

Contests over Political Authority*

Catherine Hafer[†]

11/02/2006

Abstract

This paper analyzes a model of standoff in the contest for political authority. Two players have a mix of common and contrary interests; the resolution of the dispute must be self-sustaining, i.e. there is no external enforcement of agreements; and the players are uncertain about each other's resolve, i.e., about the relative strength of their interests in one outcome over another. The equilibrium solution of the model provides insights into the length of the standoff, its ultimate outcome, and the efficiency of the outcome itself. I show that there exists no informative equilibrium in the unmediated cheap-talk extension of the standoff game, and that, in the case with mediation, only non-stationary incentive-compatible mechanisms can exist and only under a restricted set of explicitly characterized conditions. Even when transfers are possible, stationary pure-strategy ICDMs do not exist.

*Comments Welcome. First draft: 12/06/2004. The preliminary draft of this paper was written when I was a visitor at CMS-EMS at the Kellogg School of Business, Northwestern University. I benefitted from comments from participants in the political economy seminars at Kellogg and at Stanford GSB, the Wednesday seminar at NYU, and at the Princeton Conference on Political Institutions and Economic Policy, and especially from David Austen-Smith, David Baron, Bruce Bueno de Mesquita, Ethan Bueno de Mesquita, Ernesto Dal Bó, Eric Dickson, Avinash Dixit, Tim Feddersen, Sven Feldmann, Sandy Gordon, Bard Harsted, Dimitri Landa, and Adam Meirowitz.

[†]Assist. Prof. of Politics, NYU, e-mail: catherine.hafer@nyu.edu

1 Introduction

This paper analyzes a model of contested political authority: a situation in which two disputants who wish to agree ultimately on a single (joint) course of action each insist on different courses of action, in spite of the fact that doing so is immediately costly, in the hopes that their opponent will give in first. The model captures a number of features intrinsic to such situations: the players have some degree of common interest, so that they may benefit from coordinating their actions (peaceful settlement is, all else equal, better for each player); they have some degree of contrary interests (each player prefers to be the source of authority given the benefits entailed), so that the dispute between them is not trivial; and the players are uncertain about each other's resolve, i.e., about the relative strength of their interests in one outcome over another. Finally, because the contest is for the prize of political authority, its resolution must be self-sustaining, i.e. there is no external enforcement of agreements.

The first two features - a mix of common and contrary interests - are ubiquitous, and characterize both political disputes, in which people with different policy preferences benefit from coordinating on one public policy, and common economic ones, in which two firms that may profit from trade have opposing preferences over the division of such profits between them. In the case of two firms that wish to do business, many divisions of the profits are possible, in part because the firms can typically rely on some form of third-party contract enforcement to insure that they abide ex post by whatever agreement they make ex ante. However, when the contest in question is for political authority, we cannot assume that the agreement - if it were reached - would be enforced by an already existing common authority, and so the feasible final resolutions must include only those agreements to which both parties could credibly commit to carrying out. The requirement that final resolutions be credible in this way implies a considerable limitation on the set of viable agreements, ruling out a variety of arrangements that otherwise might have been possible.

The requirement that the settlement of authority be self-enforcing also implies a constraint on the analytical model that could be used to capture the underlying distributive game. In particular, it rules out the class of standard bargaining models, which, in effect, assume perfect enforcement of the negotiated outcomes. Indeed, one of the key findings is that in the contest over political authority there may not be enough flexibility - enough room for given and take - to achieve a resolution of the dispute through bargaining. Rather, the parties may find themselves in a costly standoff, each insisting (at least initially) on a different outcome and hoping that the other will be the first to give in.

The conjunction of these factors alone is not sufficient to produce a standoff, however, because if it is common knowledge between the disputants which one of them will ultimately win, the other will concede immediately. Thus a standoff necessarily

requires that the participants have some ongoing uncertainty about which of them is willing to hold out longer for her more-preferred outcome, because each of them must have some hope of winning in order to continue the dispute. A model of standoff should, then, incorporate both the possibility of conflict under uncertainty and the possibility of immediate resolution when that uncertainty is removed.

The model presented in this paper is based on a dynamic extension of an incomplete-information asymmetric coordination game. This game captures both the common and the contrary aspects of the interests of parties contesting political authority. After introducing this game, I examine the equilibria that incorporate standoff, e.g. each player initially insisting upon its own most preferred outcome by taking the action that corresponds to it. The game thus restricted is strategically equivalent to an incomplete-information war of attrition, and I exploit this strategic equivalence to obtain the explicit solution for the unique symmetric equilibrium. Because equilibrium play ultimately reveals the type of the less resolute player, and produces the long-term outcome preferred by the more resolute player, neither player has an incentive to rekindle the dispute—the resolution of the dispute is self-sustaining. The equilibrium solution of the model provides insights into the length of the dispute, its ultimate outcome, and the efficiency of the outcome itself (apart from the costs of achieving it).

An additional advantage of the model is that, like all war-of-attrition games, the complete-information version has an asymmetric equilibrium in which the weaker player concedes immediately to the stronger one. I exploit this fact to examine the extent to which mediated and unmediated cheap-talk communication can be used successfully to end or to avert a standoff as a function of features of the strategic environment.

The key results of the model are, on the whole, negative. I show that unmediated cheap-talk communication does not lead to an improvement in social welfare. Of the possible mechanisms without transfers, only non-stationary flat mechanisms are sometimes incentive-compatible. Mechanisms with transferable utility can be welfare-enhancing, but such mechanisms are sustainable only over limited time: parties to the agreement eventually become unwilling to honor it.

The paper proceeds as follows. In Section 2, I provide a brief discussion of the relation between the model and the existing literature. Section 3.1 presents the symmetric equilibrium of the basic model of standoff without communication. Section 3.2 analyzes the possibility of welfare-enhancing unmediated cheap-talk. Section 3.3 moves to the analysis of mediation without transfers, and Section 3.4 considers the effects of allowing transfers. Section 4 concludes with a brief discussion.

2 Relation to Literature

The model presented below can be seen as a natural complement to the model of political compromise by Dixit, Grossman, and Gul (2000).¹ In the latter, the party in power in a given period is assumed to determine unilaterally the division of benefits between the two parties, and the focus is on the dynamic evolution of the allocation of spoils between the in- and out-of-power parties. In particular, Dixit, Grossman, and Gul ask what arrangements for sharing the political and economic benefits can be supported in equilibrium, and how those arrangements depend on exogenously given probabilities of each party obtaining and retaining office. In contrast, in the model presented below the likelihood of each party's victory is determined by the parties' choices, which depend in equilibrium on their privately known valuations of the possible outcomes, and in each period the outcome is jointly determined by the parties' actions. By assumption, neither party can unilaterally obtain its most preferred outcome for even one period, and preventing an opponent from obtaining their most preferred outcome imposes lower payoffs on both parties, in essence destroying part of the pie. In this model, the focus is on establishing political authority in an environment in which it is initially absent, by force if necessary, rather than on managing its exploitation where it is already well established and can be peacefully transferred from party to party.

The model in this paper is related to two branches of the recent work on appropriative conflict. In the economics literature, Dixit (2004) presents a recent state of the art discussion of the approaches to analyzing economic activity in a decentralized legal and political environment. Hafer (2006) is a model of the state of nature in which parties engage in dynamic continuous-time wars of attrition over a productive resource. In political science, the closely related work is on crisis bargaining games (e.g., Banks 1990, Bueno de Mesquita, Morrow, and Zorick 1997, Powell 2004), in which parties attempt to negotiate a resolution to a conflict under the threat of either party's unilateral initiation of a war, and one or both parties are uncertain about the payoffs associated with war.

In the model presented here, the parties are not only uncertain about the value of going to war (because of their uncertainty about their opponenets' types), they are also uncertain of their opponents' valuation of the settlement. Both the payoffs to war and the payoffs associated with a given settlement depend on aspects of the player's type, though in different ways. This fact has important implications for the viability of a negotiated long-term resolution in which one player is asked to accept the other's authority, because a player's continued willingness to accept the authority of the other player (in spite of its own changing beliefs about its opponent's type) reveals information about its own type in turn. Furthermore, the fact of that information revelation and its potential consequences for subsequent renegotiation or conflict may

¹For a related model, see also Alesina (1988).

limit the players' ability to reach a negotiated settlement (Nalebuff 1987). In this sense, peace under a negotiated settlement is an informative process, just as conflict is.

Prominent political economy applications of incomplete information war of attrition include models of delayed fiscal stabilization (Alesina and Drazen 1991, Drazen and Grilli 1993, Casella and Eichengreen 1996, Spolaore 2004). Relative to these models, the basic model in Section 3.1 generalizes to allow private information over all payoffs, retains the potential for selecting inefficient long-run outcome and shows how global changes in cost of conflict may change the outcome of conflict. The results in the subsequent sections may be interpreted as identifying the extent to which these inefficiencies survive in the presence of cheap-talk communication - a common feature of the empirical settings relevant to these models.

The analysis of the cheap-talk extension of the basic model constitutes a contribution to the literature on cheap-talk games with uncertainty over payoffs, which is mostly of recent vintage (see e.g., Ben-Porath 2002 and Baliga and Morris 1998, 2002). Baliga and Morris (1998, 2002) consider the conditions that guarantee the impossibility of information transmission that would be equilibrium payoff-relevant in the subsequent game. They show that if the game satisfies two general conditions, then, in any equilibrium, all types of each player are indifferent between any of the equilibrium messages sent by any type of that player (though such equilibria may still be informative). These conditions are *type independence* and *common induced preferences*. A game has common induced preferences if, for any pair of a player's possible types, the two types have the same preferences over any pair of distributions of the player's opponent's actions. The game considered above satisfies type independence and violates common induced preferences. While all types prefer the opponent to drop out sooner rather than later, not all types will necessarily have the same preferences over a pair of distributions of stopping times such that in one, both early and late stopping times are more likely than they are in the other.

The results can be instructively compared to those of Banks and Calvert (1992), who ask whether efficient outcomes can be obtained through mediated or unmediated cheap talk in a static incomplete-information battle-of-the-sexes game. They find that, while mediated cheap talk can improve efficiency, unmediated cheap talk cannot. I find that unmediated cheap talk is also fruitless in the dynamic game and, relative to their results for the static game, efficiency gains through mediation are more difficult to sustain. Banks and Calvert also show that, for a one-shot (discrete-time) incomplete-information Battle of the Sexes, there is always an ICDM that is ex ante efficiency improving compared to the symmetric equilibrium of the game without communication. Baliga and Sjostrom (2004) obtain a non-monotonic cheap-talk equilibrium strategy in a game with some common interest. Neither of these results survives in the model analyzed here.

3 The Model

Suppose that two players face an asymmetric coordination problem in continuous time with incomplete information. In particular, suppose that each player has two actions, $\{X, Y\}$, and at every point in time each player must take one of these actions. The matrix below shows the payoff per unit of time for each player from each of the four possible action profiles. Suppose that it is common knowledge that player 1 (weakly) prefers outcome (X, X) to (Y, Y) and strictly prefers (Y, Y) to either (X, Y) or (Y, X) . Player 2 prefers (Y, Y) to (X, X) , but she also prefers either of these outcomes to either of the others. Suppose, however, that each player is uncertain of the precise payoffs of her opponent (a_i, b_i) , and that they have a common prior over the distribution of types. The payoff matrix corresponds to that of an incomplete information battle-of-the-sexes game, but with the caveat that, because the game takes place in continuous time, the payoffs it shows are flows rather than lump sums. Each player's payoff for the entire infinite-horizon game can be evaluated at any moment in time as the discounted present value of all future payoffs, given the common discount rate r .

	Player 2	
	X	Y
Player 1	X	Y
X	a_1, b_2	$0, 0$
Y	$0, 0$	b_1, a_2

$$a_i \geq b_i > 0, i \in \{1, 2\}$$

This dynamic extension of the battle-of-the-sexes game can be understood to entail a model of standoff. Suppose that at the beginning of the game, each player must choose an action, X or Y . Actions are observable, and at any moment, either player can switch to the other action. If each player begins by taking the action that corresponds to her own preferred outcome, i.e. 1 chooses action X and 2 chooses action Y , then each player's strategy can be described by the time at which she switches, conceding victory to her opponent. (If a player chooses the action that corresponds to the other player's preferred outcome at the start, then she can be said to "switch" at time $t = 0$.) At each moment in time, a player obtains additional information about her opponent's type from the fact that her opponent has not yet conceded. Described in these terms, the possible actions and the players' beliefs are the same as in an incomplete-information war of attrition. As I will demonstrate below, the payoffs of the dynamic battle of the sexes can be expressed as payoffs to a war of attrition.

The players' utilities can be expressed as the discounted present values of their future streams of payoffs. If player i quits at t_1 , while player j stays in, the discounted present value of player i 's payoff at $t = 0$ is

$$u_i(t_1, t_2 > t_1) = \int_{t_1}^{\infty} b_i e^{-rt} dt = \frac{b_i}{r} e^{-rt_1}.$$

Similarly, if player j quits at t_2 , while player i stays in, the discounted present value of player i 's payoff at $t = 0$ is

$$u_i(t_1, t_2 < t_1) = \frac{a_i}{r} e^{-rt_2}.$$

In terms of the war of attrition, the value of the prize (won at time 0) is the difference between the present value of the preferred equilibrium and of the opponent's preferred equilibrium, $(a_i - b_i) \int_0^\infty e^{-rt} dt$. The cost of spending another unit of time resisting is the difference between the payoff for that unit of time in discordant play and the payoff for that unit in the opponent's preferred equilibrium, b_i .

3.1 Pure Standoff

The solution concept is Perfect Bayesian Equilibrium, which requires that at each moment in time, a player choose optimally to switch to the other action or to continue, given her beliefs; and that at each moment in time, she update her beliefs about her opponent's type via Bayes' Rule based on the fact that her opponent has not yet conceded.

The first result provides a characterization of the symmetric equilibrium outcome:²

Theorem 1 *The standoff game has a unique symmetric equilibrium which*

(1) *selects the preferred outcome of the player i with the higher value of the ratio $\theta_i := \frac{a_i}{b_i}$*

(2) *has a standoff of duration $s^*(\theta_i) = \frac{1}{r} \int_{\theta_i}^{\theta_i} (\theta - 1) \frac{p(\theta)}{1-P(\theta)} d\theta$.*

Proof. Letting $p_j(t)$ be the probability density of opponent j 's quitting time, the expected payoff for player i from action t_i is

$$\begin{aligned} E[u_i(t_i)] &= (1 - \Pr(t_j \leq t_i)) \frac{b_i}{r} e^{-rt_i} + \int_0^{t_i} p_j(t) \frac{a_i}{r} e^{-rt} dt \\ &= \frac{b_i}{r} e^{-rt_i} (1 - \Pr(t_j \leq t_i)) + \frac{a_i}{r} \int_0^{t_i} p_j(t) e^{-rt} dt. \end{aligned} \quad (1)$$

²Note that, as in all incomplete-information wars of attrition, there are also two asymmetric equilibria: player 1 "fights forever" and player 2 "quits immediately;" and 1 "quits immediately" and 2 "fights forever." Recall that, because the players' types are *not* common knowledge, their types cannot be invoked to select one equilibrium over the other.

Then the first-order condition is

$$-p_j(t_i) \frac{b_i}{r} e^{-rt_i} + (1 - \Pr(t_j \leq t_i))(-b_i e^{-rt_i}) + p_j(t_i) \frac{a_i}{r} e^{-rt_i} = 0$$

In order to solve the first-order condition, we must express $\Pr(t_j \leq t_i)$ in terms of primitives, e.g. as a probability of a type (a_j, b_j) that chooses to quit before time t_i . This requires being able to identify a type that corresponds to t , but because type is two-dimensional, multiple types may (and, in equilibrium, do) choose the same stopping time t .³ This problem can be surmounted by re-expressing payoffs in terms of the ratio of the difference between the value of the prize and the opportunity cost of continuing the conflict, $\frac{a}{b}$. This is algebraically equivalent to dividing the first order condition by b . Let θ represent $\frac{a}{b}$. Figure 1 shows the level curves of θ in the (b, a) space. We solve for players' optimal stopping times as a function of θ .⁴

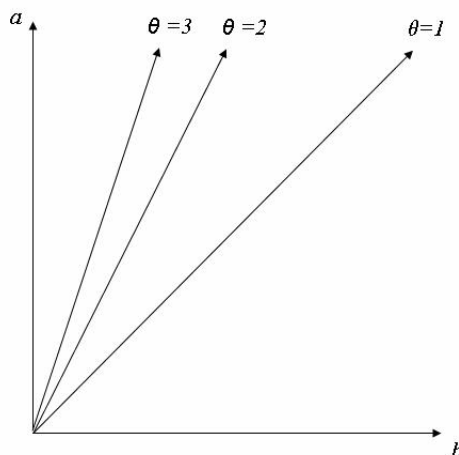


Figure 1: Level curves of θ .

To solve for i 's best response, we must be able to express the opponent's type as a function of her action, i.e. $\Theta(t_j)$. Such a function exists only if her optimal choice

³In the game with non-transferable utility, one can normalize b_1 and b_2 without loss of generality, reducing players' types to one dimension. However, later results consider environments in which transfers are possible, requiring that the players' utilities be expressed in terms of a common scale and thus requiring two-dimensional types to maintain generality.

⁴Note that once the problem is expressed in terms of θ , it constitutes a variation of the incomplete information war of attrition in Fudenberg and Tirole (1991, pp. 216-19) with discounting, and it can be solved in the same way.

of action in equilibrium is strictly monotonic. To see that it is weakly monotonic, recall that, from the definition of Bayesian Nash equilibrium, the equilibrium action of type θ of player i always yields at least as high a payoff for that type of that player than does the equilibrium action of some other type. Thus, if t'_i is the equilibrium action of an agent of type (a'_i, b'_i) , and hence of type θ'_i , and if t''_i is the equilibrium action of an agent of type (a''_i, b''_i) , and hence of type θ''_i , we have the following two inequalities:

$$\begin{aligned} & \frac{b'_i}{r} e^{rt'_i} (1 - \Pr(t_j < t'_i)) + \frac{a'_i}{r} \int_0^{t'_i} p_j(t) e^{-rt} dt \\ & \geq \frac{b'_i}{r} e^{rt''_i} (1 - \Pr(t_j < t''_i)) + \frac{a'_i}{r} \int_0^{t''_i} p_j(t) e^{-rt} dt; \\ & \frac{b''_i}{r} e^{rt''_i} (1 - \Pr(t_j < t''_i)) + \frac{a''_i}{r} \int_0^{t''_i} p_j(t) e^{-rt} dt \\ & \geq \frac{b''_i}{r} e^{rt'_i} (1 - \Pr(t_j < t'_i)) + \frac{a''_i}{r} \int_0^{t'_i} p_j(t) e^{-rt} dt. \end{aligned}$$

Multiplying each inequality by r and dividing by b'_i and b''_i respectively, these inequalities can be expressed in terms of θ'_i and θ''_i , respectively. Because these inequalities are of the same sign, the sum of their right-hand sides must be greater than the sum of the left-hand sides. Collecting terms and reducing, we have, then, the following true inequality:

$$\left(\frac{a'}{b'} - \frac{a''}{b''} \right) \int_0^{t'_i} p_j(t) e^{-rt} dt \geq \left(\frac{a'}{b'} - \frac{a''}{b''} \right) \int_0^{t''_i} p_j(t) e^{-rt} dt.$$

It follows, then, that if $\frac{a'}{b'} > \frac{a''}{b''}$, then $t' \geq t''$. To see that the equilibrium strategy must be strictly increasing and continuous in $\frac{a}{b}$, suppose instead that there is an interval of values of $\frac{a}{b}$ that choose to stop at time t , i.e. that $\Pr(t_j = t) > 0$. Then there must be some $\varepsilon > 0$ such that $\Pr(t_i \in (t - \varepsilon, t)) = 0$. But then $t - \varepsilon$ strictly dominates t for player j . Hence the equilibrium strategy must be strictly increasing in type $\theta := \frac{a}{b}$. Exploiting the symmetry of the game (i.e. the common prior), part 1 is established.

Let $s(\theta)$ be time chosen by type θ , i.e., θ 's strategy. Because $s(\theta)$ is continuous and strictly increasing, the inverse of $s(\theta)$, $\Theta(s)$, exists and is continuous. The

optimization problem may then be written

$$s_i \in \arg \max \left[\frac{1}{r} \left((1 - P(\Theta_j(s_i)))e^{-rs_i} + \frac{a_i}{b_i} \int_0^{s_i} p(\Theta_j(s)) \frac{\partial \Theta_j(s)}{\partial s} e^{-rs} ds \right) \right],$$

where $\Theta_j(s_i)$ is the type of the opponent that would play s_i in equilibrium, and $P(\theta)$ and $p(\theta)$ are the cumulative distribution and the density, respectively, of the opponent's type θ . Replacing $\frac{a_i}{b_i}$ with θ_i , the first-order condition is

$$-p_j(\Theta_j(s_i)) \frac{\partial \Theta_j(s_i)}{\partial s} \frac{1}{r} e^{-rs_i} - (1 - P_j(\Theta_j(s_i)))e^{-rs_i} + p_j(\Theta_j(s_i)) \frac{\partial \Theta_j(s_i)}{\partial s} \frac{\theta_i}{r} e^{-rs_i} = 0.$$

Collecting terms and cancelling, we get

$$p_j(\Theta_j(s_i)) \frac{\partial \Theta_j(s_i)}{\partial s} \frac{\theta_i - 1}{r} = (1 - P_j(\Theta_j(s_i))).$$

Exploiting symmetry and isolating $\frac{\partial s(\theta_i)}{\partial \theta}$ yields $\frac{\partial s(\theta_i)}{\partial \theta} = \frac{p(\theta_i)}{1 - P(\theta_i)} \frac{\theta_i - 1}{r}$. Integrating, we obtain a family of solutions that vary only with respect to the constant of integration, k :

$$s^*(\theta_i) = \frac{1}{r} \int_{\theta_i}^{\theta_i} (\theta - 1) \frac{p(\theta)}{1 - P(\theta)} d\theta + k.$$

Given strict monotonicity, if $s(\theta_l)$ were strictly positive, then θ_l could increase her utility by unilaterally switching to $s = 0$. Therefore, $s(\theta_l) = 0$ in equilibrium and, thus, $k = 0$. Hence the equilibrium strategy is unique. ■

The argument of the proof proceeds by exploiting the possibility of collapsing the two dimensions of players onto a single dimension induced by the ratio of the benefit of winning over the opportunity cost of continuing the standoff, θ . The symmetric equilibrium of the standoff game can, then, be characterized as the symmetric equilibrium of the incomplete-information continuous-time war of attrition between opponents defined by their type θ_i .

The model of standoff predicts that the player who has a higher value on that dimension will ultimately prevail, and the coordination outcome that she prefers will be the long-run outcome. Once a player has conceded, neither player has an incentive to renew the dispute by changing her choice of action again; the player who caved has revealed her type and, more to the point, she has revealed that she is of a less resolute type than the winner. Her optimal action, then, is to continue playing the action corresponding to the victor's preferred outcome, and the resolution of the dispute is final.

Because the equilibrium solution of the model of standoff identifies just one of the action profiles in the underlying battle of the sexes as the long-run equilibrium behavior, it could be used as an equilibrium selection device in battle-of-the-sexes games. As such, it would dictate that we focus on the outcome most preferred by the player with higher θ .

3.2 Standoff and Unmediated Cheap-talk

Recall that the players' uncertainty about each other's types was a necessary ingredient of standoffs. We may wonder, then, if allowing the players to communicate can reduce their uncertainty and affect the duration or occurrence of a standoff. In this subsection I analyze the effects of introducing the possibility of unmediated cheap talk on equilibrium play in the standoff game analyzed above. In *the cheap-talk extension of the game*, players i and j can simultaneously send and observe messages m_i and m_j , respectively, at any point in the standoff game. The messages themselves have no direct effect on the players' utilities, although they may, in principle, have an indirect effect via changes in their opponent's beliefs and behavior.

As before, I assume for the cheap-talk extension of the game that, when it becomes common knowledge that $\theta_i > \theta_j$, the players play (as of that point in time) the subgame-perfect Nash equilibrium in which j 's strategy is "quit immediately" and i 's is "fight forever." Because players differ in both their valuation of the prize and in their costs of competing for it, the relevant characteristic of the players is θ , as discussed above.

The next theorem shows that unmediated cheap-talk in the cheap-talk extension of the game of standoff yields no efficiency gains. I proceed by showing that this holds with respect to the possibility, first, of *monotonic* and, then, of *non-monotonic* informative equilibria. An equilibrium is monotonic if the agents' cheap-talk strategies are monotonic in their type, i.e., for an n -message equilibrium, there exist $n - 1$ cutpoints in the type space such that no two types that are separated by any cutpoint prefer to send the same message. An equilibrium is non-monotonic if agents' cheap-talk strategies are not monotonic in their type.

Theorem 2 *There exists no informative equilibrium in the unmediated cheap-talk extension of the standoff game.*

Proof. To prove the theorem, I first show that its claim holds for the cheap-talk extension of the standoff game in which communication occurs before playing the standoff game, and then argue that the result extends if we allow for the possibility of communication at any point in that game.

Consider first the possibility of a monotonic informative equilibrium in the pre-play cheap-talk extension of the standoff game without mediation. Because the existence of such equilibrium with more than two messages implies the existence of one

with only two messages, to prove the impossibility of informative monotonic equilibria, it is sufficient to show that there exists no monotonic two-message equilibrium. Consider then a partition of the type space by $\hat{\theta}$, such that $\forall \theta < \hat{\theta}$, $m = m^1$ and $\forall \theta > \hat{\theta}$, $m = m^2$. Suppose that, if $m_i > m_j$, then j quits immediately, and $u_i(\cdot) = \frac{a_i}{r}$ and $u_j(\cdot) = \frac{b_j}{r}$. If $m_i = m_j$, then they each update their beliefs and play the war of attrition as analyzed in the previous subsection.

This symmetric strategy profile cannot be an equilibrium because $\theta < \hat{\theta}$ prefers to defect to m^2 . To see that this is so, we will first establish that, if her opponent (who is using this strategy) sends m^2 , then she will choose $s = 0$ in the subsequent war of attrition. Suppose $\theta_2 > \hat{\theta}$ and $\theta_1 < \hat{\theta}$, and both play m^2 . (Assume the informative equilibrium but player 1 is deviating in the cheap-talk stage). Player 1 wants to choose t_1 such that

$$-(1 - P(s^{-1}(t_1))) + p(s^{-1}(t_1)) \frac{\partial s^{-1}(t_1)}{\partial t} \frac{1}{r} (\theta_1 - 1) = 0.$$

Recall that $s(\theta|(m^2, m^2))$ is such that $s(\hat{\theta}|(m^2, m^2)) = 0$. Hence,

$$p(s^{-1}(0)) \frac{\partial s^{-1}(0)}{\partial t} \frac{1}{r} (\hat{\theta} - 1) = 1 - P(s^{-1}(0)).$$

Note that the left-hand side is increasing in θ , while the right-hand side is constant. It follows that if $\hat{\theta}$ is not getting a sufficient expected benefit from staying in, then $\theta < \hat{\theta}$ is not either. Hence, $\forall \theta < \hat{\theta}$, $s(\theta|(m^2, m^2)) = 0$ and $E[u_1(0)] = \frac{b_1}{r}$.

Note that if $\theta_i < \hat{\theta}$ sends m^1 and her opponent sends m^2 , then θ_i quits immediately and obtains the same payoff, $\frac{b_1}{r}$. It follows that which message yields higher expected utility depends entirely on the payoffs obtained when the opponent sends m^1 , i.e., when she is also of a type $\theta < \hat{\theta}$. According to the proposed equilibrium, if i sends message m^2 and her opponent sends m^1 , then her opponent quits immediately and i obtains her highest possible payoff $\frac{a_1}{r}$. On the other hand, if she were to send m^1 , she would then have to play asymmetric war of attrition with a lower expected payoff. Hence, she will always want to choose m^2 , and so there is no monotonic informative equilibrium.

Consider next the possibility of a two-message non-monotonic equilibrium. If such an equilibrium exists, then, for some θ', θ'' such that $\theta' < \theta''$, all $\theta < \theta'$ and all $\theta > \theta''$ send message m^1 , and all $\theta \in (\theta', \theta'')$ send message m^2 . Suppose, without loss of generality, that $m_1 = m^1$ and $m_2 = m^2$. Because strategy is increasing in θ for each player, and because each opponent must be expected to quit with positive probability at every time $s < s(\theta^h)$, there must be a time

$$\hat{s} \equiv \lim_{\theta_1 \rightarrow -\theta'} s(\theta_1|m_1 = m^1) = \lim_{\theta_1 \rightarrow +\theta''} s(\theta_1|m_1 = m^1)$$

such that $\theta_1 < \theta'$ if and only if $s(\theta_1|m^1) < \hat{s}$ and $\theta_1 > \theta''$ if and only if $s(\theta_1|m^1) > \hat{s}$. Then, if player 1 continues at time \hat{s} , it becomes common knowledge that $\theta_1 > \theta'' > \theta_2$ and player 2 quits. But then the type θ' strictly prefers m^1 to m^2 , and θ'' does also, which yields a contradiction.

Now consider the possibility of a non-monotonic informative equilibrium with any number of messages. For any symmetric nonmonotonic messaging strategy $m(\theta)$, there exists a pair of distinct messages m^1, m^2 and distinct types θ', θ'' that satisfy the following conditions: 1) $\theta' < \theta''$; 2) $s(\theta'|m^1) = s(\theta''|m^1)$; 3) there exists a $\theta \in \{\theta|m(\theta) = m^2\}$ such that $\theta > \theta'$; 4) for every $\theta \in \{\theta|m(\theta) = m^2\}$, $\theta < \theta''$. But then the corresponding partial strategy profile is strategically equivalent to the two-message non-monotonic strategy profile, and, by the same argument, cannot be part of an equilibrium strategy profile.

Finally, to see that allowing symmetric unmediated cheap-talk at any other point in the game will not admit informative communication in equilibrium, first note that, correcting for beliefs, the game is strategically equivalent at any point in time provided that neither player has conceded. Since the proofs rely only on the continuity of $P(\theta)$ and the finiteness of its support, both of which are preserved as the game progresses, informative unmediated cheap talk is not possible at any time. ■

The key intuition for this theorem is that all types always wish to appear to be as high a type (as resolute in the standoff) as possible, in order to convince their opponent to quit sooner. Because each player can quit at any time, the costs of initially failing to coordinate can be made arbitrarily small by either player, acting unilaterally. Hence the benefits associated with the possibility of getting one's more-preferred outcome by claiming to be a high type always outweigh the benefits of immediate coordination.

3.3 Mediated Settlements without Transfers

Having established that cheap talk alone cannot affect the occurrence or duration of standoffs, the natural next step is to consider mediation as a potential means of shortening or eliminating them. Consider augmenting the game with a pre-play communication stage in which each player communicates privately with a mediator, who recommends a course of action. Assume that all messages, e.g. claims about one's type, are unverifiable, and that the adoption of the mediator's recommended course of action is strictly voluntary. The main result—that mediation is an ineffective means of eliminating conflict and thereby achieving more efficient outcomes—is obtained by application of the Revelation Principle (Myerson 1985), which states that any outcome associated with equilibrium behavior in any such augmentation of a game is also associated with some *incentive-compatible direct mechanism* (ICDM) based on the same game. A *direct mechanism* consists of a communication protocol and a decision rule for the mediator. The pre-play communication stage has the following form: first, each player sends a private message to the mediator, who is an impartial

non-strategic actor, and the set of messages available to each player is identical to the set of possible types. The mediator then privately recommends a strategy to each player, based on their messages and the mediator’s commonly known decision rule. A mechanism is said to be *incentive-compatible* if truthfully revealing their types (to the mediator) and adopting the recommended strategies constitutes equilibrium behavior.

First, is it possible for the mediator to establish one of the long-run outcomes that might result from a standoff, namely both players choosing X , or both of them choosing Y , indefinitely? The advantage of mediation would then be the avoidance of the costs of the actual standoff, during which both players receive their lowest payoffs. Let a *stationary* ICDM be an ICDM in which the recommended strategies are stationary, i.e. each player is to take the same action, X or Y , at every point in time, independent of the history of play and of calendar time. If we continue to restrict attention to the symmetric equilibrium in the incomplete information war of attrition, then:

Theorem 3 *There is no pure-strategy stationary incentive-compatible mechanism.*

Proof. Applying the Revelation Principle, it is sufficient to show that there is no pure-strategy stationary ICDM. Call a mechanism in which the mediator disregards the messages and randomizes between telling both players “ X ” and telling them “ Y ” a *flat mechanism*. First, I show that no flat mechanism is incentive-compatible. Because the mediator’s messages to the players are independent of their messages to the mediator, and because the mediator has no other private information about the players’ types (since messages are unverifiable), the mediator’s messages do not alter the players’ beliefs about the other’s type. Because i ’s expected discounted present value of the war of attrition is at least as great as $\frac{b_i}{r}$, there is always at least one player—the one who is advised to take the action associated with her less-preferred outcome—will not voluntarily adopt the suggested behavior unless either 1) she is of a type θ s.t. $s(\theta) = 0$; or 2) we assume an asymmetric, rather than a symmetric, equilibrium in any subsequent incomplete-information war of attrition, i.e. we assume that a specified player “holds out” forever and that the other player “gives in” immediately, independent of their types.

Consider, then, the remaining possibility of a *responsive mechanism* - i.e., a mechanism that is responsive to the players’ messages. If the mechanism is deterministic, then the players will be able to infer from the mediator’s advice which player is of the higher type, and the actions that correspond to that player’s preferred outcome will be an equilibrium. But then every type wants to mimic the highest possible type, and the mechanism is not incentive compatible. So a responsive ICDM must be probabilistic. But if it is not common knowledge which type is higher, then, even though the mediator’s advice may convey some information that changes the players’ beliefs, it is still the case that i ’s expected discounted present value of the war of attrition is bounded below by $\frac{b_i}{r}$, and thus the player who is advised to take the

action associated with her less-preferred outcome will prefer to deviate if $s(\theta) > 0$. Hence the mechanism is not incentive-compatible. ■

Because the implementation of power-sharing (non-stationary) mechanisms is often complicated by the factors that are outside the present model - e.g., commitment problems associated with the exclusive access to government resources once in power - Theorem 3 is a rather grim result.

The next theorem provides a necessary and sufficient condition for the existence of non-stationary implementable mechanisms.

Theorem 4 (1) *Non-stationary incentive-compatible mechanisms exist if and only if*

$$\frac{1 + \bar{\theta}}{2\bar{\theta}} \geq P(\theta_{\min}) + \int_{\theta_{\min}}^{\bar{\theta}} p(\theta) e^{-rs^*(\theta)} d\theta; \quad (2)$$

(2) *Condition (2) is satisfied if $p(\theta)$ is increasing and sufficiently convex.*

Proof. (1) The proof proceeds by analyzing four collectively exhaustive types of strategy profiles. I first derive the necessary and sufficient conditions for the existence of an equilibrium strategy profile of the first type. I then show that the necessary and sufficient conditions for the existence of an equilibrium strategy profile of each of the other types must be more demanding.

Restrict attention to flat mechanisms (the extension of the argument to responsive mechanisms is straightforward). Because the mediator's advice is independent of the players' messages, truthful revelation is incentive-compatible, and the mediator's messages convey no new information to the players. Thus a direct mechanism is incentive-compatible if the discounted present value of following the mediator's advice is higher than the expected discounted present value of the war of attrition for every possible type of each player, given the players' prior beliefs.

From Theorem 1 and $E[u_i(t_i)]$ (see expression (1)), the expected discounted present value of deviating (to the war of attrition) is

$$\frac{1}{r} (b_i e^{-rs(\theta_i)} (1 - P(\theta_i)) + a_i \int_{\theta_{\min}}^{\theta_i} p(\theta') e^{-rs(\theta')} d\theta' + a_i P(\theta_{\min})), \quad (3)$$

where $s(\theta_i) = \frac{1}{r} \int_{\theta_{\min}}^{\theta_i} (\theta' - 1) \frac{p(\theta')}{1 - P(\theta')} d\theta'$.

Consider first a strategy of the following form. Let n be a natural number and let $T \in \mathbb{R}_{++}$ be a length of time. Let $t \in \mathbb{R}_+$ be a point in calendar time. Then $\forall t$,

choose

$$\begin{cases} X & \text{if } \exists n, \text{ s.t. } (n-1)T \leq t < nT \\ Y & \text{else.} \end{cases} \quad (4)$$

The discounted present value of following the mediator's advice, given that the opponent does also, is lowest for player 2 at $t = (n-1)T$ for n odd, and for player 1 at $t = (n-1)T$ for n even. At such a point, the discounted present value is

$$\begin{aligned} & \int_0^T b_i e^{-rt} dt + \int_T^{2T} a_i e^{-rt} dt + \int_{2T}^{3T} b_i e^{-rt} dt + \dots \\ &= b_i \int_0^\infty e^{-rt} dt + (a_i - b_i) \left(\int_T^{2T} e^{-rt} dt + \int_{3T}^{4T} e^{-rt} dt + \dots \right) \\ &= \frac{1}{r} (b_i - (a_i - b_i) \sum_{k=1}^\infty (-1)^k e^{-rkT}) \\ &= \frac{1}{r} (b_i + (a_i - b_i) \frac{1}{1 + e^{rT}}). \end{aligned} \quad (5)$$

The strategy profile suggested by the mediator is a PBE iff

$$\begin{aligned} \frac{1}{r} (b_i + (a_i - b_i) \frac{1}{1 + e^{rT}}) &\geq \frac{1}{r} b_i e^{-rs(\theta_i)} (1 - P(\theta_i)) \\ &\quad + \frac{1}{r} (a_i \int_{\theta_{\min}}^{\theta_i} p(\theta') e^{-rs(\theta')} d\theta' + a_i P(\theta_{\min}) + a_i P(\theta_{\min})) \end{aligned}$$

for all (a_i, b_i) . After some algebraic manipulations, we get the following equivalent inequality expressed in terms of θ_i :

$$1 + (\theta_i - 1) \frac{1}{1 + e^{rT}} \geq e^{-rs(\theta_i)} (1 - P(\theta_i)) + \theta_i \int_{\theta_{\min}}^{\theta_i} p(\theta') e^{-rs(\theta')} d\theta' + \theta_i P(\theta_{\min}). \quad (6)$$

The LHS of (6) is increasing and linear in θ_i . Using the fact that $\frac{\partial s}{\partial \theta} = \frac{1}{r} (\theta - 1) \frac{p(\theta)}{1 - P(\theta)}$, the derivative of the RHS reduces to $\int_{\theta_{\min}}^{\theta_i} p(\theta') e^{-rs(\theta')} d\theta'$, and so the RHS is increasing in θ_i and convex $\forall \theta_i \in (\theta_{\min}, \bar{\theta})$. Since (6) holds with equality at $\theta_i = \theta_{\min} = 1$, it follows that (6) holds $\forall \theta_i$ iff it holds for $\theta_i = \bar{\theta}$. Substituting $\theta_i = \bar{\theta}$ into (6) and

recognizing that $P(\bar{\theta}) = 1$, (6) implies

$$1 + (\bar{\theta} - 1) \frac{1}{1 + e^{rT}} \geq \bar{\theta} \int_{\theta_{\min}}^{\bar{\theta}} p(\theta) e^{-rs(\theta)} d\theta.$$

Multiplying by $(1 + e^{rT})$ and re-arranging terms, we obtain

$$e^{rT} \left(1 - \bar{\theta} \int_{\theta_{\min}}^{\bar{\theta}} p(\theta) e^{-rs(\theta)} d\theta - \bar{\theta} P(\theta_{\min})\right) + \bar{\theta} \left(1 - \int_{\theta_{\min}}^{\bar{\theta}} p(\theta) e^{-rs(\theta)} d\theta - P(\theta_{\min})\right) \geq 0. \quad (7)$$

From $r > 0$ and $s(\theta) \geq 0$, it follows that $e^{-rs(\theta)} \leq 1$. Hence,

$$\int_{\theta_{\min}}^{\bar{\theta}} p(\theta) e^{-rs(\theta)} d\theta < \int_{\theta_{\min}}^{\bar{\theta}} p(\theta) d\theta,$$

and so the second term in (7) is always positive.

Because $e^{rT} > 1$ with $\lim_{T \rightarrow 0} e^{rT} = 1$, it follows that $\exists T$ s.t. the strategies described constitute a PBE iff condition (2) in the statement of the theorem holds.

Consider a similar strategy profile in which $\forall n \in \mathbb{N}$

$$\begin{cases} (X, X) & \text{if } (n-1)(T_1 + T_2) \leq t < (n-1)T_2 + nT_1 \\ (Y, Y) & \text{if } (n-1)T_2 + nT_1 \leq t < n(T_1 + T_2). \end{cases} \quad (8)$$

Suppose $T_2 < T_1$. Then, holding type constant, player 2's present discounted value at $t = (n-1)(T_1 + T_2)$ is less than player 1's at $t = (n-1)T_2 + nT_1$. Thus, the profile is a weak PBE iff player 2 of type $\bar{\theta}$ prefers not to defect at $t = 0$. While her discounted present value of the war of attrition (expression (3)) is the same as above, her present discounted value of following the mediator's advice is

$$\begin{aligned} & \int_0^{T_1} b_2 e^{-rt} dt + \int_{T_1}^{T_1+T_2} a_2 e^{-rt} dt + \int_{T_1+T_2}^{2T_1+T_2} b_1 e^{-rt} dt + \int_{2T_1+T_2}^{2(T_1+T_2)} a_2 e^{-rt} dt + \dots \\ &= \frac{1}{r} [b_2 + (a_2 - b_2) \frac{1}{e^{r(T_1+T_2)} - 1} (1 - e^{rT_2})]. \end{aligned}$$

Given that $T_2 < T_1$, and hence $T_2 < \frac{1}{2}(T_1 + T_2)$, this is strictly less than (5) for $T = \frac{1}{2}(T_1 + T_2)$. Thus the necessary and sufficient condition for the existence of an equilibrium of form (8) is always strictly more demanding than the condition for the

existence of an equilibrium of form (4). Since the game is symmetric and strategically equivalent $\forall t$, the results are the same for $T_1 < T_2$.

Consider a strategy profile that may be characterized by an infinite sequence of points in time, $\{t^n\}_{n=1}^\infty$, such that $\forall n \in \mathbb{N}$

$$\begin{cases} (X, X) & \text{if } t^{n-1} \leq t < t^n \text{ for } n \text{ odd} \\ (Y, Y) & \text{if } t^{n-1} \leq t < t^n \text{ for } n \text{ even.} \end{cases}$$

Now the minimum present discounted value of following this strategy, given that the other player does also, occurs at some $t \in \{t^n : n \text{ is even}\}$ for player 1 and at some $t \in \{t^n : n \text{ is odd}\}$ for player 2, but their present discounted values may vary across these values of t . Since the strategy profile is an equilibrium iff every type of each player prefers not to deviate at every point in time, it is an equilibrium iff the highest type of each player prefers not to deviate at her time of lowest present discounted value. But this condition is strictly more difficult to satisfy than the necessary and sufficient condition for existence for some equilibrium of form (8).

Finally, observe that since $b_1 > 0$ for all types, any strategy profile that requires the play of (X, Y) or (Y, X) for a positive measure of time has a lower present discounted value than some strategy profile in which only (X, X) and (Y, Y) are played.

(2) First observe that $\lim_{\bar{\theta} \rightarrow \infty} \frac{1+\bar{\theta}}{2\bar{\theta}} = \frac{1}{2}$ and that LHS of condition (2) is decreasing in $\bar{\theta}$. Second, observe that $e^{-rs^*(\theta)}$ is decreasing in θ (since from Theorem 1, $s^*(\theta)$ is increasing in θ), that $e^{-rs^*(\theta_{\min})} = e^0 = 1$, and that $\lim_{s^*(\theta) \rightarrow \infty} e^{-rs^*(\theta)} = 0$. Thus,

$\int_{\theta_{\min}}^{\bar{\theta}} p(\theta)e^{-rs^*(\theta)}d\bar{\theta} + P(\theta_{\min}) < 1$, and $\int_{\theta_{\min}}^{\bar{\theta}} p(\theta)e^{-rs^*(\theta)}d\bar{\theta} + P(\theta_{\min})$ decreases as probability mass is redistributed from lower θ to higher θ . hence condition (2) is satisfied if $p(\theta)$ is increasing and sufficiently convex. ■

The condition for the existence of an ICDM is obtained from the conditions under which both players will choose to abide voluntarily (at every point in time) with recommendations to switch back and forth at equal time intervals between the most favored action profiles of each of the players, (X, X) and (Y, Y) . Because this mechanism dictates that each player's favorite outcome be played for the same duration, regardless of type, the players are willing to reveal their types to the mediator. Furthermore, playing one action profile for a longer period than the other would necessarily make the incentive compatibility constraint more stringent for the less favored player, and would consequently reduce the range of parameter values for which such a mechanism could be sustained. Likewise, any mechanism that dictated (X, Y) or (Y, X) with positive probability could not be sustained for as wide a range of parameter values. The existence of an incentive compatible mechanism of this sort relies on the ability of the players to switch between action profiles arbitrarily quickly, to ensure that the willingness of each player to wait through her less favored action profile in order to obtain the higher payoff of her more favored one.

Condition (2) is satisfied if the probability density function $p(\theta)$ is increasing and sufficiently convex, or, more generally, if the probability density of the lowest types θ is sufficiently small. Intuitively, an incentive-compatible mechanism exists only if the war of attrition is sufficiently costly, in expectation, which is the case when low types (who would quit quickly) are sufficiently rare.

3.4 Mediated Settlements with Utility Transfers

The final result characterizes what happens when we allow for the possibility of a mediator suggesting a vector of transfers as a part of a settlement. As above, such transfers have to be self-enforcing in that the net transfer payer has to be willing to continue paying them given the learning involved in each party's accepting the settlement.

Theorem 5 *There is no pure-strategy stationary incentive-compatible mechanism employing transfers.*

Proof. Player i prefers to pay v rather than starting a war of attrition if

$$a_i - v \geq a_i \int_0^{s_i} p_j(s) e^{-rs} ds + b_i e^{-rs_1} (1 - \Pr(s_j \leq s_i)),$$

where s_i is i 's equilibrium strategy in the war of attrition, as defined by (??). Let \hat{v} be the highest transfer that i is willing to pay in equilibrium:

$$\hat{v}(a_i, b_i) = a_i \left(1 - \int_0^{s_i} p_j(s) e^{-rs} ds\right) - b_i e^{-rs_1} (1 - \Pr(s_j \leq s_i)).$$

Similarly, j accepts v rather than starting a war of attrition if

$$b_j + v \geq a_j \int_0^{s_j} p_i(s) e^{-rs} ds + b_j e^{-rs_j} (1 - \Pr(s_i \leq s_j)).$$

Let \underline{v} be the minimum transfer that j would be willing to accept in equilibrium:

$$\underline{v}(a_j, b_j) = a_j \left(1 - \int_0^{s_j} p_i(s) e^{-rs} ds\right) - b_j (1 - e^{-rs_j} (1 - \Pr(s_i \leq s_j))).$$

We next establish three useful claims:

Claim 1: For any $(a', b'), (a'', b'')$ such that $\hat{v}(a', b') = \hat{v}(a'', b'')$, $b'' > b'$ implies $a'' > a'$.

Proof of Claim 1. Let $v' \equiv \hat{v}(a', b')$ and $s' \equiv s(\frac{a'}{b'})$. Consider (a', b'') s.t. $b'' > b'$, and let $s'' = s(\frac{a'}{b''})$. Given that $\hat{v}(a, b) = 0$ for all (a, b) s.t. $a = b$ and that two different level curves never intersect, it is sufficient to show that $\hat{v}(a', b'') < v'$.

By definition of equilibrium, player i of type (a', b'') obtains at least as high an expected utility in the war of attrition from playing s'' as from playing s' . Thus:

$$\begin{aligned}\hat{v}(a', b'') &= a'(1 - \int_0^{s''} p_j(s)e^{-rs} ds) - b''e^{-rs''}(1 - \Pr(s_j \leq s'')) \\ &\leq a'(1 - \int_0^{s'} p_j(s)e^{-rs} ds) - b''e^{-rs'}(1 - \Pr(s_j \leq s')).\end{aligned}$$

Because $b'' > b'$,

$$\begin{aligned}&a'(1 - \int_0^{s'} p_j(s)e^{-rs} ds) - b''e^{-rs'}(1 - \Pr(s_j \leq s')) \\ &< a'(1 - \int_0^{s'} p_j(s)e^{-rs} ds) - b'e^{-rs'}(1 - \Pr(s_j \leq s')) = v'.\end{aligned}$$

Thus, $\hat{v}(a', b'') < v'$, which establishes the claim.

Claim 2: For any $(a', b'), (a'', b'')$ such that $\underline{v}(a', b') = \underline{v}(a'', b'')$, $b'' < b'$ implies $a'' > a'$.

Proof of Claim 2. Let (a', b') be such that $\underline{v}(a', b') = v'$ and let $s' \equiv s(\frac{a'}{b'})$. Consider (a', b'') s.t. $b'' < b'$ and let $s'' = s(\frac{a'}{b''})$. Proceeding analogously to the proof of Claim 1, we show that $v' < \underline{v}(a', b'')$.

By definition of equilibrium, a player of type (a', b'') obtains at least as high an expected utility in the war of attrition from playing s'' as from playing s' . Thus:

$$\begin{aligned}v' &= a' \int_0^{s'} p_i(s)e^{-rs} ds - b'(1 - e^{-rs'}(1 - \Pr(s_i \leq s'))) \\ &< a' \int_0^{s''} p_i(s)e^{-rs} ds - b''(1 - e^{-rs''}(1 - \Pr(s_i \leq s''))).\end{aligned}$$

Because $b'' < b'$,

$$\begin{aligned} & a' \int_0^{s'} p_i(s) e^{-rs} ds - b''(1 - e^{-rs'}(1 - \Pr(s_i \leq s'))) \\ & \leq a' \int_0^{s''} p_i(s) e^{-rs} ds - b''(1 - e^{-rs''}(1 - \Pr(s_i \leq s''))) = \underline{v}(a', b''). \end{aligned}$$

Thus, $v' < \underline{v}(a', b'')$, establishing the claim.

Claim 3:

Agents who are willing to pay \hat{v} are disproportionately more likely to be types with higher θ and agents who are willing to accept \underline{v} are disproportionately more likely to be types with lower θ relative to the initial distribution.

Proof of Claim 3. The diagram shows right triangles formed by the a -axis, lines parallel to the b -axis, and level curves of θ . The argument is formulated geometrically. For any \hat{a} s.t. $0 < \hat{a} < \bar{a}$, a constant function $f(b) = \hat{a}$ creates pairs of similar triangles $\triangle AEF$ and $\triangle ABC$, and $\triangle AEG$ and $\triangle ABD$. By the properties of similar triangles, $\frac{AF}{AC} = \frac{AE}{AB}$ and $\frac{AG}{AD} = \frac{AE}{AB}$. Combining these expressions, $\frac{AF}{AC} = \frac{AG}{AD}$. Thus, for any $H \neq G$ on segment GD , $\frac{FC}{AC} > \frac{HD}{AD}$ and for any $I \neq G$ on segment AG , $\frac{FC}{AC} < \frac{ID}{AD}$.

Because a and b are independently distributed and $\theta \geq 1$ for all θ , $p(a|\theta') = p(a|\theta'')$ for all θ' and θ'' . It follows that for any \hat{a} and any θ', θ'' , $\Pr(a \leq \hat{a}|\theta') = \Pr(a \leq \hat{a}|\theta'')$. The claim then follows from the fact that, by Claim 1, any level curve of \hat{v} is upward sloping, and by Claim 2, any level curve of \underline{v} is downward sloping in (b, a) -space.

As the paying agent updates negatively about the recipient's θ and the recipient updates positively about the payer's θ , the payer's strategy in the war of attrition increases and the recipient's decreases (see e.g., Hafer 2006). Thus, the payer's equilibrium expected utility from playing the war of attrition increases and the recipient's decreases. As a result, for any (a, b) , $\hat{v}(a, b)$ increases and $\underline{v}(a, b)$ decreases in subsequent periods. Since v is fixed, this implies that, as long as the truce is maintained, the set of types who prefer defection is enlarged in each period in which the truce is maintained. Thus, all truces are eventually violated. ■

The impossibility of implementing a lasting truce in which the player that obtains its more preferred outcome must compensate its opponent through ongoing transfers follows from the fact that the initial willingness of the parties to abide by such an arrangement credibly communicates information about their payoffs. Players who pay a given transfer rather than initiating active conflict do so precisely because the difference between their payoff from their most preferred outcome a and their expected equilibrium payoff from conflict is greater than the transfer v . But an indication that they have high values of a is also an indication that they are more likely than a randomly drawn member of the population of types to have high values of θ . Similarly, players who accept the transfer rather than initiating conflict do so

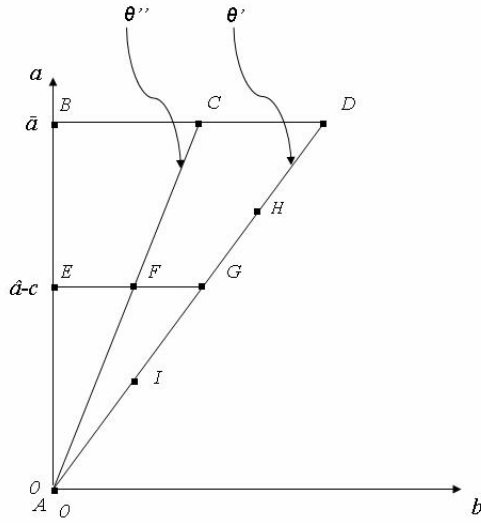


Figure 2: Proof of Claim 3

because the difference between their expected equilibrium payoffs from conflict and their payoffs b from their less preferred outcome is less than the transfer v . But this indicates that they have high values of b , which is also an indication that they are more likely than a randomly drawn member of the population of types to have low values of θ . Thus by abiding by the truce, the player that is paying the transfer is learning that her opponent is less willing to fight than she had previously believed, and the recipient of the transfer is learning that her opponent is more willing to fight than she had previously believed. Thus the players update in the course of the truce in a way that encourages the payer to initiate conflict and encourages the recipient to surrender quickly once that conflict has begun. Furthermore, the types of recipient that correspond to the lowest values of θ will surrender immediately.

Still, the outcome is not “all bad” from the welfare standpoint. The following Corollary underscores that mediation with transfers yields ex ante efficiency gains by delaying conflict and decreasing its expected duration.

Corollary 1 *There exist stationary mechanisms with transfers that can be implemented for positive and finite time with positive probability. In such mechanisms, conflict is delayed, the expected duration of conflict is shorter than in the symmetric equilibrium of the unmediated game, and in some instances no conflict occurs at all after the cessation of the transfers.*

These efficiency improvements result because some pairs of types will, in fact, prefer to engage in peaceful side payments for a positive number of periods before one

of them initiates conflict. Although the learning that occurs while the parties abide by the truce leads ultimately to its dissolution, the time that this learning requires delays the onset of conflict. The information that is revealed about the parties reduces the expected duration of the conflict once it commences because it reduces the amount of time that recipients are willing to devote to conflict before surrendering. Indeed, knowing that it is certain to lose, a recipient with a type corresponding to the lowest θ will choose not to resist at all when its opponent chooses to stop paying the transfer.

The much more positive results obtained by Jackson and Sonnenschein (forthcoming) are an interesting point of comparison to the results obtained here. They consider an environment in which multiple agents face a series of identical collective decision problems in which they have conflicting interests and private information about the strength of their own preferences over outcomes and identify an incentive compatible mechanism that implements efficient (in the limit) outcomes by linking the decision problems through endogenously determined restrictions on the agents' sets of possible messages. The crucial difference between their environment and the one examined above is that, in their model, each agent's preferences are a new draw for each problem, whereas here we have assumed that preferences are fixed across periods. Fixing preferences makes it physically impossible to reveal information consistently across periods in the Jackson-Sonnenschein mechanism and it undermines the incentive to do so, since what a player's behavior reveals about her type in one period can be used against her in future periods. The truth is, no doubt, usually in the middle: in most applications, one would expect type to be neither completely persistent nor completely independent across periods. The stark difference between their results and those obtained here suggests the value of exploring the case of imperfectly persistent types.

4 Discussion

References

- [1] Abreu, Dilip and David Pearce. 2006. "Reputational Wars of Attrition with Complex Bargaining Postures." Princeton University Mimeo.
- [2] Abreu, Dilip and David Pearce. 2002. "Bargaining, Reputation, and Equilibrium Selection in Repeated Games." Princeton University Mimeo.
- [3] Alesina, Alberto. 1988. "Credibility and Policy Convergence in a Two-Party System with Rational Voters." *American Economic Review* 78, 796-805.
- [4] Alesina, Alberto and Alan Drazen. 1991. "Why Are Stabilizations Delayed?" *American Economic Review* 81 (December): 1170-1188.

- [5] Baliga, Sandeep and Thomas Sjoström. 2004. "Arms Races and Negotiations." *Review of Economic Studies*.
- [6] Baliga, Sandeep and Stephen Morris. 1998. "Cheap Talk and Coordination with Payoff Uncertainty." Cowles Foundation Discussion Paper #1403.
- [7] Baliga, Sandeep and Stephen Morris. 2002. "Cheap Talk, Coordination, and Spillovers." *Journal of Economic Theory*.
- [8] Banks, Jeffrey S. 1990. "Equilibrium Behavior in Crisis Bargaining Games." *American Journal of Political Science* 34 (3), 599-614.
- [9] Banks, Jeffrey S. and Randall L. Calvert. 1992. "A Battle-of-the-Sexes game with Incomplete Information." *Games and Economic Behavior* 4, 347-72.
- [10] Ben-Porath, Elchanan. 2002. "Cheap Talk in Games with Incomplete Information." Hebrew University Mimeo.
- [11] Bueno de Mesquita, Bruce. 2000. "Popes, Kings, and Endogenous Institutions: the Concordat of Worms and the Origins of Sovereignty." *International Studies Review* 2 (2), 93-118.
- [12] Bueno de Mesquita, Bruce, James Morrow, and Ethan Zorick. 1997. "Capabilities, Perception, and Escalation." *American Political Science Review* 91 (1), 15-27.
- [13] Casella, Alessandra and Barry Eichengreen. 1996. "Can Foreign Aid Accelerate Stabilization?" *The Economic Journal* 106, 605-619.
- [14] Dixit, Avinash K. 2004. *Lawlessness and Economics: Alternative Modes of Governance*. Princeton: Princeton University Press.
- [15] Dixit, Avinash, Gene M. Grossman, and Faruk Gul. 2000. "The Dynamics of Political Compromise." *Journal of Political Economy* 108 (3), 531-568.
- [16] Drazen, Alan and Vincent Grilli. 1993. "The Benefits of Crises for Economic Reforms." *American Economic Review* 82, 598-607.
- [17] Fudenberg, Drew and Jean Tirole. 1991. *Game Theory*. Cambridge: MIT Press.
- [18] Fudenberg, Drew and Jean Tirole. 1986. "A Theory of Exit in Duopoly." *Econometrica* 54 (4), 943-60.
- [19] Hafer, Catherine. 2006. "On the Origins of Property Rights: Conflict and Production in the State of Nature." *Review of Economic Studies* 73 (1), 119-144.
- [20] Hsieh, Chang-Tai. 2000. "Bargaining Over Reform." *European Economic Review* 44 (9), 1659-76.

- [21] Jackson, Matthew O. and Hugo F. Sonnenschein. 2005. "Overcoming Incentive Constraints by Linking Decisions." *Econometrica*.
- [22] Myerson, Roger. 1985. "Bayesian Equilibrium and Incentive Compatibility." In L. Hurwicz, D. Schmeidler, and H. Sonnenschein, eds., *Social Goals and Social Organization*. Cambridge: Cambridge University Press, 229-59.
- [23] Nalebuff, Barry. 1987. "Credible Pretrial Negotiation." *The Rand Journal of Economics* 18 (2), 198-210.
- [24] Powell, Robert L. 2004. "Bargaining and Learning While Fighting." *American Journal of Political Science* 48 (2), 344-61.
- [25] Spolaore, Enrico. 2004. "Adjustments in Different Government Systems." *Economics and Politics* 16 (2), 117-46.