only applies to stage 2, not to stage 1.

$t > 0$ must hold, for otherwise North would be better off setting $g_N = 0$. In equilibrium, there must be appeasement: South transfers a positive amount of butter to North and there is no war. The equilibrium transfer is given by (4).

It remains only to calculate exactly how many guns will be produced in equilibrium. To simplify this calculation, we assume North is quite poor,

$$x_N < \frac{1}{3} x_S, \tag{5}$$

and has a low cost of war,

$$c_N < \sqrt{x_N (x_S + x_N)} - 2x_N . \tag{6}$$

Proposition 1 *Suppose (5) and (6) hold. The following is a subgame perfect equilibrium outcome. In stage 1, South chooses*

$$g_S = \sqrt{x_N (x_S + x_N)} - x_N \tag{7}$$

and North chooses

$$g_N = x_N . \tag{8}$$

In stage 2, South transfers

$$t = \sqrt{x_N (x_S + x_N)} - x_N - c_N > 0 \tag{9}$$

and there is no war.

The proof is in the Appendix. Notice that North allocates *all* of its productive resources to weapons production. This extreme result is due to our strong assumptions.[8] But the intuition is clear: the poor tend to specialize in fighting ability because "the poor have a *comparative advantage* in conflict as opposed to production" (Hirshleifer, 1994, page 7).[9]

In equilibrium, South allocates a fraction

---

[8] Equilibria where the poor put all their resources into weapons production are called "banditry equilibria" by Neary (1997).

[9] Hirshleifer (1991) discusses the fact that, as the poor specialize in fighting, the outcome is not that "the rich get richer and the poor poorer", but rather a redistribution of wealth from the rich to the poor (which he calls the "paradox of power").

$$\frac{g_S}{x_S} = \sqrt{\frac{x_N}{x_S}(1+\frac{x_N}{x_S})} - \frac{x_N}{x_S} \tag{10}$$

of its productive resources to weapons production. This fraction is less than 1/2 but increasing in x_N / x_S. Thus, if North's productive capacity x_N increases slightly, South shifts resources into weapons production. Since North produces no butter, $b_S + b_N$ is decreasing in x_N. But $b_S + b_N$ is the global surplus (the total utility in the two countries). A small increase in North's productivity reduces global welfare because North becomes more threatening and global weapons production increases by more than the increase in North's productivity. Helping North become more productive can be *counter* productive in this sense.[10]

Notice that the equilibrium transfer is independent of South's cost of war, c_S. Of course, the costlier a war is to South, the more South is willing to transfer, but this is irrelevant. The decision to start a war is up to North, and its decision is based purely on its *own* cost-benefit calculation, not on how costly a war would be to South. Therefore, c_S doesn't enter into (4). To put it differently, if North wants to extort a big transfer from South, rather than trying to convince South that war would be very costly to South, North should try to convince South that war would not be very costly to North.[11]

This model can explain arms races but not wars. Since bargaining at stage 2 is not subject to any frictions or "transaction costs", the Coase theorem applies, and the players will surely negotiate a peaceful outcome. To explain wars, we must introduce some kind of friction so the Coase theorem no longer applies. Brito and Intriligator (1985) introduced incomplete information. This is considered in Section 2.2. In Section 3.3 we consider another kind of friction: limits on side-payments.

2.2 Incomplete Information and Signaling

Suppose South doesn't know North's *type*. With probability p, North is a *tough type* with cost $c_N < x_S$ just as before. But with probability $1-p$, North is a *weak type* with cost $\overline{c}_N > x_S$. North, of course, knows his own true type. If North is weak, he will never declare war because his cost is too high to make it worthwhile. But if North is tough, he might declare war unless he gets a sufficiently high

[10] Major improvements in North's productivity are a different story, because they will change the nature of the equilibrium. As North becomes less and less poor relative to South, its "comparative advantage in conflict" will tend to disappear and North will no longer devote all its resources to weapons production.

[11] North can benefit from a high c_S if the bargaining game is different. Suppose it is North who makes a take-it-or-leave-it proposal (demands a transfer from South). If South refuses, there is war. In this case, the responsibility to decide between war and peace is shifted to South, and North will be able to extract a very large amount of butter if c_S is very big.

transfer, just as in Section 2.1. Since South does not know North's type but can observe g_N, we have a *signaling game*. Formally, the game now has a stage 0, where North (but not South) learns North's type. After that, stages 1 and 2 are the same as before.[12]

This game has no pure strategy separating equilibrium for generic parameter values.[13] If p is close to one, then a pure strategy pooling equilibrium exists. In the pooling equilibrium, South cannot distinguish the two types, but since he thinks North is probably tough, he is willing to appease. He pays the transfer given by (4), just as in Section 2.1, and there is no war.

The more interesting case occurs if p is fairly small, i.e., South thinks that North is probably weak. In this case, there exists a semi-separating equilibrium, where even if North arms himself in stage 1, South may refuse to make any transfer because he thinks North is probably a weak type who is just bluffing. If North is actually tough, a war may ensue.

Proposition 2 *Suppose (5) and (6) hold, and in addition*

$$c_N + pc_S < (1-p)\left(\sqrt{x_N(x_S + x_N)} - x_N\right).$$

There exists a semi-separating equilibrium. North chooses $g_N = x_N$ *if he is tough. If North is weak, then he chooses* $g_N = x_N$ *("bluffing") with probability* q*, where*

$$q = \frac{p}{1-p}\frac{c_N + c_S}{\sqrt{x_N(x_S + x_N)} - x_N - c_N}, \tag{11}$$

and he chooses $g_N = 0$ *with probability* $1-q$. *South chooses*

---

[12] This signaling game formulation differs from Brito and Intriligator (1985). They assumed South could commit to a policy of screening. The main insights are the same in both formulations.

[13] The proof of this is by contradiction. Let $g_N(c)$ be the amount of guns that North produces when his cost type is $c \in \{c_N, \overline{c}_N\}$. In separating equilibrium $g_N(\overline{c}_N) \neq g_N(c_N)$. When South observes $g_N = g_N(\overline{c}_N)$ he knows North is weak, so South will not make any transfer, and there is no reason for the weak type to build any weapons. Hence, $0 = g_N(\overline{c}_N) < g_N(c_N)$, and the weak type of North gets a utility of x_N. When South observes $g_N = g_N(c_N)$ he knows North is tough, and South offers a transfer t given by (4) which prevents war. But North's weak type could imitate the tough type, choose $g_N = g_N(c_N)$, and get the same transfer (since South would think North is tough). Similarly, North's tough type could mimic the weak type. For neither deviation to be profitable, we must have $t = g_N(c_N)$. The tough type of North will choose $g_N(c_N)$ to maximize his payoff from war, given by (2), since even if there is no war, the payoff is the same, by (4). South will choose g_S to maximize his payoff given $g_N(c_N)$ and $g_N(\overline{c}_N) = 0$. Only for knife-edge parameter values will simultaneously solving these maximization problems yield $t = g_N(c_N)$ which satisfies (4).

$$g_S = \sqrt{x_N(x_S + x_N)} - x_N \, . \tag{12}$$

When South sees $g_N = x_N$ *he offers a transfer* $t > 0$ *with probability* α*, where*

$$t = \sqrt{x_N(x_S + x_N)} - x_N - c_N \tag{13}$$

and

$$\alpha = \frac{1}{\sqrt{\frac{x_S}{x_N} + 1} - \frac{c_N}{x_N} - 1} \, . \tag{14}$$

Both types of North accept this transfer. With probability $1-\alpha$*, South "calls North's bluff" and offers no transfer when he sees* $g_N = x_N$*, and then North declares war if and only if he is tough. In equilibrium, a war occurs with probability*

$$\omega = p(1-\alpha) = p - \frac{p}{\sqrt{\frac{x_S}{x_N} + 1} - \frac{c_N}{x_N} - 1} > 0 \, . \tag{15}$$

The proof is in the Appendix.

The semi-separating equilibrium resembles a game of poker. There is a positive probability that North's weak type will bluff, i.e., produce guns in the hope that South will fold and choose appeasement. But South calls North's bluff with positive probability; North then backs down if he is weak, but declares war if he is tough.

From (15) we get $\frac{\partial \omega}{\partial p} = 1 - \alpha > 0$. Thus, in the parameter region where the semi-separating equilibrium exists, the probability of war increases with the probability that North is tough, as is intuitive. Next, notice that $\frac{\partial g_S}{\partial x_S} > 0$ and $\frac{\partial \alpha}{\partial x_S} < 0$ so that $\frac{\partial \omega}{\partial x_S} > 0$. Thus, if South becomes more productive (i.e., if x_S increases), then he produces more guns and becomes more reluctant to choose appeasement, making war more likely. In the previous subsection we saw that an increase in North's productive capacity may reduce the total production of butter; now we see that an increase in South's productive capacity may increase the risk of war.

South wins the war with probability

$$\frac{g_S}{g_S + g_N} = 1 - \frac{1}{\sqrt{\frac{x_S}{x_N} + 1}} \, ,$$

which is increasing in $\frac{x_S}{x_N}$. In fact, if c_N is small enough that the term $\frac{c_N}{x_N}$ is negligible, or if c_N is proportional to x_N (as is plausible) so that $\frac{c_N}{x_N}$ is a constant, then ω is also increasing in $\frac{x_S}{x_N}$. Intuitively, the more unbalanced is the situation, in the sense that South is relatively more productive than North ($\frac{x_S}{x_N}$ is big), the more likely South is to win a war, the more likely South is to call North's bluff, and the more likely it is that a war occurs.

Proposition 2 confirms that private information may lead to war. It may be objected that the stage 2 bargaining game is quite special, with South making a take-it-or-leave-it offer. Perhaps some other bargaining game could be designed such that war never occurs in equilibrium? Of course, with unlimited freedom to design stage 2, this can be achieved: let a "world government" either block North from going to war, or compel South to make a transfer that North surely prefers to war. But this would be a very unrealistic game. It seems that any realistic stage 2 should satisfy the following criteria: (i) North must have the option to declare war, and (ii) any transfer from South to North should be voluntary. In addition, in stage 1, (iii) each country should be free to produce as many guns as it wants. But let us here assume military strength is exogenously given, so stage 1 is redundant and we can neglect criterion (iii).[14] Moreover, assume both types of North have the same military strength, so that South cannot infer North's true type by observing North's military strength.

Thus, we fix $g_S > 0$ and $g_N > 0$ and focus on stage 2. Assume the expression in (4) is positive, so the tough type will go to war if he is not offered a sufficient transfer. Is there any bargaining game that can be used at stage 2, satisfying criteria (i) and (ii), which eliminates the risk of war? In general, the answer is no. To see this, notice that, since South does not know North's type, the equilibrium of any such game must satisfy an incentive-compatibility condition for North: no type can profitably gain by imitating the other type. This has a very strong implication: if there is no war in equilibrium, each type of North must get exactly the same transfer (for otherwise, the type that gets the smaller transfer will have a profitable deviation). Let t^* denote this transfer. Does there exist t^* such that (i) North's tough type prefers to accept t^* rather than go to war, and (ii) South prefers to pay t^* to both types, rather than to refuse to pay which would imply a probability of war of at most p (since the weak type of North would never declare war)? If p is small enough, the answer is clearly no, because (ii) forces t^* to be small, but then (i) cannot hold. For more formal treatments of this kind of argument (in different models), see Bester and Wärneryd (2006) and Fey and Ramsay (2009, 2011).

[14] This simplification is due to space limitations. A similar argument will go through when gun production is endogenous.

III. Commitment Problems, Limited Transfers and Salami Tactics

In Section 2, North's acceptance of South's proposal ended the game. In this section, we consider the role of side-payments when the future matters. We assume no player has any private information, and consider whether commitment problems can cause wars under complete information. We will maintain, in this section, the assumption that a war in effect ends the game; the winner takes everything and the loser is wiped out. In Section 4, we will consider instead the possibility that bargaining continues after a war has started. In other words, there are two kinds of commitment problems: the inability to commit to future actions in peacetime, and the inability to commit to future actions after a war has started. We consider the first in this section, and the second in Section 4.

Section 3.1 shows that even if it is not possible to commit to future actions, there is no war if utility is transferable (no restrictions on side-payments). Following Fearon (1996), Section 3.2 shows how we can relax the transferable utility assumption: there is no war if the *productive asset* is transferable, as long as the CSF is continuous in asset ownership. In Section 3.3, we restrict the ability to make side-payments even further: South has an asset which will be productive both now and in the future, but the asset itself cannot be transferred. Following Acemoglu and Robinson (2001), we assume transfers must come from the current output; South is unable to borrow against its asset. We show that this can make war unavoidable.

In Section 3.3, the productive asset is valuable, but the future is not modelled explicitly. This can be justified by assuming that if North does not defeat South today, then North will be unlikely to defeat South in the future, perhaps because North's strength is declining, so that North has to take the asset today or never. Powell (2004a) discusses how such anticipated shifts in the environment, combined with limits on current side-payments, may cause wars.

3.1 Transferable Utility

Since war is costly, with transferable utility (unrestricted side-payments) there always exists a side-payment (a transfer of utility) that both South and North prefer to war; with complete information, there is no ambiguity about how big this side-payment should be. Therefore, no war can occur in equilibrium (in accordance with the Coase theorem). Here we verify this in a rather general dynamic model, where the players lack the ability to commit to future actions.

There are infinitely many periods denoted $\tau = 1,2,3\ldots$ The discount factor is δ. There is complete information and no upper bound on side-payments.[15] (It

[15] Powell (1993) and Jackson and Morelli (2009) show how wars may occur in the (Markov perfect)

does not matter if productive assets are transferable so we make no assumption about this.) Let x_τ denote the *state of the world* at the end of period τ, which is inherited at the beginning of period $\tau+1$. Let X denote the set of all possible states.

In each period τ, player $i \in \{N,S\}$ takes an action $z_{i\tau}$. For example, $z_{i\tau}$ may be player *i*'s decision of how many guns to produce, how much to consume and how much to invest in order to increase the capital stock, etc. The feasible set of actions $Z_i(x_{\tau-1})$ in period τ can depend on the state $x_{\tau-1}$. Within period τ there are two stages.

Stage 1: Productive decisions. Each player $i \in \{N,S\}$ chooses $z_{i\tau} \in Z_i(x_{\tau-1})$. These decisions are simultaneous and are publicly observed.

Stage 2: Bargaining. South proposes to transfer an amount of butter $t \geq 0$ to North. North either accepts this proposal or declares war.

The important point is that, as in Fearon (1996), if North accepts a proposal then he maintains the option of either extracting future transfers or declaring war in some future period. This model differs from Fearon (1996) because transfers are in terms of butter. Since utility is linear in butter there is "transferable utility". (In Section 3.2 we will instead make Fearon's (1996) assumption that transfers consist of resources that influence the state variable.)

If there is no war in period τ, then player $i \in \{N,S\}$ gets utility (not including any transfer) $u_i(z_{i\tau})$ in period τ, and the state variable evolves according to the equation

$$x_\tau = F(x_{\tau-1}, z_{N\tau}, z_{S\tau}).$$

If there is a war then player i suffers a cost c_i, and he wins the war with probability $\rho_i(x_{\tau-1}, z_{N\tau}, z_{S\tau})$. This is the CSF in this model. Naturally,

$$\rho_N(x_{\tau-1}, z_{N\tau}, z_{S\tau}) + \rho_S(x_{\tau-1}, z_{N\tau}, z_{S\tau}) = 1.$$

After a war, the strategic interaction becomes trivial because the loser is in effect wiped out. The winner will make consumption and investment decisions to maximize his payoff, knowing that he controls all resources forever. If player i wins the war then he gets continuation payoff $W_i(x_{\tau-1}, z_{N\tau}, z_{S\tau})$, while the loser's continuation payoff is normalized to 0.

Notice that the specification of a state can include the period, e.g., we may have $x_\tau = (k_{N\tau}, k_{S\tau}, \tau)$ where $k_{i\tau}$ is country *i*'s capital stock at the end of period τ.

---

equilibria of infinite horizon guns-versus-butter models with limits on side-payments.

This allows, in particular, the CSF to shift over time, allowing one player to become more powerful for some exogenous reason (see Powell, 2004a, for the importance of such "power shifts" when side-payments are limited).

Consider Markov perfect equilibrium of this game. Let V_i denote player i's value function. At the beginning of stage 1 of period τ, player i's expected continuation payoff is $V_i(x_{\tau-1})$.

Suppose decisions $(z_{N\tau}, z_{S\tau})$ are made in state $x_{\tau-1}$ and consider the bargaining stage. If North accepts South's proposed transfer t, North's continuation payoff will be

$$t + u_N(z_{N\tau}) + \delta V_N(F(x_{\tau-1}, z_{N\tau}, z_{S\tau})).$$

If North rejects the proposal, there is war; North's expected continuation payoff is

$$\rho_N(x_{\tau-1}, z_{N\tau}, z_{S\tau})\, W_N(x_{\tau-1}, z_{N\tau}, z_{S\tau}) - c_N.$$

North will accept South's proposal if and only if

$$\begin{aligned} &t + u_N(z_{N\tau}) + \delta V_N(F(x_{\tau-1}, z_{N\tau}, z_{S\tau})) \\ &\geq \rho_N(x_{\tau-1}, z_{N\tau}, z_{S\tau})\, W_N(x_{\tau-1}, z_{N\tau}, z_{S\tau}) - c_N. \end{aligned} \tag{16}$$

If this holds for $t = 0$, then clearly South will propose $t = 0$ and there is no war. So suppose (16) requires $t > 0$. Then South can avoid war by proposing

$$\begin{aligned} t = {} &\rho_N(x_{\tau-1}, z_{N\tau}, z_{S\tau})\, W_N(x_{\tau-1}, z_{N\tau}, z_{S\tau}) - c_N \\ &- \delta V_N(F(x_{\tau-1}, z_{N\tau}, z_{S\tau})) - u_N(z_{N\tau}). \end{aligned} \tag{17}$$

Does South want to make this proposal to avoid war? If he does, South's continuation payoff will be

$$\begin{aligned} &u_S(z_{S\tau}) - t + \delta V_S(F(x_{\tau-1}, z_{N\tau}, z_{S\tau})) \\ &= u_S(z_{S\tau}) + u_N(z_{N\tau}) - \rho_N(x_{\tau-1}, z_{N\tau}, z_{S\tau})\, W_N(x_{\tau-1}, z_{N\tau}, z_{S\tau}) \\ &\quad + \delta V_N(F(x_{\tau-1}, z_{N\tau}, z_{S\tau})) + c_N + \delta V_S(F(x_{\tau-1}, z_{N\tau}, z_{S\tau})) \end{aligned} \tag{18}$$

where the equality uses (17). If there is war, South's expected continuation payoff will be

$$\rho_S(x_{\tau-1}, z_{N\tau}, z_{S\tau})\, W_S(x_{\tau-1}, z_{N\tau}, z_{S\tau}) - c_S. \tag{19}$$

Thus, South strictly prefers to pacify North if (18) exceeds (19), which is true if and only if

$$\begin{aligned} & u_N(z_{N\tau}) + u_S(z_{S\tau}) + \delta[V_N(F(x_{\tau-1}, z_{N\tau}, z_{S\tau})) + V_S(F(x_{\tau-1}, z_{N\tau}, z_{S\tau}))] \\ > & \rho_S(x_{\tau-1}, z_{N\tau}, z_{S\tau}) W_S(x_{\tau-1}, z_{N\tau}, z_{S\tau}) + \rho_N(x_{\tau-1}, z_{N\tau}, z_{S\tau}) W_N(x_{\tau-1}, z_{N\tau}, z_{S\tau}) \\ & - c_N - c_S . \end{aligned}$$

The left hand side is the sum of the continuation payoffs if there is no war this period; the right hand side is the continuation payoff for the country that wins a war (recall that the loser gets zero), minus the total cost of war. Thus, this inequality holds if peace yields a greater global surplus than war. Any reasonable specification of underlying technology and preferences will satisfy this inequality. Otherwise war would be surplus-maximizing, which is not the case that we are interested in.

To summarize, war cannot happen in this model because South will always use appeasement to prevent war. This is just a restatement of the Coase theorem: if wars destroy surplus then there is always some transfer that will make both better off. This is true even though the players cannot commit to future actions, as long as utility is perfectly transferable (no restrictions on side-payments). Section 3.2 discusses Fearon's (1996) result that war may be avoided even if transfers consist of resources that determine future bargaining power.

3.2 Transfers that Change the State

We now consider a simplified version of the infinite-horizon model of Section 3.1, modified in one key respect: now transfers consist of productive resources which directly determine bargaining power. Thus, if North receives an initial transfer, he becomes more powerful, and may return later to extract even more. Anticipating this, wouldn't South refuse to make the initial transfer to North, and couldn't this refusal lead to war? This issue was studied by Fearon (1996), who found that war will be avoided as long as transfers do not cause *discontinuous* changes in bargaining power (see also Powell, 1996).

To simplify, assume the total amount of resources to be divided among the two countries is constant and normalized to 1. For example, the two countries may contest a fixed land area of size 1. There is no guns-or-butter decision. Stage 1 is then redundant and only the bargaining stage is nontrivial. The resource is perfectly durable and is divisible; any part of it can be transferred from one country to the other. It is convenient here to assume that in the bargaining stage it is North who makes a take-it-or-leave-it offer; if South doesn't accept then there is war. As in Section 3.1, if South accepts then North can still come back and demand more in the future.

We let the state variable $x_{\tau-1}$ denote North's share of the resource coming into period τ (so $1-x_{\tau-1}$ is South's share). The resource generates military power: the probability North wins a war in period τ is $\rho(x_{\tau-1})$, where ρ is the CSF. South wins with probability $1-\rho(x_{\tau-1})$. A player is militarily stronger the more of the resource he controls, so $\rho(x_{\tau-1})$ is increasing in $x_{\tau-1}$. Assume for now that ρ is continuous. Transfers must be in terms of the resource as it is the only good available.

In the bargaining stage of period τ, North makes a demand x_τ. If South accepts, then the state changes to x_τ, and since the players derive utility directly from the resource, they get utilities $u_N(x_\tau)=x_\tau$ and $u_S(x_\tau)=1-x_\tau$ in period τ. If South rejects, then there is a war which effectively ends the game: the loser is wiped out, and the winner holds the total resource of 1 forever. Thus, if there is war in period τ, North's expected payoff is

$$\rho(x_{\tau-1})(1+\delta+\delta^2+\ldots)-c_N=\frac{\rho(x_{\tau-1})}{1-\delta}-c_N.$$

For South, it is similarly

$$\frac{1-\rho(x_{\tau-1})}{1-\delta}-c_S. \qquad (20)$$

Consider a Markov perfect equilibrium where V_i denotes player i's value function. If South accepts North's demand, South gets utility $u_S(x_\tau)=1-x_\tau$ in period τ and $\delta V_S(x_\tau)$ after that, so his continuation payoff is

$$1-x_\tau+\delta V_S(x_\tau). \qquad (21)$$

If he rejects, his continuation payoff is given by (20).

If North demands nothing, $x_\tau=0$, South strictly prefers to accept and get continuation value $1+\delta V_S(0)$. To see this, notice that since South can always refuse any demand, we must have

$$V_S(0)\geq\frac{1-\rho(0)}{1-\delta}-c_S.$$

Thus, the expected payoff from accepting the demand $x_\tau=0$ satisfies

$$1+\delta V_S(0)\geq 1+\delta\left(\frac{1-\rho(0)}{1-\delta}-c_S\right)>\frac{1-\rho(x_{\tau-1})}{1-\delta}-c_S \qquad (22)$$

where the second inequality comes from $\delta < 1$ and $1 \geq 1 - \rho(0) \geq 1 - \rho(x_{\tau-1})$. This proves that South strictly prefers to accept a zero demand. Therefore, by (22), North will make a positive demand in equilibrium. If (21) is bigger than (20) for $x_\tau = 1$, then North will demand everything and possess the resource forever. If this is not the case, then there must exist a demand x_τ with $0 < x_\tau < 1$ that makes South indifferent between accepting and rejecting. To see this, note that by (22), and as (21) is smaller than (20) for $x_\tau = 1$, we have

$$1 + \delta V_S(0) > \frac{1 - \rho(x_{\tau-1})}{1-\delta} - c_S > \delta V_S(1) .$$

The continuity of ρ implies that V_S is continuous, so there is $x_\tau \in (0,1)$ such that

$$1 - x_\tau + \delta V_S(x_\tau) = \frac{1 - \rho(x_{\tau-1})}{1-\delta} - c_S . \tag{23}$$

Thus, (20) equals (21), so South is indifferent between accepting and rejecting the demand x_τ. By standard arguments, North makes this demand in equilibrium and South accepts. South's continuation payoff, given by (23), must equal $V_S(x_{\tau-1})$. Thus,

$$V_S(x_{\tau-1}) = \frac{1 - \rho(x_{\tau-1})}{1-\delta} - c_S .$$

By the same argument applied to the next period,

$$V_S(x_\tau) = \frac{1 - \rho(x_\tau)}{1-\delta} - c_S$$

If we plug this into (23) we find that

$$1 - x_\tau + \delta\left[\frac{1 - \rho(x_\tau)}{1-\delta} - c_S\right] = \frac{1 - \rho(x_{\tau-1})}{1-\delta} - c_S$$

If for example the CSF is $\rho(x) = x$, then this equation yields

$$1 - x_\tau + \delta\left[\frac{1 - x_\tau}{1-\delta} - c_S\right] = \frac{1 - x_{\tau-1}}{1-\delta} - c_S$$

or

$$x_\tau = x_{\tau-1} + c_S(1-\delta)^2 > x_{\tau-1}$$

until x_τ hits 1. Thus, for any initial state x_0, x_τ will increase monotonically and eventually reach 1 and then stay there forever (although it may take a long time to reach 1 if δ is close to 1; or if South's cost of conflict is small). Eventually, North gets all of the resource. He does not get everything right away, only piece by piece, a process known as "salami tactics". War never breaks out since bargaining is efficient. The assumption that North has all the bargaining power (he makes take-it-or-leave-it offers) is not important for the result that there is no war in equilibrium. However, the assumptions we make about bargaining power will obviously influence the long-run distribution of the resource.

Even maintaining the assumption that North makes take-it-or-leave it offers, the result that North eventually takes everything does not hold for general CSFs. Indeed, if x^{**} satisfies

$$x^{**} - \rho(x^{**}) = (1-\delta)c_S$$

then the stationary outcome $x_\tau = x^{**}$ for all τ is a Markov perfect equilibrium outcome. But in any case, there is never war if the CSF is continuous.

Fearon (1996) showed that if the CSF is discontinuous then there may be war in equilibrium. To see this simply, suppose the CSF is as follows:

$$\rho(x) = \begin{cases} \rho_L & \text{if } x < x^* \\ \rho_H & \text{if } x \geq x^* \end{cases}$$

Assume

$$x_0 < x^* < \rho_L < \rho_H$$

where x_0 is the initial state. North's expected payoff from a war in period 1 is then

$$\frac{\rho(x_0)}{1-\delta} - c_N = \frac{\rho_L}{1-\delta} - c_N .$$

Assume δ is close enough to 1 so that

$$\frac{\rho_L}{1-\delta} - c_N > \frac{x^*}{1-\delta} . \tag{24}$$

Suppose there is no war in equilibrium. Then we must have $x_t \geq x^*$ in some period t, for otherwise North's expected payoff is less than $\frac{x^*}{1-\delta}$ and he would prefer a war, by (24). Let t be the first period such that $x_t \geq x^*$. Thus, $x_{t-1} < x^* \leq x_t$. North's expected payoff from a war in period $t+1$ will be

$$\frac{\rho(x_t)}{1-\delta} - c_N = \frac{\rho_H}{1-\delta} - c_N .$$

So, in period $t+1$, North's continuation payoff must be at least $\frac{\rho_H}{1-\delta} - c_N$, and South's continuation payoff will be at most

$$\frac{1}{1-\delta} - \left(\frac{\rho_H}{1-\delta} - c_N \right) = \frac{1-\rho_H}{1-\delta} + c_N$$

since the total surplus (in the absence of war) is $\frac{1}{1-\delta}$. Thus, in period t, South's expected payoff along the equilibrium path is at most

$$1 - x_t + \delta \left(\frac{1-\rho_H}{1-\delta} + c_N \right). \tag{25}$$

Since $x_{t-1} < x^*$, if there is war in period t, then South's expected payoff is

$$\frac{1-\rho_L}{1-\delta} - c_S .$$

If δ is sufficiently close to 1 then this expression exceeds (25), so South prefers war. This contradicts the hypothesis that there is no war in equilibrium. Thus, if δ is close to 1 there must be war in equilibrium. Intuitively, South prefers to fight when his chance of winning is $1-\rho_L$ rather than appeasing North, because any resource transfer that North would prefer to war would cause South's chance of winning a war to drop discontinuously to $1-\rho_H$. This would reduce South's future bargaining power so much that he prefers war. More generally, by invoking continuity or discontinuity of the CSF, this theory can account for both periods of peace and periods of war. But such explanations have many degrees of freedom, a situation which could be addressed by a theory of the sources of discontinuities.

3.3 Transfers that are Limited by Current Output

Even with complete information, a war may occur if the players are sufficiently constrained in their ability to make side-payments. In Section 2, the constraint

$t \leq b_S$ was irrelevant: there was no conceivable reason to transfer more than 100% of South's current stock of butter. But if the dispute concerns *future* butter production, then transferring all of the butter that South *currently* has may not be enough to pacify North. This might still not be a problem if (as in Section 3.2) some part of South's productive capacity, represented by x_S, could be transferred to North. For example, if South's superior productivity is due to its land or physical capital, then some part of its land or capital could be given to North. If South's superior productivity is due to its human capital, perhaps embodied in skilled engineers, then South could send its engineers to train Northern engineers, or Northerners could be invited to study in the South. But it may be that human capital is not transferable in this way, and the only way for North to seize South's human capital is to conquer South (assuming South's engineers can be productively employed by North after the war). Then, we would have a situation where the productive asset could be seized in a war, but it could not otherwise be transferred. Similarly, if the productive asset is land, there may be a taboo against transferring land during peacetime.

Whatever the reason for it, let us assume South's productive asset cannot be transferred unless there is a war. Moreover, South cannot credibly commit to make any transfers to North *in the future*. The same commitment problem makes international bankers unwilling to lend to South: South cannot credibly commit to repay the loan and, since it is a sovereign state, it cannot be forced to do so. Therefore, South is unable to borrow against its asset. These assumptions are inspired by Acemoglu and Robinson's (2001) model, where the poor can stage a revolution which allows them to expropriate some of the rich's assets. In their model, the rich would like to use transfers to buy off the poor, but this may be impossible because the transfers are limited by the current output of the economy.

Consider again the (complete information) two-stage game of Section 2, but with one modification. The countries are contesting not just the currently available butter, but also South's productive resources; implicitly, because they can be used to produce butter in the future. But we do not here model the future explicitly: we collapse the future into the value of the productive asset, which North must capture today or never. This could be justified by assuming that if North does not defeat South today, then North will be unlikely to defeat South in the future, perhaps because North's strength is declining, or (following Acemoglu and Robinson, 2001) its opportunity cost of conflict becomes very high. Powell (2004a) discusses how such anticipated shifts in the environment, combined with limits on current side-payments, may cause wars.

If there is a war, the winner takes both all of the currently available butter $b_S + b_N$ and all of South's productive resource x_S; the loser gets nothing. Each unit of South's productive resource is worth η to its possessor. Transfers can only be shipments of butter: South's productive resource cannot be transferred unless there is war. South does not have a line of credit, so the transfer t must satisfy $t \le b_S$. There is complete information.

North's expected payoff from war is

$$\frac{g_N}{g_S + g_N}[b_S + b_N + \eta x_S] - c_N . \tag{26}$$

If North accepts South's transfer, North's consumption of butter is $b_N + t$. Therefore, the proposal is accepted if

$$b_N + t \ge \frac{g_N}{g_S + g_N}[b_S + b_N + \eta x_S] - c_N . \tag{27}$$

We omit the calculations, but it should be clear that if we make assumptions similar to those in Section 2, i.e., North is relatively poor and c_N is small, then North sets $(b_N, g_N) = (0, x_N)$. Then (27) becomes

$$t \ge \frac{x_N}{g_S + x_N}[b_S + \eta x_S] - c_N . \tag{28}$$

Since $g_S \le x_S$, the right hand side of (28) exceeds x_S if η is sufficiently big. But since $b_S \le x_S$, (28) would require $t > b_S$, which is impossible. Therefore, war must occur if η is sufficiently big, that is, if the non-transferable resource is sufficiently valuable.[16] A similar result is obtained if the resource is transferable but indivisible, since if η is sufficiently big then South prefers to go to war rather than give up *all* of x_S.

To summarize Section 3, if utility is perfectly transferable, or if productive assets can be transferred and the probability of winning a war as a function of these assets is continuous, then with complete information there is no war in equilibrium. This is true even when the players lack the ability to commit to future actions. But if there are limits on transfers, in particular when transfers are limited by current output, or if transfers must be in terms of productive assets that cause discontinuous changes in the CSF, then there may be war in equilibrium. Simply put, the Coase theorem breaks down when there is no transfer which is feasible today and which both parties prefer to a war today (issue indivisibility). And, as emphasized by Powell (2004a, 2006), limits on transfers are especially likely to cause war when the

[16] War is actually Pareto efficient in this simple model, because (due to the restrictions on transfers) it is impossible to compensate North for not going to war. In a more explicitly dynamic model with commitment problems and limits on side-payments, such as Acemoglu and Robinson (2001), a war may occur even if it is not Pareto efficient.

parties expect shifts in the environment; for example, if North expects to become weaker in the future. In this scenario (as in Section 3.3), all of South's current output may not be enough to buy off North today, and the commitment problem makes it impossible for South to credibly promise to pay more in the future (when North will be weaker, and therefore unable to extract transfers by force).

IV. Bargaining During War

A different kind of commitment problem arises if a war which is due to incomplete information (as in Section 2.2) can last a long time and negotiations may continue during the war. The question now becomes how long the war will last before a truce is negotiated. Formal models of negotiations during wartime have been provided by Powell (2004b), Fearon (2007), Slanchev (2003), Smith and Stam (2004) and Heifetz and Segev (2005). Our discussion in this section is based on Powell (2004b) and Fearon (2007).

To be clear, while Section 3 asked whether commitment problems can cause wars under complete information, we now ask a different question: how long will wars caused by incomplete information last if the players lack the ability to commit to future actions? To see the nature of this commitment problem, assume North has private information about his type. Suppose that, as in the equilibrium of Section 2.2, only the tough type declares war. Then once the war has started, South knows that North is tough. The Coase theorem would then suggest that the two parties immediately agree on a (possibly very large) transfer that makes the tough type willing to stop fighting. The transfer may consist of a share of South's resources. But this raises the issue of bluffing: if a very short war is not very costly to the weak type, then he may declare war as well in order to collect the transfer.[17] Thus, we obtain a contradiction of the hypothesis that only the tough type declares war. We therefore conclude that in equilibrium the weak type must declare war with some probability. But as Powell (2004b) made clear, the Coase conjecture (Gul, Sonnenschein and Wilson, 1986) suggests that it is anyway true that if proposals to terminate the war can be made in rapid succession, then any war must end quickly. In equilibrium, North will extract a transfer after a very short war, even if he is weak. To prevent this from happening, South would like to commit to keep fighting for a long time. But if South cannot make such a commitment, then it seems the incomplete information model cannot explain why some wars last a long time (Fearon, 2007, Powell, 2006). However, this difficulty is due to the assumption that North's private

[17] If even a very short war is extremely costly to the weak type, then if the tough type declares war a truce can be negotiated quickly without the fear of bluffing. But the model would be inconsistent with the fact that, in reality, many wars last a long time.

information only relates to his cost of war. For other types of private information, the Coase conjecture no longer holds.

Powell (2004b) and Fearon (2007) argued that wars may last a long time if the private information concerns the likelihood of winning or losing a war in any given time period. Here, let us make a slightly different assumption: North has private information about how long he can fight a war before collapsing. Specifically, North's tough type is able to fight longer than the weak type. South doesn't know how long North is able to fight, i.e., he doesn't know North's true type. The Coase conjecture is not valid here because, in Fearon's (2007) terminology, South can "screen by fighting". South will not have to make a large transfer to the weak type since an alternative strategy is available: since South can outlast the weak type, if South simply fights long enough then the weak type must eventually accept to terminate the war with a zero transfer (or face a collapse, which amounts to the same thing). The availability of such a strategy guarantees that, for a wide range of parameters, wars will occur in equilibrium and they will not be short, regardless of how frequently proposals to terminate the war can be made. To show this formally, we will make a number of simplifying assumptions. In particular, we assume that once a settlement has been agreed upon, and a transfer is made to North, the war stops and cannot be re-started.[18] (Fearon, 2007, does not make this assumption.)

Again we will suppress the guns-or-butter decision to simplify the exposition. There is one unit of a perfectly durable resource which is initially owned by South, $x_S = 1$. The resource is divisible and part of it can be transferred from South to North. (There is no other good that can be transferred.) Each unit of the resource yields its owner a utility of 1 per unit of time. North has private information about his type; he is tough with probability p and weak with probability $1-p$, where $0 < p < 1$. The private information concerns not only the cost of fighting a war (as in Section 2.2) but also how long North is able to fight before he collapses. North's weak type can fight for at most k_W units of time, while the tough type can fight for at most k_T units of time. South is commonly known to be able to fight for at most k_S units of time. Assume $k_W < k_S < k_T$, so South can outlast the weak type, but not the tough type. South's cost of fighting a war is commonly known to be $c > 1$ per unit of time. For simplicity, assume North's weak type has the same cost of fighting as South does, while the tough type's cost of fighting is 0. This assumption is made for the sake of exposition and could be easily generalized to give South and North's weak type different costs from war, and a positive cost to the tough type.

We actually do not need to specify the exact rules of the bargaining that takes place while the war is being fought. But to fix thoughts, it may be helpful to

---

[18] In a more complex model, this assumption could be justified along the lines of Beviá and Corchón (2010): after receiving a share of the resource, a war is less attractive to North because he now has more to lose (as he holds some of the resource) and South has less to take.

consider the following. Time is divided into small "bargaining periods", each lasting Δ units of time. Thus, each unit of time consists of $1/\Delta$ bargaining periods. At the very beginning of the game, at time 0, South proposes to transfer some amount of the resource to North (possibly zero amount, but not less than zero). If North accepts then there is no war, the transfer is made and the game ends. They will derive utility from the amount of the resource they have forever. If North rejects, then there will be war during the first bargaining period, i.e., for Δ units of time (at a cost $c\Delta$ for South and for North's weak type, since their cost is c per unit of time, but at no cost for the tough type). From then on, as long as the war continues, at the beginning of each bargaining period South proposes a (non-negative) transfer to North. If North accepts, the transfer is made, and the war (and the game) ends. (They will keep what they have forever because war cannot be re-started.) If North rejects, then the war continues during this bargaining period (again at a cost $c\Delta$ for South and for North's weak type) and in the next bargaining period the process is repeated.[19] In the terminology of bargaining theory, war is an "inside option" as it does not terminate negotiations. The war goes on (with the participants incurring the costs of war) until a proposal is accepted, with one important exception: if the war lasts for k_W units of time (i.e., for k_W/Δ bargaining periods), then if North is weak he collapses and the game ends (with South keeping all of the resource forever). Similarly, if the war lasts for k_S units of time then South collapses, all of the resource is transferred to North (who keeps it forever) and the game ends. (The tough type would collapse at time k_T, but this will never happen as South cannot keep fighting after time $k_S < k_T$.)

Let δ be the discount factor applied to one unit of time (not to one bargaining period). Assume $\delta < 1$. Owning all of the resource forever is worth $1/(1-\delta)$ since it yields a utility if 1 per unit of time. Because North knows that South can fight for at most k_S units of time, North's tough type's expected payoff in equilibrium must be at least

$$\delta^{k_S}\frac{1}{1-\delta}. \tag{29}$$

Moreover, at any time $\tau < k_S$ the tough type's continuation payoff is at least

$$\delta^{k_S-\tau}\frac{1}{1-\delta}$$

since he will win for sure if he fights until time k_S. Therefore, to get the tough

[19] This is just an example of how events may occur within each bargaining period. Many other protocols would yield the same results, so we need not be very specific.

type to accept an offer at time τ, South must offer him at least a share $\delta^{k_S-\tau}$ of the resource. Discounted back to time 0, the offer is equivalent to a share $\delta^{\tau}\delta^{k_S-\tau}=\delta^{k_S}$.

Fix any sequential equilibrium of the game. Let Ψ denote the probability that, in this equilibrium, North's tough type will fight a war that lasts for at least k_W units of time. In other words, if North is tough then he will accept an offer *before* time k_W with probability $1-\Psi$. The Coase conjecture would imply that Ψ goes to zero as Δ goes to zero. We aim to show this is false. The weak type can imitate the tough type's behavior until time k_W (at which time the weak type must accept any offer, to prevent a collapse). Therefore, the weak type's expected payoff in this equilibrium must be at least

$$(1-\Psi)\left(\frac{\delta^{k_S}}{1-\delta}\right)-\frac{1-\delta^{k_W}}{1-\delta}c. \tag{30}$$

This is true because if the tough type accepts an offer at time $\tau<k_W$, he must have been offered at least a share $\delta^{k_S-\tau}$ of the resource, as explained above. And then the weak type would get the same, if he imitates the tough type. The second term in (30) is just the cost of fighting until time k_W.

Suppose South's strategy is to make no positive offer during the first k_W units of time, but after that he will offer North all of the resource (an offer North would surely accept) if the war still continues. That is, North gets everything once he has proved conclusively that he is tough, which requires fighting for k_W units of time. (Strictly speaking, he should fight for just a little bit longer than that.) Since North's weak type can fight for at most k_W units of time, this strategy guarantees that South will keep all of the resource forever if North is weak (which he is with probability $1-p$). Moreover, it guarantees that a war will last at most k_W units of time even if North is tough. Since this is a feasible strategy, in equilibrium South's expected payoff must be at least

$$(1-p)\frac{\delta^{k_W}}{1-\delta}+\frac{1-\delta^{k_W}}{1-\delta}(1-c). \tag{31}$$

The second term in (31) is just the cost of fighting until time k_W while simultaneously deriving utility from the resource. Recall that $c>1$ so this term is negative.[20]

Since the total surplus available is $1/(1-\delta)$, feasibility requires that p times

[20] Thus, we assume South benefits from the resource during the war, but this assumption is not important.

the expression in (29) plus $1-p$ times the expression in (30) plus the expression in (31) equals at most $1/(1-\delta)$. This implies

$$p\delta^{k_S} + (1-p)(1-\Psi)\delta^{k_S} - (1-p)(1-\delta^{k_W})c$$
$$+(1-p)\delta^{k_W} + (1-\delta^{k_W})(1-c) \le 1\,. \tag{32}$$

This inequality, which must hold at any sequential equilibrium, refutes the Coase conjecture. Indeed, for any $\varepsilon > 0$, if δ is sufficiently close to 1 then (32) forces Ψ to be at least $1-\varepsilon$. This means that if the players are very patient and North is tough, then with probability close to 1 there will be a war which lasts for as long as the weak type is able to fight, i.e., k_W units of time. Since North is tough with probability $p > 0$, the ex ante probability of war is at least $p(1-\varepsilon)$. More importantly, the ex ante expected *duration* of war is at least $p(1-\varepsilon)k_W$ units of time, where ε is close to 0 for δ close to 1. A lower bound on the expected duration of war can be derived from (32) even if δ is not close to 1. The bound will be positive for a wide range of parameters.[21]

Importantly, the lower bound on the expected duration of war derived from (32) is independent of how frequently South can make proposals to terminate the war, i.e., it does not depend on Δ. Indeed, the lower bound is in units of time, and it applies regardless of the bargaining protocol (one player could make take-it-or-leave-it offers, or they could make alternating offers, etc.). Thus, the Coase conjecture is in general not valid for this model. Incomplete information may lead to wars that last a long time, even when the ability to commit to future actions is lacking, and even though there are no impediments to intense negotiations during the war.

The details of the sequential equilibrium strategies will depend, of course, on the details of the bargaining protocol. Here we will just observe that the game in effect has finite length: it can be truncated at time k_S since a war cannot continue after that. We can then make the strategy spaces finite by assuming the resource is only finitely divisible (i.e., there are only finitely many ways in which the resource can be divided up). The game is then a finite sequential game with perfect recall, so a sequential equilibrium exists by a standard result (Kreps and Wilson, 1982).

V. Final Thoughts

Thucydides (1989) claimed that there are three motives for war: greed, fear and honor. Hobbes (1886) elaborated on the three motives as follows:

[21] Naturally, it is not always positive. For example, we can drive it to zero by holding δ fixed and making c very big.

> "So that in the nature of man, we find three principal causes of quarrel. First, competition; secondly, diffidence; thirdly, glory. The first maketh men invade for gain; the second, for safety; and the third, for reputation. The first use violence, to make themselves masters of other men's persons, wives, children, and cattle; the second, to defend them; the third, for trifles..." (Hobbes, 1886, page 64).

This survey emphasized the first motive: greed. In the basic model, North arms itself in order to extract resources from South and declares war if South does not offer sufficient "tribute". But some arms races and wars seem to be due not primarily to a struggle for resources, but rather due to a struggle for security.[22] This brings up Thucydides's second motive for going to war: fear. Rather than discussing it here, we refer to Baliga and Sjöström (2012) for a discussion of some game-theoretic models of conflicts caused by fear.

The models we have discussed here (as well as those discussed in Baliga and Sjöström, 2012) assume players maximize their expected gains. In equilibrium, expectations are consistent. These strong assumptions are a reasonable starting point which serves to clarify the conceptual issues involved. In effect, in the models each player practises *Raison d'état*, a cold calculation of costs and benefits without ideological or moral constraints. In history, Cardinal Richelieu was a master of this practise. His enemy Emperor Ferdinand II believed instead that foreign policy should be based on the Catholic faith, which prevented him from making deals with Protestants, even when it could have been highly advantageous to his empire.[23] In contrast, Richelieu (although himself a Catholic) supported the Protestant King of Sweden when he went to war against the Emperor. The Emperor's fixed moral principles clearly put him at a disadvantage in this struggle. This example illustrates that wars are not always fought in order to maximize (earthly) gains, and it brings up Thucydides's third motive: honor. This may be a residual category for motives that do not .t comfortably in the economist's standard models. The economist's comparative advantage lies in analyzing the cold logic of a Richelieu, not the morality of a Ferdinand II, and naturally our models reflect this.

[22] For example, see Morgenthau's (1967) discussion of World War I (and the arms race that preceded it).

[23] "It would be a great folly for one to try to strengthen a kingdom, which God alone has granted, with means that God hates" (Ferdinand's adviser Lamormaini, quoted in Kissinger, 1994, page 60).

References

Abreu, D. and F. Gul (2000), "Bargaining and Reputation," *Econometrica*, Vol. 68, 85-117.

Acemoglu, D. and J. A. Robinson (2001), "A Theory of Political Transitions," *American Economic Review*, Vol. 91, 938-963.

Baliga, S. and T. Sjöström (2008), "Strategic Ambiguity and Arms Proliferation," *Journal of Political Economy*, Vol. 116, 1023-1058.

______________________ (2012), "The Hobbesian Trap," in Oxford Handbook of the Economics of Peace and Conflict, edited by M. Garfinkel and S. Skaperdas. Oxford University Press.

Bester, H. and K. Wärneryd (2006), "Conflict and the Social Contract," *Scandinavian Journal of Economics*, Vol. 108, 231-249.

Beviá, C. and L. Corchón (2010), "Peace Agreements without Commitment," *Games and Economic Behavior*, Vol. 68, 469-487.

Brito, D. and M. Intriligator (1985), "Conflict, War and Redistribution," *American Political Science Review*, Vol. 79, 943-957.

Chatterjee, K. and L. Samuelson (1988), "Bargaining under Two-Sided Incomplete Information: The Unrestricted Offers Case," *Operations Research*, Vol. 36, 605-618.

Corchón, L. (2007), "The Theory of Contests: A Survey," *Review of Economic Design*, Vol. 11, 69-100.

Fearon, J. (1994), "Domestic Political Audiences and the Escalation of International Disputes," *American Political Science Review*, Vol. 88, 577-592.

__________ (1995), "Rationalist Explanations for War," *International Organization*, Vol. 49, 379-414.

__________ (1996), "Bargaining over Objects that Influence Future Bargaining Power," mimeo, Stanford University.

__________ (2007), "Fighting Rather than Bargaining," mimeo, Stanford University.

Fey, M. and K. Ramsay (2009), "Mechanism Design Goes to War: Peaceful Outcomes with Interdependent and Correlated Types," *Review of Economic Design*, Vol. 13, 233-250.

___________________ (2011), "Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict," *American Journal of Political Science*, Vol. 55, 149-169.

Garfinkel, M. (1990), "Arming as a Strategic Investment in a Cooperative Equilibrium," *American Economic Review*, Vol. 80, 50-68.

Gul, F., H. Sonnenschein, and R. Wilson (1986), "Foundations of Dynamic Monopoly and the Coase Conjecture," *Journal of Economic Theory*, Vol. 39, 155-190.

Heifetz, A. and E. Segev (2005), "Escalation and Delay in Protracted International Conflicts," *Mathematical Social Sciences*, Vol. 49, 17-37.

Hirshleifer, J. (1988), "The Analytics of Continuing Conflict," *Synthese*, Vol. 76, 201-233.

__________ (1991), "The Paradox of Power," *Economics and Politics*, Vol. 3, 177-200.

__________ (1994), "The Dark Side of the Force," *Economic Inquiry*, Vol. 32, 1-10.

Hobbes, T. (1886), Leviathan, Second Edition, London: Ballantyne Press.

Jackson, M. and M. Morelli (2007), "Political Bias and War," *American Economic Review*, Vol. 97, 1353-1373.

________________________ (2009), "Strategic Militarization, Deterrence and War between Nations," *Quarterly Journal of Political Science*, Vol. 4, 279-313.

Kissinger, H. (1994), Diplomacy, New York: Simon and Schuster.

Kreps, D. and R. Wilson (1982), "Sequential Equilibria," *Econometrica*, Vol. 50, 863-894.

Morgenthau, H. (1967), Politics Among Nations: The Struggle for Power and Peace, Fourth Edition, New York: Alfred A. Knopf.

Myerson, R. and M. Satterthwaite (1983), "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, Vol. 29, 265-281.

Neary, H. (1997), "Equilibrium Structure in an Economic Model of Conflict," *Economic Inquiry*, Vol. 35, 480-494.

Powell, R. (1993), "Guns, Butter and Anarchy," *American Political Science Review*, Vol. 87, 115-132.

_______ (1996), "Uncertainty, Shifting Power, and Appeasement," *American Political Science Review*, Vol. 90, 749-764.

_______ (2004a), "The Inefficient Use of Power: Costly Conflict with Complete Information," *American Political Science Review*, Vol. 2, 231-241.

_______ (2004b), "Bargaining and Learning while Fighting," *American Journal of Political Science*, Vol. 48, 344-361.

_______ (2006), "War as a Commitment Problem," *International Organization*, Vol. 60, 169-203.

Schelling, T. (1966), Arms and Influence, New Haven: Yale University Press.

Skaperdas, S. (1992), "Cooperation, Conflict and Power in the Absence of Property Rights," *American Economic Review*, Vol. 82, 720-739.

__________ (1996), "Contest Success Functions," *Economic Theory*, Vol. 7, 283-290.

Skogh, G. and C. Stuart (1982), "A Contractarian Theory of Property Rights and Crime," *Scandinavian Journal of Economics*, Vol. 84, 27-40.

Slantchev, B. (2003), "The Power to Hurt: Costly Conflict with Completely Informed States," *American Political Science Review*, Vol. 97, 123-133.

Smith, A. and A. Stam, (2004), "Bargaining and the Nature of War," *Journal of Conflict Resolution*, Vol. 48, 783-813.

Thucydides (1989), The Peloponnesian War: The Complete Hobbes Translation, Edition 1989, Chicago: University of Chicago Press.

Tullock, G. (1980), "Efficient Rent Seeking," In Toward a Theory of the Rent-seeking Society, edited by J. M. Buchanan, R. D. Tollison and G. Tullock. College Station: Texas A&M Press.

Appendix

1. Proof of Proposition 1

If $t > 0$ is given by (4) then North's payoff is

$$x_N - g_N + t = \frac{g_N}{g_S + g_N}(x_S + x_N) - g_N - c_N .$$

The derivative of this expression with respect to g_N is

$$\frac{g_S}{(g_S + g_N)^2}(x_S + x_N) - 1 \qquad (33)$$

which, after substituting for g_S from (7), can be shown to be positive for all $g_N \le x_N$ as long as (5) holds. Therefore, if North sets $g_N > 0$ he should set $g_N = x_N$. Moreover, (6) implies that this is better for North that $g_N = 0$. So $g_N = x_N$ is North's best response to (7).

Now, consider South's best response to $g_N = x_N$. Clearly, it is useless to set g_S so high that (3) holds strictly when $t = 0$. Thus,

$$g_S \le \frac{x_N(x_S - c_N)}{c_N + x_N}$$

and the transfer will be given by (4). Since $g_N = x_N$, South's payoff will be

$$b_S - t = \frac{g_S}{g_S + x_N}(x_S + x_N) - g_S + c_N . \qquad (34)$$

This expression is maximized with respect to g_S when (7) holds. The transfer is positive because

$$t = \sqrt{x_N(x_S + x_N)} - x_N - c_N > 0$$

by (6). Thus, (7) is South's best response to $g_N = x_N$.

2. Proof of Proposition 2

It is easy to verify that under our assumptions, (11) and (14) imply $0 < q < 1$

and $0 < \alpha < 1$. Next, we need to check that each player plays optimally. First, suppose South observes $g_N = x_N$. He thinks North is tough with probability $\frac{p}{p+(1-p)q}$. Since South randomizes in equilibrium, he must be indifferent between appeasement and no appeasement. If South chooses appeasement, he pays t given by (13) which, by the same argument as before, is the minimum transfer required to persuade the tough type not to declare war. South's payoff is

$$b_S - t = \frac{g_S}{g_S + x_N}(x_S - g_S) + c_N . \tag{35}$$

If there is no appeasement, then either North is tough, declares a war, and S gets $\frac{g_S}{g_S+x_N}(x_S - g_S) - c_S$; or else North is weak and South gets $x_S - g_S$. So South's expected payoff is

$$\frac{p}{p+(1-p)q} \times \left[\frac{g_S}{g_S + x_N}(x_S - g_S) - c_S \right] + \frac{(1-p)q}{p+(1-p)q}(x_S - g_S) . \tag{36}$$

Using (11) and (12), we find that (35) equals (36). Thus, South is indeed indifferent between appeasement and no appeasement. For his choice of g_S, the argument is the same as in the proof of Proposition 1.

Next, consider the weak type of North. If he sets $g_N = 0$, his payoff is x_N. If he sets $g_N = x_N$, he receives t with probability α, and 0 with probability $1-\alpha$. His expected payoff is αt. Now (12), (13) and (14) imply that $\alpha t = x_N$. Thus, the weak type is indifferent between $g_N = 0$ and $g_N = x_N$. This implies that the tough type certainly prefers $g_N = x_N$.

Finally, if North chooses anything else than $g_N = 0$ or $g_N = x_N$, we may assume South believes North is weak and therefore South offers no transfer. Given this, nothing except $g_N = 0$ or $g_N = x_N$ can be an optimal choice for North.