Coordination, Standoff, and Communication*

Catherine Hafer†

PRELIMINARY

02/20/2006, first draft: 12/06/2004

Abstract

This paper analyzes a model of political standoff based on a continuous-time dynamic extension of an incomplete-information asymmetric coordination game (a battle-of-the-sexes). 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 (at least, for the relevant dispute); and the players are uncertain about each other’s resolve, i.e., about the relative strength of their interests in one outcome over another. I demonstrate that this game is strategically equivalent to an incomplete-information war of attrition, and exploit this strategic equivalence to characterize 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, 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). I then analyze the cheap-talk extension of the game with and without mediation. 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.

*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 the Graduate School of Business at Stanford University, and especially from David Austen-Smith, David Baron, Ethan Bueno de Mesquita, Tim Feddersen, Sven Feldmann, Sandy Gordon, Bard Harsted, and Dimitri Landa, and Adam Meirowitz.
†Assist. Prof. of Politics, NYU, e-mail: catherine.hafer@nyu.edu
1 Introduction

This paper analyzes a model of political standoff: 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; they have some degree of contrary interests, so that the dispute between them is not trivial; the resolution of the dispute must be self-sustaining, i.e. there is no external enforcement of agreements (at least, for the relevant dispute); 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 first two features 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, in a variety of political interactions, third-party contract enforcement is unavailable, and so the feasible final resolutions include only those agreements to which both parties could credibly commit to carrying out. The requirement that final resolutions be credible in this way thus excludes a variety of arrangements that otherwise might have been possible, e.g. those involving “side-payments” or transfers of benefits from one group to another. Given this limitation on the set of viable agreements, there may not be enough flexibility—enough room for given and take—to achieve a resolution of the dispute through bargaining. In such circumstances,
the parties may find themselves in a 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.

Prominent examples include Ukrainian presidential election, Middle East conflict, Fiscal policy stabilization. DISCUSSION TO BE ADDED.

The model of standoff presented in this paper is based on a dynamic extension of an incomplete-information asymmetric coordination game (a battle-of-the-sexes). This game captures both the common and the contrary aspects of the interests of the parties to a standoff. After introducing this game, I demonstrate that it is strategically equivalent to an incomplete-information war of attrition, and exploit this strategic equivalence to obtain an 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 and The Revelation Principle (Myerson 1985) to examine the extent to which mediated and unmedi-
ated cheap-talk communication can be used successfully to end or to avert a standoff as a function of features of the strategic environment (e.g. the distribution of types, the players’ time preferences).

2 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 \((a_i, b_i, c_i)\), 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 \(r\).

\[
\begin{array}{c|cc}
 & X & Y \\
\hline
X & a_1, b_2 & c_1, c_2 \\
Y & c_1, c_2 & b_1, a_2 \\
\end{array}
\]

\[a_i \geq b_i > c_i, \ i \in \{1, 2\}\]

This dynamic extension of the battle-of-the-sexes game is a model of standoff. At the beginning of the game, each player must choose an action, \(X\) or \(Y\). Actions are

---

1Hence this game may be thought of as a generalized dynamic extension of the game analyzed in Banks and Calvert (1992).
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, and thus the two games are, in fact, strategically equivalent.

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

$$u_i(t_1, t_2 > t_1) = \int_0^\infty c_i e^{-rt} dt + \int_{t_1}^\infty (b_i - c_i) e^{-rt} dt = \frac{c_i}{r} + \frac{b_i - c_i}{r} e^{-r t_1}.$$ 

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

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

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, $(b_i - c_i)$. 

5
2.1 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.

Our first result provides a characterization of the symmetric equilibrium outcome:

**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 \( \frac{a_i - c_i}{b_i - c_i} \);
2. has a standoff of duration \( s^*(\theta_i) = \frac{1}{r} \int_{\theta_i}^{\theta_i} (\theta - 1) \frac{p(\theta)}{1 - P(\theta)} d\theta \), where player \( i \) has the lower value of \( \theta_i := \frac{a_i - c_i}{b_i - c_i} \).

**Proof.** The expected payoff for player \( i \) from action \( t_i \) is

\[
E[u_i(t_i)] = (1 - \Pr(t_j < t_i)) \left[ \frac{c_i}{r} + \frac{b_i - c_i}{r} e^{-rt_i} \right] + \int_0^{t_i} \Pr(t_j = t) \left[ \frac{c_i}{r} + \frac{a_i - c_i}{r} e^{-rt} \right] dt
\]

\[
= \frac{c_i}{r} + \frac{b_i - c_i}{r} e^{-rt_i} (1 - \Pr(t_j < t_i)) + \frac{a_i - c_i}{r} \int_0^{t_i} \Pr(t_j = t) e^{-rt} dt. \tag{1}
\]

Then the first-order condition is

\[
- \Pr(t_j = t_i) \frac{b_i - c_i}{r} e^{-rt_i} + (1 - \Pr(t_j < t_i))(-b_i - c_i)e^{-rt_i} + \Pr(t_j = t_i) \frac{a_i - c_i}{r} e^{-rt_i} = 0
\]

\[\text{Note that, as in all incomplete-information wars of attrition, there are also two asymmetric equilibria: player 1 “fights forever” and player 2 “quits immediately;” and 1 “quits immediately” and 2 “fights forever.” Recall that, because the players’ types are not common knowledge, their types cannot be invoked to select one equilibrium over the other.}\]
In order to solve the first-order condition, we must express $\Pr(t_j < t_i)$ in terms of primitives, e.g. as a probability of a type $(a_j, b_j, c_j)$ that chooses to quit before time $t_i$. This requires being able to identify a type that corresponds to $t$, but because type is three-dimensional, multiple types may (and, in equilibrium, do) choose the same stopping time $t$. 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 cost of continuing the conflict, $(a_i - b_i) - (b_i - c_i)$, to the cost of continuing the conflict, $(b_i - c_i)$, i.e. in terms of $\frac{a-c}{b-c}$. This is algebraically equivalent to dividing the first order condition by $(b - c)$. Let $\theta$ represent $\frac{a-c}{b-c}$, and solve for players’ optimal stopping times as a function of $\theta$.

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 of action in equilibrium is strictly monotonic. To see that it is, recall that, from the definition of Bayesian Nash equilibrium, the equilibrium action of type $\theta$ 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 $(a_i', b_i', c_i')$, and hence of type $\theta_i'$, and if $t_0$ is the equilibrium action of an agent of type $(a_i'', b_i'', c_i'')$, and hence of type $\theta_i''$, we have the following two inequalities:

$$\frac{c_i'}{r} + \frac{b_i'}{r} \int_0^{t_i'} e^{-rt} (1 - \Pr(t_2 < t_i')) \, dt + \frac{a_i' - c_i'}{r} \int_0^{t_i'} \Pr(t_2 = t) e^{-rt} \, dt \geq \frac{c_i''}{r} + \frac{b_i''}{r} \int_0^{t_i''} e^{-rt} (1 - \Pr(t_2 < t_i'')) \, dt + \frac{a_i'' - c_i''}{r} \int_0^{t_i''} \Pr(t_2 = t) e^{-rt} \, dt;$$

$$\frac{c_i'}{r} + \frac{b_i'}{r} \int_0^{t_i'} e^{-rt} (1 - \Pr(t_2 < t_i')) \, dt + \frac{a_i' - c_i'}{r} \int_0^{t_i'} \Pr(t_2 = t) e^{-rt} \, dt \geq \frac{c_i''}{r} + \frac{b_i''}{r} \int_0^{t_i''} e^{-rt} (1 - \Pr(t_2 < t_i'')) \, dt + \frac{a_i'' - c_i''}{r} \int_0^{t_i''} \Pr(t_2 = t) e^{-rt} \, dt.$$
Multiplying each inequality by \( r \), canceling \( \frac{a'}{r} \) and \( \frac{a''}{r} \), respectively, and dividing by \((b'_i - c'_i)\) and \((b''_i - c''_i)\), respectively, these inequalities can be expressed in terms of \( \theta'_i \) and \( \theta''_i \), 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:

\[
\left( \frac{a' - c'}{b' - c'} - \frac{a'' - c''}{b'' - c''} \right) \int_0^{t'_1} \Pr(t_2 = t)e^{-rt}dt \geq \left( \frac{a' - c'}{b' - c'} - \frac{a'' - c''}{b'' - c''} \right) \int_0^{t''_1} \Pr(t_2 = t)e^{-rt}dt.
\]

It follows, then, that if \( \frac{a' - c'}{b' - c'} > \frac{a'' - c''}{b'' - c''} \), then \( t' \geq t'' \). Hence equilibrium strategy must be increasing in type \( \theta = \frac{a - c}{b - c} \). Exploiting the symmetry of the game (i.e. the common prior), part 1 is established.

Let \( s(\theta) \) be time chosen by type \( \theta \), i.e., \( \theta \)'s strategy. As in all continuous-time incomplete-information wars of attrition, \( s(\theta) \) is continuous and atomless (Fudenberg and Tirole 1986). Hence the inverse of \( s(\theta) \), \( \Theta(s) \), exists. The optimization problem may then be written

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

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 \( \theta \). Replacing \( \frac{a_i - c_i}{b_i - c_i} \) with \( \theta_i \), 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, \( k \):

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

Given strict monotonicity, if \( s(\theta_l) \) were strictly positive, then \( \theta_l \) 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 model of standoff predicts that the player who has a higher ratio of the benefit of winning over the opportunity cost of continuing the standoff 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 \( \theta \).

Suppose now that the players have a common payoff from a failure to coordinate, i.e. that they know \( c_1 = c_2 = c \). Theorem 1 still holds if we substitute \( P(\theta|c) \) and
$p(\theta|c)$ for $P(\theta)$ and $p(\theta)$; in particular, the winner is still the player with higher $\theta$. However, changing the common value $c$ may change the ordering of the players in $\theta$-space, as can be seen from the following table of $\theta$-types, calculated for the given values of $(a, b)$ and $c$.

<table>
<thead>
<tr>
<th>$(a, b)$</th>
<th>(11, 6)</th>
<th>(9, 5)</th>
<th>(7, 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c = 0$</td>
<td>$\frac{5}{6}$</td>
<td>$\frac{4}{5}$</td>
<td>$\frac{3}{4}$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>$2\frac{1}{4}$</td>
<td>$2\frac{1}{3}$</td>
<td>$2\frac{1}{2}$</td>
</tr>
</tbody>
</table>

**Proposition 1** If the players have a common payoff $c$ from a failure to coordinate, then increasing $c$:

1. increases the amount of time each player is willing to devote to conflict;
2. increases the actual duration of the conflict;
3. affects who wins the conflict if $c$ increases past the threshold $\frac{a_1b_2 - a_2b_1}{(a_1 + b_2) - (b_1 + a_2)}$ and either $a_1 + b_2 > b_1 + a_2$ and $a_1b_2 > a_2b_1$ or $a_1 + b_2 < b_1 + a_2$ and $a_1b_2 < a_2b_1$. In such cases, the outcome selected switches from the more to the less classically efficient one.

**Proof.** Part 1 follows from $\frac{\partial s}{\partial \theta} > 0$ and $\frac{\partial \theta}{\partial c} > 0$. Part 2 is immediate from part 1. From Proposition 1, changing the order of the players in $\theta$-space changes the outcome. That ordering changes if there exists $c$ such that $\theta_i = \theta_j$. If $a_1 + b_2 > b_1 + a_2$, then $(X, X)$ is the classically efficient outcome and $\theta_1 < \theta_2$ if and only if $c > \frac{a_1b_2 - a_2b_1}{(a_1 + b_2) - (b_1 + a_2)}$. If $a_1 + b_2 < b_1 + a_2$, then $\theta_1 > \theta_2$ if and only if $c > \frac{a_1b_2 - a_2b_1}{(a_1 + b_2) - (b_1 + a_2)}$.

2.2 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 $\theta$, 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.

Call the cheap-talk extension of the standoff game in which the communication occurs before the play of the standoff game the *pre-play cheap-talk extension of the standoff game*. To prove the theorem, I first show that the claim of the theorem holds for the pre-play cheap-talk extension of the standoff game, and then argue that the result extends to any cheap-talk extension of the game. The argument in relation to the pre-play cheap-talk follows as the conjunction of the following two lemmata:

**Lemma 1** There exists no monotonic informative equilibrium in the pre-play cheap-talk extension of the standoff game without mediation.

**Proof.** Because the existence of an informative monotonic 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 \( u_i(.) = \frac{a_i}{r} \) and \( u_j(.) = \frac{b_j}{r} \). 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 $\theta$, 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 $E[u_1(0)] = \frac{b}{r}$.

Note that if $\theta_i < \hat{\theta}$ sends $m^1$ and her opponent sends $m^2$, then $\theta_i$ quits immediately and obtains the same payoff, $\frac{b_i}{r}$. 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 $\frac{a_i}{r}$. 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.

The next lemma extends the result in Lemma 1 to non-monotonic equilibria:

**Lemma 2** There exists no informative non-monotonic equilibrium in the pre-play cheap-talk extension of the standoff game without mediation.

**Proof.** Consider first the possibility of a two-message non-monotonic equilibrium. If such an equilibrium exists, then, for some $\theta', \theta''$ 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 \( \theta \) 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

\[
\hat{s} \equiv \lim_{\theta_1 \to -\theta_0} s(\theta_1|m_1 = m^1) = \lim_{\theta_1 \to +\theta_0} s(\theta_1|m_1 = m^1)
\]

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 \( \theta' \) strictly prefers \( m^1 \) to \( m^2 \), and \( \theta'' \) 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 \( \theta', \theta'' \) 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. ■

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. QED.

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.

2.3 Standoff and Mediated Cheap Talk

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 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 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 \( b \), 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 \( \theta \) 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 \( \frac{b_i}{r_i} \), 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

\[
1 + \theta - 2\int_{\theta_{\min}}^{\theta} p(\theta)e^{-\theta R^{\theta_{\min}}(x-1)p(x)} dx > 0; \tag{2}
\]

(2) comparative statics on the condition (2) to be added.

**Proof.** 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

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

(3)

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

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

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

(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

\[
\int_0^T b_i e^{-rt} dt + \int_T^{2T} a_i e^{-rt} dt + \int_{2T}^{3T} b_i e^{-rt} dt + ... \\
= \int_0^{\infty} b_i e^{-rt} dt + (a_i - b_i)(\int_T^{2T} e^{-rt} dt + \int_{3T}^{4T} e^{-rt} dt + ...) \\
= \frac{1}{r}(b_i - (a_i - b_i) \sum_{k=1}^{\infty} (-1)^k e^{-rkT}) \\
= \frac{1}{r}(b_i + (a_i - b_i) \frac{1}{1 + e^{rT}}).
\]

(5)
The strategy profile suggested by the mediator is a PBE iff

\[
\frac{1}{r}(b_i + (a_i - b_i)\frac{1}{1 + e^{rT}}) \geq \frac{1}{r}(c_i + (b_i - c_i)e^{-rs(\theta_i)}(1 - P(\theta_i)) + (a_i - c_i) \int_{\theta_{\min}}^{\theta_i} p(\theta') e^{-rs(\theta')} d\theta')
\]

for all \((a_i, b_i, c_i)\). Multiplying by \(r\), subtracting \(c_i\), dividing by \((b_i - c_i)\), and expressing the resulting inequality in terms of \(\theta_i\), we get

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

(6)

The LHS is increasing and linear in \(\theta_i\). 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 \(\theta_i\) 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

\[
e^{rT}(1 - \bar{\theta}) \int_{\theta_{\min}}^{\bar{\theta}} p(\theta) e^{-rs(\theta)} d\theta + \bar{\theta}(1 - \int_{\theta_{\min}}^{\bar{\theta}} p(\theta) e^{-rs(\theta)} d\theta) \geq 0.
\]  

(7)

From \(r > 0\) and \(s(\theta) \geq 0\), it follows that \(e^{-rs(\theta)} \leq 1\). Hence, \(\int_{\theta_{\min}}^{\bar{\theta}} p(\theta) e^{-rs(\theta)} d\theta < 1\), and so the second term in (7) is always positive.

Because \(e^{rT} > 1\) with \(\lim_{T \to 0} e^{rT} = 1\), 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 \( \forall n \in \mathbb{N} \)

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

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

\[
\frac{1}{r} \left[ b_2 + \left( a_2 - b_2 \right) \frac{1}{e^{r(T_1 + T_2)} - 1} \left( 1 - e^{-rT_2} \right) \right].
\]

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 $b_1 > c_i$ 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.

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 $\theta$ 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 Relation to Literature

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. 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.

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 below 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 in Banks and Calvert (1992) on the outcomes that may be obtained through mediation in a similar game are an interesting point of comparison. They 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.

Prominent political economy applications of incomplete information war of attrition include models of delayed fiscal stabilization (Alesina and Drazen (1991), Casella and Eichengreen (1995), Hsieh (1997), Spolaore (2004)). Relative to these models, the basic model in Section 2.1 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.
4 Conclusion

To be added.

References


