# Decoding Terror*

Sandeep Baliga
Northwestern University

Tomas Sjöström
Rutgers University

May 22, 2009

**Abstract**

In a conflict game with incomplete information, decisions are based on fear and greed. We find conditions under which the decision-making can be manipulated by "extremists" who send publicly observed cheap-talk messages. The power of extremists depends on the nature of the underlying conflict game. If actions are strategic complements, a "hawkish extremist" (terrorist) can increase the likelihood of conflict by sending messages which trigger mutual hostilities in a spiral of fear. If actions are strategic substitutes, a "dovish extremist" (pacifist) can send messages which cause one side to back down. But the hawkish extremist is unable to manipulate the outcome if actions are strategic substitutes, and the pacifist is equally powerless if actions are strategic complements.

## 1 Introduction

Terrorism destroys material objects and human lives, it causes pain and horror. The motivation is often to make the enemy give up something of value. For example, the terrorist group Irgun aimed to drive the British out of Palestine (Cohen [16]). But terrorism has another use: to rally support for the terrorists' cause. This idea is often associated with groups such as Hamas, the IRA, and the Basque separatist movement ETA. Terrorism may be most

*We thank Jim Jordan for useful early discussions and seminar audiences at the 2009 P.E.C.A. conference and NYU Politics for comments. We also thank Kane Sweeney for excellent research assistance. Any remaining errors are our responsibility.

effective in this regard when it provokes a violent response. According to *The Management of Savagery,*[1] Al Qaeda's objective is to provoke American attacks against the Islamic world, which will make moderate Muslims turn against the U.S. and its allies.

> "Force America to abandon its war against Islam by proxy and force it to attack directly so that the noble ones among the masses....will see that their fear of deposing the regimes because America is their protector is misplaced and that when they depose the regimes, they are capable of opposing America if it interferes." Abu Bakr Naji, *The Management of Savagery* ([1] p. 24)

Pacifists, on the other hand, want moderates to renounce all violence. For example, the Campaign for Nuclear Disarmament (C.N.D.) was initiated by Bertrand Russell during the Cold War. The objective of this "ban the bomb" movement was unilateral nuclear disarmament under the slogan "better red than dead".

> "If no alternative remains except Communist domination or the extinction of the human race, the former alternative is the lesser of two evils," Russell quoted in Rees [44]

These examples suggest that extremists send messages in order to influence decision-makers at home and abroad. These "messages" can be very

[1] This document, apparently composed by strategic thinkers within Al Qaeda, describes how a conflict with the Islamic world will destroy the American empire.

> "It is just as the American author Paul Kennedy says: 'If America expands the use of its military power and strategically extends more than necessary, this will lead to its downfall.' " (Naji [1], p. 18).

> "[N]ote that the economic weakness resulting from the burdens of war or from aiming blows of vexation (al-nikāya) directly toward the economy is the most important element of cultural annihilation since it threatens the opulence and (worldly) pleasures which those societies thirst for. Then competition for these things begins after they grow scarce due to the weakness of the economy. Likewise, social iniquities rise to the surface on account of the economic stagnation, which ignites political opposition and disunity among the (various) sectors of society in the central country." (Naji [1], p. 20).

costly to send (and receive). Indeed, according to the theory of the "propaganda of the deed", attributed to the Italian revolutionary Carlo Pisacane, messages *must* be costly (i.e., violence rather than words) in order to be effective (Hoffman [25], p. 5). Costless "cheap-talk" may drown in background noise without even being noticed by decision-makers. However, to understand the "pure" logic of extremist communication, a useful starting point is a model without noise, where effective communication does not have to be costly.

Following the literature, we will distinguish two kinds of conflicts.

> "World War I was an unwanted spiral of hostility...World War II was not an unwanted spiral of hostility-it was a failure to deter Hitler's planned aggression." Joseph Nye (p. 111, [39].).

Stag hunt and chicken are stylized representations of these two kinds of strategic interactions (Jervis [31]).[2] In stag hunt games, aggression feeds on itself and escalates into conflict, as in Hobbes's "state of nature" or Jervis's "spiralling model". Chicken is a model of preemption and deterrence, where toughness makes the opponent back down. We will study the ability of extremists to manipulate both kinds of conflicts.

The formal model is based on the conflict game of Baliga and Sjöström [2]. There are two countries, $A$ and $B$. In country $i \in \{A, B\}$, a decision-maker called player $i$ chooses a dovish action $D$ or a hawkish action $H$. Player $i$ may be interpreted as the median voter, or some other pivotal political decision-maker in country $i$. The hawkish action might be an act of war, accumulation of weapons, or any other aggressive action. It may involve selecting a hawkish agent who will take aggressive actions against the other country. For example, the median voters in Israel and Palestine have to decide whether to support Hamas or Fatah, or Likud or Kadima, respectively.

Player $i \in \{A, B\}$ can be a dominant strategy dove, a dominant strategy hawk, or a "moderate" whose best response depends on the opponent's action. Player $A$ doesn't know player $B$'s type, and vice versa. Baliga and Sjöström [2] showed how fear of the opponent can make moderates choose the hawkish action. Now our main purpose is to study how cheap-talk messages sent by extremists enhance or dampen this spiral of fear. In addition, we

[2] Baliga and Sjöström [4] show how strategic pre-commitments to influence bargaining over disputed territory can generate these two games.

generalize the model by allowing actions to be strategic substitutes as well as complements.

Strategic complements captures the logic of escalation, while strategic substitutes captures the logic of deterrence. Baliga and Sjöström [4] provide a formal model of how the commitment to costly conflict contained in the hawkish action determines if actions are strategic substitutes or complements. For example, suppose $H$ represents an invasion of a disputed territory. If player $i$ chooses $H$ and player $j$ chooses $D$, then player $i$ has an advantageous bargaining position, and is likely to end up with most of the disputed territory, while player $j$ gets very little. If nobody invades the disputed territory, then it is divided more equitably. Whether actions are strategic substitutes or complements is decided by what happens if both countries invade the disputed territory. If this means a high probability of a war which neither side wants, actions are strategic substitutes. But if the probability of a war is low, actions may be strategic complements instead. For details, see Baliga and Sjöström [4].

If the conflict game has strategic complements, then the moderates are "coordination types" who behave as in a stag hunt game: they want to match the action of the opponent. This can trigger an escalating spiral of fear, as in the classic work of Schelling [45] and Jervis [31]. But if the conflict game has strategic substitutes, then the moderates are "opportunists" (anti-coordination types) who behave as in a game of chicken: they choose $H$ if they think the opponent will choose $D$, but are intimidated and back down (choose $D$) if they believe the opponent will choose $H$. Whether actions are strategic complements or substitutes, fairly mild assumptions on the distribution of types guarantee that in the absence of cheap-talk the conflict game has a unique equilibrium, referred to as the *communication free* equilibrium.

In reality, extremist groups such as Hamas or the C.N.D. try to influence decision makers. Osama Bin Laden wants to provoke conflict between the U.S. and the Muslim world. But why would decision makers allow themselves to be manipulated ? To study this question, we add a third player called "the extremist" to the conflict game. Before players $A$ and $B$ make their decisions, the extremist sends a publicly observed cheap-talk message. The extremist should be thought of as the leader of an extremist movement located in, or with influence in, country $A$. The extremist's true preferences are commonly known. We consider two cases: a hawkish extremist ("terrorist") who wants player $A$ to choose $H$, and a dovish extremist ("pacifist") who wants player $A$ to choose $D$. (Both kinds of extremists want player $B$ to choose $D$.)

It is plausible that an extremist leader with influence in country $A$ knows something about the preferences of country $A$'s pivotal decision-maker. Extremists moving about the population may discover the opinion of the representative citizen (median voter). A political leader may be influenced by variables such as the state of the economy, the degree of religious fervor among the citizens, etc. The citizens themselves, including the extremists among them, may know more about these variables than outsiders. Finally, an extremist leader may know the extent to which his movement has been successful in directly influencing the pivotal decision-maker. To simplify the exposition, we assume the extremist in fact knows player $A$'s true type.

We are interested in equilibria where cheap-talk is effective, in the sense that the extremist's message influences the equilibrium decisions of players $A$ and $B$. Under fairly mild assumptions, there is a *unique* equilibrium with effective cheap-talk, referred to as the *communication equilibrium*. If cheap-talk is effective, then some message $m_1$ will make player $B$ more likely to choose $H$. A hawkish extremist is willing to send message $m_1$ only if player $A$ also becomes more likely to choose $H$. Such co-varying actions must be strategic complements. On the other hand, a dovish extremist is willing to send $m_1$ only if player $A$ becomes more likely to choose $D$. Such negative correlation occurs when actions are strategic substitutes. This argument implies that if the underlying game has strategic complements, then only a hawkish extremist can communicate effectively. By sending a "hawkish message", which we interpret informally as terrorism, the hawkish extremist triggers an unwanted (by players $A$ and $B$) spiral of fear and hostility, making both players $A$ and $B$ more likely to choose $H$. But if the underlying game has strategic substitutes, then only a dovish extremist can communicate effectively. By sending a "dovish message", which we interpret informally as a peace protest, the dovish extremist makes player $B$ more aggressive (i.e., more likely to choose $H$) and causes player $A$ to back down (choose $D$).

With strategic complements, terrorism occurs when player $A$ is a "weak moderate" who would have chosen $D$ in the communication-free equilibrium. Terrorism causes him to choose $H$ instead. In contrast, terrorism is clearly counter-productive if player $A$ is a dominant strategy hawk (who always chooses $H$ anyway). Thus, the absence of terrorism is actually "bad news" about player $A$'s type, in the sense that the conditional probability that player $A$ is a dominant strategy hawk increases.[3] This "bad news" makes player $B$ more likely to choose $H$ than in the communication-free equilibrium (although not as likely as following a terror act).

These arguments imply that, with strategic complements, players $A$ and $B$ are more likely to choose $H$ in the communication equilibrium (whether or not terrorism actually occurs) than in the communication-free equilibrium. Because each decision-maker always wants the other to choose $D$, the communication-free equilibrium interim Pareto dominates the communication equilibrium for players $A$ and $B$. Eliminating the hawkish extremist would make all types of players $A$ and $B$ strictly better off. This includes player $A$'s most hawkish types, whose preferences are actually aligned with the hawkish extremist. It is true that when the preferences are aligned in this way, the extremist will choose not to engage in terrorism, but this very decision alarms player $B$. Without the hawkish extremist, conflict would not be inflamed in this way.

When the underlying conflict game has strategic substitutes, it is only the pacifist (dovish extremist) who can communicate effectively. A peace rally occurs when player $A$ is a "tough moderate" who would have chosen $H$ in the communication-free equilibrium. The peace rally makes player $A$ choose $D$ instead, which makes player $B$ more willing to choose $H$. Thus, the communication equilibrium has a "better red than dead" flavour: following a peace demonstration in country $A$, player $B$ becomes more aggressive, and player $A$ backs down. In fact, whether or not a peace rally occurs, player $B$ is more likely to choose $H$ in the communication equilibrium than in the communication-free equilibrium, and this unambiguously makes player $A$ worse off. Thus, player $A$ would like to ban peace protests in his country if he could. On the other hand, because they induce player $A$ to choose $D$, peace protests make player $B$ better off.

Finally, we consider what happens if player $B$ can make (publicly observed) offensive or defensive investments before the conflict game is played. When the conflict game has strategic complements, player $B$ "over-invests"

[3] The fact that the *absence* of terrorism is informative is reminiscent of Sherlock Holmes's "curious incident of the dog in the night-time" (Conan Doyle [14]):

Gregory (Scotland Yard detective): "Is there any other point to which you would wish to draw my attention?"

Holmes: "To the curious incident of the dog in the night-time."

Gregory: "The dog did nothing in the night-time."

Holmes: "That was the curious incident."

is defensive capability in order to become "soft". With strategic substitutes, the strategic effect is more subtle. Intuition suggests that it is optimal to invest in offensive rather than defensive weapons, in order to convince the opponent to back down. This intuition is not valid in the presence of a dovish extremist. When player $B$'s defensive capability increases, the dovish extremist in country $A$ becomes more inclined to engage in peace protests, and as we have seen, this is good for player $B$. As a result, player $B$ actually over-invests in defensive capability even with strategic substitutes.

A number of articles have studied signaling and terrorism. Most closely related to our work is Jung [32], who also considers communication by a third party (a hawkish "Ministry of Propaganda") in a version of the Baliga and Sjöström [2] model. In Jung's model, messages are not cheap-talk: the Ministry of Propaganda cares about maintaining a reputation for being accurate, so its payoff depends directly on the messages it sends. The leader of one country has two possible types, and the Ministry of Propaganda knows the true type, while the other leader has only one possible type. In the absence of the third party, there would be multiple equilibria. Communication serves to refine the set of equilibria, and for this purpose it is crucial that messages are not cheap-talk. In equilibrium, however, communication is not effective in our sense: both leaders choose $H$ regardless of type (which is also an equilibrium outcome in the absence of the third party). In contrast, we study equilibria with effective cheap-talk, which do not replicate the outcome of any communication-free equilibrium. This requires two-sided incomplete information and a richer type-space.

Bueno de Mesquita and Dickinson [12] and de Figueiredo and Weingast [18] develop models of provocation based on signalling. Kydd and Walter [35] study "spoiling" where terrorists force an opponent to exit peace negotiations. For these authors, it is important that extremists' actions have type-dependent costs, as in the classic literature on signaling games (Spence [46]). In contrast, we show how extremists can manipulate decision-makers by pure cheap-talk.

The seminal paper on cheap-talk is Crawford and Sobel [17]. In the language of the cheap-talk literature, our model has one sender and multiple receivers. In previous work on such models (Farrell and Gibbons [21], Goltsman and Pavlov [24]) there is no strategic interaction between the receivers, which is the main focus of our work. Many articles study cheap-talk in two-player games, with no third party trying to manipulate the outcome.

For example, Farrell and Gibbons [22] and Matthews and Postlewaite [38] study cheap-talk before bargaining and auctions. Ordershook and Palfrey [40] study the impact of debate before voting and agenda-setting. Matthews [37] gives veto power to the sender and finds, like we do, that at most two messages are sent in equilibrium.

Depending on whether actions are strategic substitutes or complements, violence can either deter or lead to more violence. A growing empirical literature on the Israeli-Palestinian conflict addresses this point, although the findings are not very conclusive. Jaeger and Paserman ([27], [28], [29]) find that Palestinian violence or suicide attacks lead to increased violence by Israel, but Israeli violence either has no effect or possibly a deterrent effect. Jaeger et al. [30] find that major events in the conflict, such as the First Intifada, radicalized young Palestinians, but more moderate Israeli violence does not have a permanent effect.

There is a vast literature on terrorism which is less related to our work, including studies on the link between economic conditions and terrorism (e.g., Krueger [34]), the link between the quality of terrorist recruits and the state of the economy (Berrebi [9], Bueno de Mesquita [10], Benmelech and Berrebi [5], Benmelech et al. [6]), public goods provision by terrorist organizations (Berman [7], Iannaccone and Berman [26], Berman and Laitin [8]), and the optimal choice of targets for terrorism and counter-terrorism (Enders and Sandler [20], Bueno de Mesquita [11] and Powell [42] and [43]). Bueno de Mesquita's [13] provides an excellent survey of these and other issues.

# 2 The Model

## 2.1 The Conflict Game without Cheap Talk

Two decision makers, players $A$ and $B$, simultaneously choose either a hawkish (aggressive) action $H$ or a dovish (peaceful) action $D$. As mentioned in the introduction, we interpret player $i \in \{A, B\}$ as the pivotal political decision-maker in country $i$. The payoff for player $i \in \{A, B\}$ is given by the following payoff matrix, where the row represents his own choice, and the column represents the choice of player $j \neq i$.

$$\begin{array}{ccc} & H & D \\ H & -c_i & \mu - c_i \\ D & -d & 0 \end{array} \tag{1}$$

We assume $d > 0$ and $\mu > 0$, so player $j$'s aggression imposes a cost on player $i$. For simplicity, $d$ and $\mu$ are the same for each player. Notice that $d$ captures the cost of being caught out when the opponent is aggressive, while $\mu$ represents a benefit from being more aggressive than the opponent. The game has *strategic complements* if $d > \mu$ and *strategic substitutes* if $d < \mu$.

Player $i \in \{A, B\}$ has a cost $c_i$ of taking the hawkish action, referred to as his "type". Neither player knows the other player's type. The two types $c_A$ and $c_B$ are random variables independently drawn from the same distribution. Let $F$ denote the cumulative distribution function, with support $[\underline{c}, \overline{c}]$. Assume $F$ is differentiable, with $F'(c) > 0$ for all $c \in (\underline{c}, \overline{c})$ . Notice that the two players are symmetric *ex ante* (before their types are drawn). When taking an action, player $A$ knows $c_A$ but not $c_B$, while player $B$ knows $c_B$ but not $c_A$.

Player $i$ is a *dominant strategy hawk* if $H$ is a dominant strategy ($\mu \geq c_i$ and $d \geq c_i$ with at least one strict inequality). Player $i$ is a *dominant strategy dove* if $D$ is a dominant strategy ($\mu \leq c_i$ and $d \leq c_i$ with at least one strict inequality). Player $i$ is a *coordination type* if $H$ is a best response to $H$ and $D$ a best response to $D$ ($\mu \leq c_i \leq d$). Player $i$ is an *opportunistic type* if $D$ is a best response to $H$ and $H$ a best response to $D$ ($d \leq c_i \leq \mu$). Notice that coordination types exist only in games with strategic complements, and opportunistic types exist only in games with strategic substitutes. Assumption 1 states that the support of $F$ is big enough to include dominant strategy types of both kinds.

**Assumption 1** If the game has strategic complements then $\underline{c} < \mu < d < \overline{c}$. If the game has strategic substitutes then $\underline{c} < d < \mu < \overline{c}$.

The possibility that the opponent might be a dominant strategy type creates a spiral or multiplier effect. With strategic complements, the possibility that the opponent is a dominant strategy hawk causes coordination types who are "almost dominant strategy hawks" to play $H$. This in turn causes "almost-almost dominant strategy hawks" to play $H$, and an escalating spiral of aggression triggers further aggression. Strategic substitutes generates a very different spiral. Opportunistic types with a cost close to $d$ are "almost dominant strategy doves". The possibility that the opponent is a dominant strategy hawk makes these "almost dominant strategy doves" back off and play $D$. This emboldens opportunistic types who are "almost dominant strategy hawks" to play $H$, and so on.

To formalize this argument, suppose player $i$ thinks player $j$ will choose $H$ with probability $p_j$. Player $i$'s expected payoff from playing $H$ is $-c_i + \mu(1-p_j)$, while his expected payoff from $D$ is $-p_j d$. Thus, if he chooses $H$ instead of $D$, his *net* gain is

$$\mu - c_i + (d-\mu)p_j \tag{2}$$

A *strategy* for player $i$ is a function $\sigma_i : [\underline{c}, \bar{c}] \to \{H, D\}$ which specifies an action $\sigma_i(c_i) \in \{H, D\}$ for each cost type $c_i \in [\underline{c}, \bar{c}]$. In Bayesian Nash equilibrium (BNE), all types maximize their expected payoff. Therefore, $\sigma_i(c_i) = H$ if the expression in (2) is positive, and $\sigma_i(c_i) = D$ if it is negative. If expression (2) is zero then type $c_i$ is indifferent, but for convenience we will assume he chooses $H$ in this case.

Player $i$ uses a *cutoff strategy* if there is a *cutoff point* $x \in [\underline{c}, \bar{c}]$ such that $\sigma_i(c_i) = H$ if and only if $c_i \leq x$. Because the expression in (2) is monotone in $c_i$, all BNE must be in cutoff strategies. Therefore, it is without loss of generality to restrict attention to cutoff strategies. Any such strategy can be identified with its cut-off point $x \in [\underline{c}, \bar{c}]$. As there are dominant strategy doves and hawks by Assumption 1, all BNE must be interior: each player chooses $H$ with probability strictly between 0 and 1.

If player $j$ uses cutoff point $x_j$, the probability he plays $H$ is $p_j = F(x_j)$. Therefore, using (2), player $i$'s best response to player $j$'s cutoff $x_j$ is to choose the cutoff $x_i = \Gamma(x_j)$, where

$$\Gamma(x) \equiv \mu + (d-\mu)F(x). \tag{3}$$

The function $\Gamma$ is the best-response function for cutoff strategies. If there is "enough uncertainty", then the spirals that underlie the best-response function generate a unique equilibrium. This is ensured by Assumption 2.

**Assumption 2** $F'(c) < \frac{1}{d-\mu}$ for all $c \in (\underline{c}, \bar{c})$.

If $F$ happens to be uniform, then there is maximal uncertainty (for a given support) and Assumption 2 is redundant. More precisely, with a uniform distribution, $F'(c) = 1/(\bar{c} - \underline{c})$, so Assumption 1 implies $F'(c) < |\frac{1}{d-\mu}|$. Of course, Assumption 2 is much weaker than uniformity.[4]

**Theorem 1** *The conflict game without cheap-talk has a unique Bayesian Nash equilibrium.*

**Proof.** Equilibria must be in cutoff strategies, and must be interior by Assumption 1. The best response function $\Gamma$, defined by (3), is continuous, with $\Gamma(\underline{c}) = \mu > \underline{c}$ and $\Gamma(\bar{c}) = d < \bar{c}$, so it has a fixed-point $\hat{x} \in [\underline{c}, \bar{c}]$. If each player uses cut-off $\hat{x}$, the strategies form a BNE. It remains to show this BNE is unique. Notice that $\Gamma'(x) = (d - \mu)F'(x)$, so the best response function is upward (downward) sloping if actions are strategic complements (substitutes). In either case, a well-known sufficient condition for uniqueness is that best-response functions have slope strictly less than one in absolute value.[5] Assumption 2 implies that $0 < \Gamma'(x) < 1$ if $d > \mu$ and $-1 < \Gamma'(x) < 0$ if $d < \mu$. Hence, the best-response functions cross at most once and there is a unique equilibrium. ■

Proposition 1 shows that there exists a unique BNE, which we refer to as the communication-free BNE, whether actions are strategic substitutes or strategic complements (as long as Assumptions 1 and 2 hold). In equilibrium, player $i$ chooses $H$ if $c_i < \hat{x}$, where $\hat{x}$ is the unique fixed point of $\Gamma(x)$ in $[\underline{c}, \bar{c}]$ (see Figure 1 for the case of strategic complements). The symmetry of the game implies that both players use the same cutoff point.

## 2.2 Cheap-Talk

We now introduce a third player, player $E$. Player $E$ is the leader of an extremist group with influence in country $A$. His payoff function is similar to player $A$'s, with one exception: player $E$'s cost type $c_E$ differs from player

[4] Assumption 2 is violated if the type distribution is highly concentrated around one point. In this case, multiple equilibria can easily exist, even if Assumption 1 holds. Notice that we are assuming types are independent. Since the complete information chicken and stag hunt games have multiple equilbria, a small amount of idiosycratic noise, as in Harsanyi's purification argument, will not refine the set of equilbria.

[5] This condition is familiar from the IO literature. With upward-sloping best-response functions, as in Bertrand competition with product differentiation, the slope should be less than one. With downward-sloping best-response functions, as in Cournot competition, the slope should be greater than negative one. See Vives [47].

$A$'s cost type $c_A$. Thus, player $E$'s payoff is obtained by setting $c_i = c_E$ in the payoff matrix (1), and letting the row represent player $A$'s choice and the column player $B$'s choice. There is no uncertainty about $c_E$. Formally, $c_E$ is common knowledge among the three players.

Player $E$ knows $c_A$ but not $c_B$. Intuitively, the extremist leader might receive some signal which indicates how much influence his group has over the pivotal political decision-maker in country $A$, or more generally about the politics or economics of country $A$. To avoid unnecessary complications, we assume the signal is perfect, so player $E$ knows $c_A$.

We consider two possibilities. First, if player $E$ is a *hawkish extremist* ("terrorist"), then $c_E < 0$. To put it differently, $(-c_E) > 0$ represents a *benefit* the hawkish extremist enjoys if player $A$ is aggressive. The hawkish extremist is guaranteed a strictly positive payoff if player $A$ chooses $H$, but he gets a non-positive payoff when player $A$ chooses $D$, so he always wants player $A$ to choose $H$. Second, if player $E$ is a *dovish extremist* ("pacifist"), then $c_E > \mu + d$. The most the dovish extremist can get if player $A$ chooses $H$ is $\mu - c_E$, while the worst he can get when player $A$ chooses $D$ is $-d > \mu - c_E$, so he always wants player $A$ to choose $D$. Notice that, holding player $A$'s action fixed, the extremist (whether hawkish or dovish) is better off if player $B$ chooses $D$.

Before players $A$ and $B$ play the conflict game described in Section 2.1, player $E$ sends a publicly observed cheap-talk message $m \in M$, where $M$ is his message space. We interpret the message as some kind of demonstration which, for simplicity, has zero real cost. In reality, the demonstration (e.g. a terrorist act) may be costly. Perhaps this is necessary to "get the attention" of real-world decision-makers. Costs incurred by terrorist victims are irrelevant to our argument, because player $E$ is not assumed to internalize these costs. More relevant are costs incurred by $E$. However, player $E$ is willing to incur a cost to influence the outcome of the game, so unless these costs are prohibitively big, they do not change the nature of our arguments. Assuming messages are pure cheap-talk (with zero cost) helps clarify how player $E$ can manipulate the outcome of the conflict game.

The time line is as follows.

1. The cost type $c_i$ is determined for each player $i \in \{A, B\}$. Players $A$ and $E$ learn $c_A$. Player $B$ learns $c_B$.

2. Player $E$ sends a (publicly observed) cheap-talk message $m \in M$.

3. Players $A$ and $B$ simultaneously choose $H$ or $D$.

Cheap-talk is *effective* if there is a positive measure of types that choose different actions at time 3 than they would have done in the unique communication-free equilibrium of Section 2.1. A Perfect Bayesian Equilibrium (PBE) with effective cheap-talk is a *communication equilibrium.* Clearly, if players $A$ and $B$ maintain their prior beliefs at time 3, then they must act just as in the unique communication-free equilibrium. Therefore, for cheap-talk to be effective, player $E$'s message must reveal some information about player $A$'s type.

A strategy for player $E$ is a function $m : [\underline{c}, \bar{c}] \rightarrow M$, where $m(c_A)$ is the message sent by player $E$ when player $A$'s type is $c_A$. Without loss of generality, we assume each player $j \in \{A, B\}$ uses a "conditional" cut-off strategy: for any message $m \in M$, there is a cut-off $c_j(m)$ such that if player $j$ hears message $m$, then he chooses $H$ if and only if $c_j \leq c_j(m)$.

**Lemma 1** *In communication equilibrium, it is without loss of generality to assume that $M$ contains only two messages, $M = \{m_0, m_1\}$, where $c_B(m_1) > c_B(m_0)$.*

**Proof.** Suppose strategy $\mu$ is part of a BNE. Because unused messages can simply be dropped, we may assume that for any $m \in M$, there is $c_A$ such that $m(c_A) = m$. Now consider any two messages $m$ and $m'$. If $c_B(m) = c_B(m')$, then the probability player $B$ plays $H$ is the same after $m$ and $m'$, and this means each type of player $A$ also behaves the same after $m$ as after $m'$. Clearly, if all players behave the same after $m$ and $m'$, having two separate messages $m$ and $m'$ is redundant. Hence, without loss of generality, we can assume $c_B(m) \neq c_B(m')$ whenever $m \neq m'$.

Whenever player $A$ is a dominant strategy type, player $E$ will send whatever message minimizes the probability that player $B$ plays $H$. Call this message $m_0$. Thus,

$$m_0 = \arg\min_{m \in M} c_B(m) \tag{4}$$

Message $m_0$ is the *unique* minimizer of $c_B(m)$, since (by the previous paragraph) $c_B(m) \neq c_B(m')$ whenever $m \neq m'$.

Player $E$ cannot always send $m_0$, because then messages would not be informative and cheap-talk would be ineffective (contradicting the definition of communication equilibrium). But, since message $m_0$ uniquely maximizes the probability that player $B$ chooses $D$, player $E$ must have some other reason for choosing $m(c_A) \neq m_0$. Specifically, if player $E$ is a hawkish extremist (who wants player $A$ to choose $H$) then it must be that type $c_A$ would choose $D$ following $m_0$ but $H$ following $m(c_A)$; if player $E$ is a dovish extremist (who wants player $A$ to choose $D$) then it must be that type $c_A$ would choose $H$ following $m_0$ but $D$ following $m(c_A)$. This is the only way player $E$ can justify sending any other message than $m_0$.

Thus, if player $E$ is a hawkish extremist, then whenever he sends a message $m_1 \neq m_0$, player $A$ will play $H$. Player $B$ therefore responds with $H$ whenever $c_B < d$. That is, $c_B(m_1) = d$. But $c_B(m) \neq c_B(m')$ whenever $m \neq m'$, so $m_1$ is unique. Thus, $M = \{m_0, m_1\}$.

Similarly, if player $E$ is a dovish extremist, then whenever he sends a message $m_1 \neq m_0$, player $A$ will play $D$. Player $B$'s cutoff point must therefore be $c_B(m_1) = \mu$. Again, this means $M = \{m_0, m_1\}$. ■

Notice that this lemma holds for both strategic substitutes and strategic complements, and for both dovish and hawkish extremists. It also does not invoke Assumption 2.

# 3 Cheap-Talk with Strategic Complements

In this section, we consider the case of strategic complements, $d > \mu$.

## 3.1 Doves can't Communicate Effectively

We first show that if player $E$ is a dovish extremist, $c_E > \mu + d$, then he cannot communicate effectively when actions are strategic complements. From Lemma 1, $M = \{m_0, m_1\}$ with $c_B(m_1) > c_B(m_0)$. Thus, player $B$ is more likely to choose $H$ after $m_1$ than after $m_0$. The dovish extremist wants both players $A$ and $B$ to play $D$, so he would only choose $m_1$ if this message causes player $A$ to play $D$. Formally, if $m(c_A) = m_1$, then we must have $c_A > c_A(m_1)$, so that type $c_A$ chooses $D$ when he hears message $m_1$. But if $c_A > c_A(m_1)$ for all $c_A$ such that $m(c_A) = m_1$, then player $B$ expects player $A$ to play $D$ for sure when player $B$ hears $m_1$, so player $B$'s cut-off point must be $c_B(m_1) = \mu$. But, with $d > \mu$, types below $\mu$ are dominant strategy types who always play $H$, so we cannot have $c_B(m_0) < \mu$, a contradiction. Thus, we have:

**Proposition 1** *If player $E$ is a dovish extremist and the game has strategic complements, then cheap-talk cannot be effective.*

When actions are strategic complements, the message $m_1$ which makes player $B$ more likely to play $H$ must also make player $A$ more likely to play $H$. But a message which triggers such a spiral of fear and hostility will never be sent by a dovish extremist, and this makes the dovish extremist unable to communicate effectively. Hence, there is no credible way for the dovish extremist to influence conflict. In particular, he cannot increase the possibility of "peace", the action profile $DD$.

## 3.2 Hawkish Cheap-Talk

Now suppose player $E$ is a hawkish extremist, $c_E < 0$, and the game has strategic complements. We will construct a communication equilibrium, where the hawkish extremist $E$ uses cheap-talk to increase the risk of conflict above the level of the communication-free equilibrium of Section 2.1. It is surprising that player $E$ can do this, because $c_E$ is commonly known. That is, it is commonly known that player $E$ wants player $B$ to choose $D$ and player $A$ to choose $H$. To understand the equilibrium intuitively, it helps to recall that $M = \{m_0, m_1\}$ by Lemma 1, where $c_B(m_1) > c_B(m_0)$, and interpret message $m_1$ as "terrorism" and message $m_0$ as "no terrorism".

Say that player $A$ is a *susceptible type* if he chooses $H$ following message $m_1$, but $D$ following $m_0$. The set of susceptible types is

$$S \equiv (c_A(m_0), c_A(m_1)].$$

The proof of Lemma 1 showed that if $m(c_A) = m_1$ then type $c_A$ must be susceptible. Since terrorism makes player $B$ more likely to choose $H$, player $E$ will only engage in terrorism if it causes player $A$ to change his action from $D$ to $H$. On the other hand, player $E$ wants player $A$ to choose $H$ and therefore strictly prefers to engage in terrorism whenever player $A$ is susceptible. That is, it is optimal for player $E$ to set $m(c_A) = m_1$ if and only if $c_A \in S$. Accordingly, message $m_1$ signals that player $A$ will choose $H$. As argued in the proof of Lemma 1, this implies $c_B(m_1) = d$. Therefore, if $m_1$ is sent then player $B$ will choose $H$ with probability $F(d)$, so player $A$ prefers $H$ if and only if

$$-c_A + (1 - F(d))\mu \geq F(d)(-d)$$

which is equivalent to $c_A \leq \Gamma(d)$. Thus, player $A$ uses cut-off point $c_A(m_1) = \Gamma(d)$, where $\Gamma$ is defined by (3).

It remains only to consider how players $A$ and $B$ behave when there is no terrorism (message $m_0$). Let $y^* = c_A(m_0)$ and $x^* = c_B(m_0)$ denote the cutoff points in this case. Thus, if $m_0$ is sent then player $B$ will choose $H$ with probability $F(x^*)$, so player $A$ prefers $H$ if and only if

$$-c_A + (1 - F(x^*))\mu \geq F(x^*)(-d)$$

which is equivalent to $c_A \leq \Gamma(x^*)$. Thus, $y^* = \Gamma(x^*)$. When player $B$ hears message $m_0$, he knows that player $A$ is not a susceptible type. That is, $c_A$ is either below $y^*$ or above $\Gamma(d)$, and player $A$ chooses $H$ in the former case and $D$ in the latter case. Therefore, player $B$ prefers $H$ if and only if

$$-c_B + \frac{1 - F(\Gamma(d))}{1 - F(\Gamma(d)) + F(y^*)}\mu \geq \frac{F(y^*)}{1 - F(\Gamma(d)) + F(y^*)}(-d) \tag{5}$$

Inequality (5) is equivalent to $c_B \leq \Omega(y^*)$, where

$$\Omega(y) \equiv \frac{[1 - F(\Gamma(d))]\,\mu + F(y)d}{[1 - F(\Gamma(d))] + F(y)}$$

Thus, $x^* = \Omega(y^*)$.

To summarize, any communication equilibrium must have the following form. Player $E$ sets $m(c_A) = m_1$ if and only if $c_A \in S = (y^*, \Gamma(d)]$. Player $A$'s cut-off points are $c_A(m_0) = y^*$ and $c_A(m_1) = \Gamma(d)$. Player $B$'s cut-off points are $c_B(m_0) = x^*$ and $c_B(m_1) = d$. Moreover, $x^*$ and $y^*$ must satisfy $y^* = \Gamma(x^*)$ and $x^* = \Omega(y^*)$. Conversely, if such $x^*$ and $y^*$ exist, then they define a communication equilibrium. We now show graphically that they do exist.

By Assumption 2, $\Gamma$ is increasing with a slope less than one. Since $F(\underline{c}) = 0$ and $F(\bar{c}) = 1$, we have $\Gamma(\underline{c}) = \mu > \underline{c}$ and $\Gamma(\bar{c}) = d < \bar{c}$. Furthermore,

$$\Gamma(d) - \mu = F(d)\,(d - \mu) < d - \mu.$$

Therefore,

$$\Gamma(d) < d. \tag{6}$$

Also,

$$\Gamma(\mu) = \mu(1 - F(\mu)) + dF(\mu) > \mu$$

as $d > \mu$. Let $\hat{x}$ be the unique fixed point of $\Gamma(x)$ in $[\underline{c}, \overline{c}]$. Clearly, $\mu < \hat{x} < \Gamma(d)$ (see Figure 2).

Figure 2 shows three curves: $x = \Omega(y)$, $y = \Gamma(x)$ and $x = \Gamma(y)$. The curves $x = \Gamma(y)$ and $y = \Gamma(x)$ intersect on the 45 degree line at the unique fixed point $\hat{x} = \Gamma(\hat{x})$. Notice that

$$\Omega'(y) = \frac{F'(y)\,(d - \mu)\,(1 - F(\Gamma(d)))}{([1 - F(\Gamma(d))] + F(y))^2}$$

so $\Omega$ is increasing. It is easy to check that $\Omega(y) > \Gamma(y)$ whenever $y \in (\underline{c}, \Gamma(d))$. Moreover,

$$\Omega(\underline{c}) = \Gamma(\underline{c}) = \mu$$

and

$$\Omega(\Gamma(d)) = \Gamma(\Gamma(d)) < \Gamma(d)$$

where the inequality follows from (6) and the fact that $\Gamma$ is increasing. These properties are shown in Figure 2. Notice that the curve $x = \Omega(y)$ lies to the right of the curve $x = \Gamma(y)$ for all $y$ such that $\underline{c} < y < \Gamma(d)$ (because $\Omega(y) > \Gamma(y)$ for such $y$), but the two curves intersect when $y = \underline{c}$ and $y = \Gamma(d)$.

As shown in Figure 2, the two curves $x = \Omega(y)$ and $y = \Gamma(x)$ must intersect at some $(x^*, y^*)$, and it must be true that

$$\hat{x} < y^* < x^* < \Gamma(d) < d \tag{7}$$

By construction, $y^* = \Gamma(x^*)$ and $x^* = \Omega(y^*)$. Thus, a communication equilibrium exists.

Both player $A$ and player $B$ are strictly more likely to choose $H$ in communication equilibrium than in communication-free equilibrium. To see this, notice that in the communication-free equilibrium, each player's cutoff is $\hat{x}$. By (7), the cut-off points are strictly higher in communication equilibrium, whether or not terrorism occurs. Thus, whenever a player would have chosen $H$ in the communication-free equilibrium, he necessarily chooses $H$ in communication equilibrium.. Moreover, after any message, there are types (of each player) that choose $H$, but who would have chosen $D$ in the communication-free equilibrium. It follows that all types of players $A$ and $B$ are made worse off by communication, because each prefers the opponent to choose $D$.

For player $E$, the welfare comparison across equilibria is ambiguous, because cheap-talk makes both players $A$ and $B$ more likely to choose $H$. Specifically, there are three cases. First, if either $c_A \leq \hat{x}$ or $c_A > \Gamma(d)$, then player $A$'s action is the same in the communication equilibrium and in the communication-free equilibrium, but player $B$ is more likely to choose $H$ in the former, making player $E$ worse off. Second, if $\hat{x} < c_A \leq y^*$, then player $A$ would have chosen $D$ in the communication-free equilibrium. In the communication equilibrium, there is no terrorism when $\hat{x} < c_A \leq y^*$, but player $A$ plays $H$ rather than $D$, because player $B$ is likely to choose $H$ (the "dog that doesn't bark" effect). Third, if $y^* < c_A \leq \Gamma(d)$, then terrorism causes player $A$ to play $H$, rather than $D$ as in the communication free equilibrium. Player $E$ gets a strictly positive payoff whenever player $A$ chooses $H$, and a non-positive payoff whenever player $A$ chooses $D$. Thus, player $E$ is better off if player $A$ switches to $H$.

The communication equilibrium is *unique* if the two curves $x = \Omega(y)$ and $y = \Gamma(x)$ have a *unique* intersection. This would be true, for example, if $F$ were concave, because in this case both $\Omega$ and $\Gamma$ would be concave. However, uniqueness also obtains without concavity, if Assumption 2 is strengthened. Assumption 2 implies $0 < \Gamma'(x) < 1$. It can be checked that if $F'(y) < \frac{1-F(\Gamma(d))}{d-\mu}$ then $0 < \Omega'(y) < 1$. In this case, the two curves $x = \Omega(y)$ and $y = \Gamma(x)$ intersect only once, as indicated in Figure 2.

In summary:

**Theorem 2** *Suppose player $E$ is a hawkish extremist and the game has strategic complements. A communication equilibrium exists. All types of players $A$ and $B$ prefer the communication-free equilibrium to the communication equilibrium. Player $E$ is better off in the communication equilibrium if and only if $\hat{x} < c_A \leq \Gamma(d)$. If $F'(c) < \frac{1-F(\Gamma(d))}{d-\mu}$ for all $c \in (\underline{c}, \overline{c})$ then the communication equilibrium is unique.*

In the communication-free equilibrium, the probability of peace, in the sense that the outcome is $DD$, is $(1 - F(\hat{x}))^2$. In the communication equilibrium, $DD$ happens with probability $(1 - \Gamma(d))(1 - F(x^*)) < (1 - F(\hat{x}))^2$. Thus, peace is less likely in the communication equilibrium than in the communication-free equilibrium.

To understand how the cut-off points can be uniformly higher with cheap-talk, we again interpret message $m_1$ as "terrorism" and message $m_0$ as "no terrorism". Terrorism occurs when player $A$ is a coordination type $c_A \in$

$[y^*, \Gamma(d)]$ who would have played $D$ in the communication-free equilibrium. Now, he plays $H$ instead, and so does player $B$ (except if he is a dominant strategy dove). The players behave aggressively following terrorism because they think the other will be aggressive, as in a "bad" equilibrium of a stag-hunt game. The fact that terrorism does *not* occur also triggers conflict, but for a different reason. In "the curious incident of the dog in the night-time" (Conan Doyle [14]), the dog did not bark at an intruder because the dog knew him well. Similarly, when player $A$'s preferences are aligned with the hawkish extremist, there is no terrorism. Hence, a terrorist who "does not bark" signals the possibility that player $A$ is a dominant strategy hawk. This information makes player $B$ want to play $H$. Accordingly, the communication equilibrium has more conflict than the communication-free equilibrium, no matter which message is sent.

There is a stark contrast between the results in Baliga and Sjöström [2], where cheap-talk *between the decision-makers* was shown to prevent conflict, and the current results. In both cases, cheap-talk truncates the distribution of types, with a separate message sent for intermediate types and another for extreme types. In Baliga and Sjöström [2], the intermediate types are "tough" coordination types. Separating them from the rest of the distribution cuts the spiral and prevents the population from being infected by fearfulness. Moreover, since the intermediate types are not dominant strategy types, they can coexist peacefully. But communication by a hawkish extremist separates out "weak" coordination types, who play $D$ in the communication-free equilibrium but $H$ in the communication equilibrium. This brings conflict when peace could have prevailed. When there is no terrorism in the communication equilibrium, the spiralling logic is even worse than before, because the absence of "weak" coordination types leads to a less favorable type-distribution.

# 4 Cheap-Talk with Strategic Substitutes

In this section, we consider the case of strategic substitutes, $d < \mu$.

## 4.1 Hawks can't Communicate Effectively

A hawkish extremist cannot communicate effectively when actions are strategic substitutes. From Lemma 1, $M = \{m_0, m_1\}$ with $c_B(m_1) > c_B(m_0)$. The hawkish extremist wants player $A$ (but not player $B$) to play $H$, so he would only send $m_1$ if this message causes player $A$ to play $H$. But if player $A$ plays $H$ for sure after $m_1$, then player $B$'s cut-off point is $c_B(m_1) = d$. But, with $d < \mu$, types below $d$ are dominant strategy types who always play $H$, so we cannot have $c_B(m_0) < d$, a contradiction. Thus, we have:

**Proposition 2** *If player $E$ is a hawkish extremist and the game has strategic substitutes, then cheap-talk cannot be effective.*

When actions are strategic substitutes, the message $m_1$ which makes player $B$ more likely to play $H$ must make player $A$ more likely to play $D$. But a message which causes player $A$ to back down in this way will never be sent by a hawkish extremist, and this makes the hawkish extremist unable to communicate effectively.

## 4.2 Dovish Cheap-Talk

Now suppose player $E$ is a dovish extremist and the game has strategic substitutes. We will construct a communication equilibrium where the dovish extremist $E$ sends informative messages. Again, it is surprising that this can be done because $c_E$ is commonly known. To understand the communication equilibrium intuitively, it helps to again recall Lemma 1, but now interpret message $m_1$ as a "peace rally" and message $m_0$ as "no peace rally". Intuitively, the peace rally will make player $B$ more aggressive, and player $A$ backs down and chooses $D$.

Again, say that player $A$ is a susceptible type if his action depends on which message is sent. But now, susceptible types switch from $H$ to $D$ when they hear message $m_1$. That is, the set of susceptible types is

$$S \equiv (c_A(m_1), c_A(m_0)].$$

The proof of Lemma 1 showed that if $m(c_A) = m_1$ then type $c_A$ must be susceptible. Intuitively, since peace demonstrations make player $B$ more likely to choose $H$, player $E$ would not engage in them unless player $A$ is a susceptible type. Conversely, whenever player $A$ is a susceptible type, the dovish extremist will engage in peace demonstrations, since he wants player $A$ to choose $D$. Therefore, $m(c_A) = m_1$ if and only if $c_A \in S$. Accordingly, message $m_1$ signals that player $A$ will choose $D$. As argued in the proof of Lemma 1, this implies $c_B(m_1) = \mu$, and player $A$'s best response to this cut-off point is $c_A(m_1) = \Gamma(\mu)$.

It remains only to consider how players $A$ and $B$ behave when there is no peace demonstration (message $m_0$). Let $y^* = c_A(m_0)$ and $x^* = c_B(m_0)$ denote the cutoff points used in this case. Arguing as for the case of strategic complements, the cut-off points must satisfy $y^* = \Gamma(x^*)$ and $x^* = \tilde{\Omega}(y^*)$, where

$$\tilde{\Omega}(y) \equiv \frac{[1 - F(y)]\,\mu + F(\Gamma(\mu))d}{[1 - F(y)] + F(\Gamma(\mu))}$$

As shown in Figure 3, $(x^*, y^*)$ is an intersection of the two curves $x = \tilde{\Omega}(y)$ and $y = \Gamma(x)$. With strategic substitutes, Assumption 2 implies

$$-1 < \Gamma'(x) < 0$$

Furthermore, $\Gamma(\underline{c}) = \mu < \bar{c}$ and $\Gamma(\bar{c}) = d > \underline{c}$, and

$$\Gamma(\mu) - d = (1 - F(\mu))\,(\mu - d)$$

where

$$0 < (1 - F(\mu))\,(\mu - d) < \mu - d.$$

Therefore,

$$d < \Gamma(\mu) < \mu \tag{8}$$

Let $\hat{x}$ be the unique fixed point of $\Gamma(x)$ in $[\underline{c}, \bar{c}]$. Clearly, $d < \hat{x} < \mu$ (see Figure 3).

Figure 3 shows three curves: $x = \tilde{\Omega}(y)$, $y = \Gamma(x)$ and $x = \Gamma(y)$. The curves $x = \Gamma(y)$ and $y = \Gamma(x)$ intersect on the 45 degree line at the fixed point $\hat{x} = \Gamma(\hat{x})$. It is easy to check that $\tilde{\Omega}(y) > \Gamma(y)$ whenever $y \in (\Gamma(\mu), \bar{c})$. Moreover,

$$\tilde{\Omega}(\bar{c}) = \Gamma(\bar{c}) = d$$

and

$$\tilde{\Omega}(\Gamma(\mu)) = \Gamma(\Gamma(\mu)) > \Gamma(\mu)$$

where the inequality follows from (8) and the fact that $\Gamma$ is decreasing. Consider now $(x^*, y^*)$ such that $y^* = \Gamma(x^*)$ and $x^* = \tilde{\Omega}(y^*)$, i.e., the intersection of the two curves $x = \tilde{\Omega}(y)$ and $y = \Gamma(x)$. Figure 3 reveals that there exists $(x^*, y^*) \in [\underline{c}, \bar{c}]^2$ such that $y^* = \Gamma(x^*)$ and $x^* = \tilde{\Omega}(y^*)$, and

$$d < \Gamma(\mu) < y^* < \hat{x} < x^* < \mu. \tag{9}$$

Thus, a communication equilibrium exists. What impact do pacifist messages have on the probability of conflict? In the communication-free equilibrium, each player's cutoff is $\hat{x}$. Now (9) reveals that with pacifist communication, player $B$'s cutoff points $x^*$ and $\mu$ are strictly greater than $\hat{x}$. Thus, communication makes player $B$ more aggressive, whatever message is actually sent. On the other hand, player $A$'s cutoff points $y^*$ and $\Gamma(\mu)$ are strictly smaller than $\hat{x}$. Thus, communication makes player $A$ less aggressive ("better red than dead"), whatever message is actually sent. Since one player becomes more and the other less aggressive, it is not possible to unambiguously say if communication is good or bad for peace.

The welfare effects are unambiguous, however. As player $A$ is more likely to play $D$ in the communication equilibrium, player $B$ is made better off. Conversely, as player $B$ is more likely to play $H$, player $A$ is made worse off. The pacifist (dovish extremist) is made better off by the peace rally when it occurs, because it prevents player $A$ from choosing $H$. On the other hand, the "dog that did not bark" effect makes player $B$ more likely to choose $H$ when there is no peace rally, and this makes player $E$ worse off.

Finally, consider whether the communication equilibrium is unique. By Assumption 2, $-1 < \Gamma'(x) < 0$. It can be checked that if $F'(c) < \frac{F(\Gamma(d))}{\mu - d}$ then $-1 < \tilde{\Omega}'(y) < 0$. In this case, the two curves $x = \tilde{\Omega}(y)$ and $y = \Gamma(x)$ intersect only once, as indicated in Figure 3. In summary:

**Theorem 3** *Suppose player $E$ is a dovish extremist and the game has strategic substitutes. A communication equilibrium exists. All of player $A$'s types prefer the communication-free equilibrium to the communication equilibrium. All of player $B$'s types have the opposite preference. Player $E$ is better off in the communication equilibrium if and only if $\Gamma(\mu) \leq c_A < \hat{x}$. If $F'(c) < \frac{F(\Gamma(d))}{\mu - d}$ for all $c \in (\underline{c}, \overline{c})$ then the communication equilibrium is unique.*

Theorem 3 is in stark contrast to Theorem 2. With strategic complements, terrorism caused both players $A$ and $B$ to become more aggressive, and hence both became worse off. With strategic substitutes, player $B$ benefits from peace rallies in country $A$, because they make player $A$ back down.

# 5 Strategic Effects of Ex Ante Investment

Suppose a decision maker can make a publicly observed investment which changes his country's military capability. He might invest in weapons that increase the chances of military victory or result in less destructive wars. At the other extreme, he might invest in technology that shoots down enemy missiles or build fortifications than make an attack difficult. These investments make an attack less costly.

Intuitively, in a game of chicken, there is a incentive to overinvest in offensive capability in order to intimidate the opponent and force him to back down. In a stag-hunt game, there is a incentive to over-invest in defensive capability in order to reassure the opponent that one is unlikely to attack out of fear. These so-called strategic effects (Fudenberg and Tirole [23]) are easy to understand when no extremist exists. In this section, we will consider the more complex strategic effects when an extremist observes the investment and can react to it.

We first generalize the model to allow for *ex ante* asymmetries. The parameters $\mu$ and $d$, and the distribution over cost-types, are now player-dependent. The payoff of player $i \in \{A, B\}$ is given by the following payoff matrix, where the row represents his own choice, and the column represents the choice of player $j$.

$$\begin{array}{ccc} & H & D \\ H & -c_i & \mu_i - c_i \\ D & -d_i & 0 \end{array} \tag{10}$$

Player $i$'s type $c_i$ is drawn from a distribution $F_i$ with support $[\underline{c}_i, \bar{c}_i]$. As before, types are independently drawn. In the communication-free equilibrium, equilibrium cutoff points $(\hat{x}_A, \hat{x}_B)$ solve the two equations

$$\hat{x}_A = \mu_A + (d_A - \mu_A) F_B(\hat{x}_B) \tag{11}$$

$$\hat{x}_B = \mu_B + (d_B - \mu_B) F_A(\hat{x}_A) \tag{12}$$

If the obvious analog of Assumption 1 holds and if $F_i'(c_i) < \left| \frac{1}{d_i - \mu_i} \right|$ for $i \in \{A, B\}$ (the analog of Assumption 2), then the communication-free equilibrium is unique by the same argument as in Proposition 1.

Consider the strategic effects when no extremist is present. Suppose player $B$, at time 0, can make a publicly observed investment which increases $\mu_B$. This may represent, for example, increased offensive capability. Suppose after the investment, the communication-free equilibrium is played (as given by (11) and (12)). The investment increases player $B$'s benefit from choosing $H$, and hence makes player $B$ appear tough (it shifts his best response curve to the right). The strategic effect of the investment is its impact on the behavior of player $A$. Fudenberg and Tirole [23] classify strategic effects in four categories: Top Dog, Puppy Dog, Fat Cat, and Lean-and-Hungry-Look. These effects differ in whether investment makes a player "soft" or "tough" or whether there is an incentive to "overinvest" or "underinvest". With strategic complements, shifting player $B$'s best response curve to the right causes both $\hat{x}_A$ and $\hat{x}_B$ to increase. Since player $B$ wants player $A$ to choose $D$, the strategic effect is negative: player $B$ prefers to under-invest in order to appear soft (Puppy Dog strategy). With strategic substitutes, the strategic effect is instead positive: player $B$ then prefers to overinvest in order to appear tough (Top Dog strategy).

Suppose instead the investment reduces $d_B$. This may represent, for example, better defensive abilities of country $B$, making it less vulnerable to an attack. This investment will raise player $B$'s benefit from choosing $D$, and hence make player $B$ appear soft (it shifts his best response curve to the left). With strategic complements, both $\hat{x}_A$ and $\hat{x}_B$ decrease. Thus, the strategic effect is positive: player $B$ prefers to overinvest in order to appear soft (Fat Cat strategy). With strategic substitutes, the strategic effect is instead negative: player $B$ underinvests in order to appear tough (Lean and Hungry Look).

We now turn to strategic effects in the presence of an extremist. Observe that Lemma 1 is still valid in the asymmetric environment. In communication equilibrium, generalized to allow for ex ante asymmetries, player $B$'s publicly observed investment influences player $A$ not only directly but also indirectly, via changes in player $E$'s behavior. Nevertheless, with strategic complements, the strategic effects turn out to be the same as discussed above: the optimal strategies are still Puppy Dog and Fat Cat, making oneself look less threatening. However, with strategic substitutes, the presence of a dovish extremist dramatically changes the strategic effects. The dovish extremist is, in a sense, an "ally" of player $B$, because peace protests make player $A$ back down. In this case, Top Dog and Lean and Hungry Look strategies can backfire for player $B$: by investing in offensive capacity (or under-investing in defensive capacity), player $B$ alarms the pacifist, who organizes fewer peace protests. The net effect may be to make player $B$ worse off. We now formalize these arguments.

## 5.1 Strategic Complements

Suppose $d_i > \mu_i$ for $i \in \{A, B\}$ and player $E$ is a hawkish extremist. Define

$$\Gamma_A(x) \equiv \mu_A + F_B(x)\,(d_A - \mu_A) \tag{13}$$

and

$$\Omega_B(y) \equiv \frac{[1 - F_A(\Gamma_A(d_B))]\,\mu_B + F_A(y)d_B}{[1 - F_A(\Gamma_A(d_B))] + F_A(y)} \tag{14}$$

Now let $x_B^* = \Omega_B(y_A^*)$ and $y_A^* = \Gamma_A(x_B^*)$. Arguing as in Section 3.2, if $F_A'(c_A) < \frac{1-F_A(\Gamma(d))}{d-\mu}$ for all $c_A$ then there exists a unique pair $(x_B^*, y_A^*)$ such that $y_A^* = \Gamma_A(x_B^*)$ and $x_B^* = \Omega_B(y_A^*)$. Moreover, $\hat{x}_B < x_B^* < d_B$ and $\hat{x}_A < y_A^* < \Gamma_A(d_B) < d_A$.The strategies are the obvious generalizations of the strategies in Section 3.2. Player $E$ sends the message $m(c_A) = m_1$ if and only if $y_A^* < c_A \leq \Gamma_A(d_B)$. Player $A$'s cut-off points are $c_A(m_1) = \Gamma_A(d_B)$ and $c_A(m_0) = y_A^*$. Player $B$'s cut-off points are $c_B(m_0) = x_B^*$ and $c_B(m_1) = d_B$. Notice that, in equilibrium, player $A$ chooses $H$ if and only if $c_A \leq \Gamma_A(d_B)$.

Suppose player $B$, at time 0, makes a publicly observed investment which increases $\mu_B$. This shifts the $\Omega_B$ function to the right: player $B$ becomes "tough". Since $\Gamma_A(d_B)$ does not depend on $\mu_B$, the set of types of player $A$ that choose $H$ does not change. However, the cutoff $y_A^*$ increases when $\Omega_B$ shifts, so message $m_1$ is sent less often. Intuitively, with a higher $\mu_B$, the hawkish extremist has less reason to send $m_1$, because player $A$ is anyway very inclined to choose $H$ when player $B$ is tough. The message $m_1$ corresponds to a "barking dog" that reveals that player $A$ will choose $H$. Because this information is valuable to player $B$, the strategic effect is negative, and player $B$ will underinvest (Puppy Dog Ploy).

Suppose instead that player $B$'s publicly observed investment reduces $d_B$. Then $\Gamma_A(d_B)$ falls, so the set of types of player $A$ that choose $H$ shrinks. This strategic effect is positive for player $B$. Moreover, the investment shifts the $\Omega_B$ function to the left, so $y_A^*$ falls, say to $y_A^* - \varepsilon$. The "bark" $m_1$ that reveals player $A$'s action now sounds for types in the interval $[y^* - \varepsilon, y^*]$. This is also positive for player $B$. Thus, both effects make player $B$ better off, so he will overinvest (Fat Cat strategy).

To summarize, with strategic complements, player $B$ has an incentive to use a Puppy Dog Ploy and a Fat Cat strategy, whether or not there is a hawkish extremist. That is, player $B$ has an incentive to underinvest in offensive capability and overinvest in defensive capability to increase information transmitted by the terrorists or to reduce the incentive of player $A$ to be aggressive.

## 5.2 Strategic Substitutes

Suppose $d_i < \mu_i$ for $i \in \{A, B\}$ and player $E$ is a dovish extremist. Define $\Gamma_A(x)$ as in (13), and

$$\Omega_B(y) \equiv \frac{[1 - F_A(y)]\,\mu_B + F_A(\Gamma_A(\mu_B))d_B}{[1 - F_A(y)] + F_A(\Gamma_A(\mu_B))}.$$

Now let $x_B^* = \Omega_B(y_A^*)$ and $y_A^* = \Gamma_A(x_B^*)$. Arguing as in Section 4.2, if $F_A'(c_A) < \frac{F_A(\Gamma(\mu))}{\mu - d}$ for all $c_A$ then there exists a unique pair $(x_B^*, y_A^*)$ such that $y_A^* = \Gamma_A(x_B^*)$ and $x_B^* = \Omega_B(y_A^*)$. Moreover, $d_A < \Gamma_A(\mu_B) < y_A^*$ and $\hat{x}_B < x_B^* < \mu_B$. The strategies are the obvious generalizations of the strategies in Section 4.2. Player $E$ sends the message $m(c_A) = m_1$ if and only if $\Gamma_A(\mu_B) < c_A \leq y_A^*$. Player $A$'s cut-off points are $c_A(m_0) = y_A^*$ and $c_A(m_1) = \Gamma_A(\mu_B)$. Player $B$'s cut-off points are $c_B(m_0) = x_B^*$ and $c_B(m_1) = \mu_B$. Notice that, in equilibrium, player $A$ chooses $H$ if and only if $c_A \leq \Gamma_A(\mu_B)$.

Suppose player $B$, at time 0, makes a publicly observed investment which reduces $d_B$. This shifts the $\Omega_B$ function to the left: player $B$ becomes "soft". Since $\Gamma_A(\mu_B)$ does not depend on $d_B$, the set of types of player $A$ that choose $H$ does not change. However, the cutoff $y_A^*$ increases when $\Omega_B$ shifts, so message $m_1$ is sent more often. Intuitively, a lower $d_B$ encourages $A$ to choose $H$ (to take advantage of the not-so-tough player $B$) but to counter that, the dovish extremist organizes peace protests. Because $m_1$ is an informative signal that reveals that player $A$ will choose $D$, the fact that $m_1$ is sent more often makes player $B$ better off (it becomes easier to exploit $A$). This means that the strategic effect is positive, and player $B$ will overinvest to look soft (Fat Cat). Recall that if the extremist is not present, the Lean and Hungry Look is optimal. Thus, the presence of the dovish extremist flips the strategic effect in the opposite direction. Intuitively, player $B$ and the pacifist have a common interest: to make player $A$ back down. The pacifist becomes more inclined to "help" player $B$ when player $B$ is soft, and this produces the Fat Cat effect.

Suppose instead that player $B$'s publicly observed investment increases $\mu_B$. Then $\Gamma_A(\mu_B)$ falls, so the set of types of player $A$ that choose $H$ shrinks.

This strategic effect is positive for player $B$. However, the investment shifts the $\Omega_B$ function to the right, so $y_A^*$ falls, say to $y_A^* - \varepsilon$. Intuitively, with a higher $\mu_B$, the dovish extremist has less reason to organize peace protests, because player $A$ is anyway more inclined to choose $D$ when player $B$ has become tough. Because $m_1$ is an informative signal that alerts player $B$ that player $A$ is about to choose $D$, the fact that $m_1$ is sent less often makes player $B$ worse off. Thus, in this case there are two strategic effects which go in opposite directions. Increasing $\mu_B$ has a direct effect on player $A$, making him more likely to back down, and this benefits player $B$. But the indirect effect (fewer peace protests) hurts player $B$. In general, we cannot say if the Top Dog strategy or Puppy Dog Ploy is optimal.

To summarize, with strategic substitutes, the presence of a dovish extremist changes the strategic effects in an interesting way. Player $B$ has less of an incentive to behave aggressively (Top Dog or Lean and Hungry Look) because this would, in effect, make the pacifists in country $A$ less "cooperative" (from player $B$'s perspective). Instead, he has an incentive to overinvest in defensive technology (Fat Cat strategy). The strategic effect of an increased offensive ability cannot be signed.

# 6 Conclusion

In a model where actions are strategic complements and conflict is caused by fear of the opponent, Baliga and Sjöström [2] found that decision makers can use cheap-talk to create peace. In the current article, we have investigated the effects of cheap-talk by a third party. We found that hawkish extremists are either bad for peace (when actions are strategic complements) or irrelevant (when actions are strategic substitutes). Dovish extremists are either irrelevant (strategic complements) or have an ambiguous impact because they make one country more aggressive while the other backs down. In all cases, informative cheap-talk has a non-convex structure: it identifies a subset of moderate (intermediate) decision makers.

Assuming strategic complements, Baliga and Sjöström [2] showed that decision makers can reduce conflict by using cheap-talk. This cheap-talk identifies "tough" moderates, who would have chosen $H$ without communication, but now feel safe to coordinate on $D$. In contrast, a hawkish extremist's message identifies "weak" moderates, who would have chosen $D$ in the communication-free equilibrium, but now choose $H$ instead. In equilibrium, player $A$ always chooses $H$ following terrorism (message $m_1$), and player $B$ uses the maximal cut-off point $c_B(m_1) = d$. When there is no terrorism, both players behave less aggressively (but still more aggressively than in the communication-free equilibrium). This suggests that "the propaganda of the deed" does not require messages to be costly. According to this interpretation, Hamas would engage in terrorism in order to provoke an Israeli counter-response, and make the Palestinian population more inclined to take hawkish actions. Al Qaeda would attack the U.S., not to force the U.S. to back down, but to encourage it to aggress, forcing Islamic leaders and populations to stop cooperating with it. Because the hawkish extremist's messages create conflict, players $A$ and $B$ have a common interest in suppressing the messages. Hence, international cooperation to eliminate terrorism might be feasible.

The case of strategic substitutes follows a very different logic. The dovish extremist's message identifies "tough" moderates who would have chosen $H$ in the communication-free equilibrium, but now back down and choose $D$. By strategic substitutes, the opponent then becomes more likely to choose $H$. But most importantly for the pacifists, the chance of a coordination failure $HH$ is reduced, and this makes them better off ("better red than dead"). Player $B$ benefits from the activity of the pacifist, but player $A$ would like to suppress it. However, the negative welfare effect of the pacifist on player $A$ is mitigated when player $B$'s ex ante investment is taken into account. For the pacifists in country $A$ to "cooperate" with player $B$, player $B$ cannot appear too threatening. Hence, player $B$'s (publicly observed) investment will be skewed towards "defensive" measures, and this is good for player $A$.

# References


[1] Abu Bakr Naji (2004): *The Management of Savagery,* Olin Institute for Strategic Studies, Harvard University.

[2] Sandeep Baliga and Tomas Sjöström (2004): "Arms Races and Negotiations," *Review of Economic Studies,* Vol. 17, No. 1, pp 129-163.

[3] Sandeep Baliga and Tomas Sjöström (2008): "Strategic Ambiguity and Arms Proliferation," *Journal of Political Economy* 116: 1023-1057

[4] Sandeep Baliga and Tomas Sjöström (2008): "Bargaining Foundations of Conflict Games," mimeo, Northwestern University

[5] Efraim Benmelech and Claude Berrebi (2006): "Attack Assignments in Terror Organizations and the Productivity of Suicide Bombers," Harvard University Typescript

[6] Efraim Benmelech, Claude Berrebi and Esteban Klor (2009): "Economic Conditions and the Quality of Suicide Terrorism," mimeo, Hebrew University

[7] Eli Berman (2003): "Hamas, Taliban, and the Jewish Underground: An Economists View of Radical Religious Militias." UC San Diego typescript.

[8] Eli Berman and David D. Laitin (2008): "Religion, Terrorism and Public Goods: Testing the Club Model." NBER Working Papers 13725.

[9] Claude Berrebi (2003): "Evidence About the Link Between Education, Poverty and Terrorism Among Palestinians." Princeton University Industrial Relations Section Working Paper #477.

[10] Ethan Bueno de Mesquita (2005): "The Quality of Terror." *American Journal of Political Science* 49(3):515–530.

[11] Ethan Bueno de Mesquita (2007) "Politics and the Suboptimal Provision of Counterterror." *International Organization* 61(1):9–36.

[12] Ethan Bueno de Mesquita and Eric Dickson (2007): "The Propaganda of the Deed: Terrorism, Counterterrorism, and Mobilization." *American Journal of Political Science* 51(B):364-381.

[13] Ethan Bueno de Mesquita (2008): "The Political Economy of Terrorism: A Selective Overview of Recent Work," *The Political Economist* 10(1):1-12.

[14] Arthur Conan Doyle (1894): *The Memoirs of Sherlock Holmes*, George Newnes, London, United Kingdom.

[15] Martha Crenshaw (1981): "The Causes of Terrorism," *Comparative Politics* 13, 379-399.

[16] Cohen, M. (1987): *The Origins and Evolution of the Arab-Zionist Conflict*, University of California Press, Berkeley and Los Angeles

[17] Vincent Crawford and Joel Sobel (1982): "Strategic Information Transmission," *Econometrica,* 50: 1431-1451.

[18] Rui de Figueiredo and Barry Weingast (2001): "Vicious Cycles: Endogenous Political Extremism and Political Violence," mimeo, Berkeley and Stanford.

[19] Avinash Dixit and Barry Nalebuff (2008): *The Art of Strategy*, W.W. Norton and Company, New York.

[20] Walter Enders and Todd Sandler (1993): "The Effectiveness of Anti-Terrorism Policies: Vector-Autoregression-Intervention Analysis." *American Political Science Review* 87:829–844.

[21] Joseph Farrell and Robert Gibbons (1989): "Cheap Talk with Two Audiences," *American Economic Review*, 79, 1214-23.

[22] Joseph Farrell and Robert Gibbons (1989): "Cheap Talk Can Matter in Bargaining," *Journal of Economic Theory*, 48, 221-37.

[23] Drew Fudenberg and Jean Tirole (1984): "The Fat-Cat Effect, the Puppy-Dog Ploy, and the Lean and Hungry Look," *American Economic Review*, 74 (2) :361-366

[24] Maria Goltsman and Gregory Pavlov (2008): "How to Talk to Multiple Audiences?," mimeo, University of Western Ontario

[25] Bruce Hoffman (2006): *Inside Terrorism*, Columbia University Press: New York.

[26] Lawrence R. Iannaccone and Eli Berman (2006): "Religious Extremists: The Good, the Bad and the Deadly." *Public Choice* 128(1–2):109–129.

[27] David Jaeger and Daniele Paserman (2006): "Israel, the Palestinian Factions and the Cycle of Violence." (with David Jaeger). *American Economic Review*, 96(2), pages 45-49.

[28] David Jaeger and Daniele Paserman (2007): "The Shape of Things to Come? Assessing the Effectiveness of Suicide Bombings and Targeted Killings," mimeo

[29] David Jaeger and Daniele Paserman (2008): "The Cycle of Violence? An Empirical Analysis of Fatalities in the Palestinian-Israeli Conflict," *American Economic Review.*

[30] David Jaeger, Esteban Klor, Sami Miaari and Daniele Paserman (2008): ""The Struggle for Palestinian Hearts and Minds: Violence and Public Opinion in the Second Intifada," mimeo, Boston University.

[31] Robert Jervis (1978): "Cooperation Under the Security Dilemma," *World Politics*, Vol. 30, No. 2., pp. 167-214.

[32] Hanjoon Michael Jung (2007): "Strategic Information Transmission through the Media," Working Paper, Lahore University

[33] Alan B. Krueger and Jitka Maleckova. 2003. "Education, Poverty, and Terrorism: Is There a Causal Connection?" *Journal of Economic Perspectives* 17(4):119–144.

[34] Alan Krueger (2007): *What Makes a Terrorist: Economics and the Roots of Terrorism.* Princeton: Princeton University Press.

[35] Andrew Kydd and Barbara F. Walter (2002): "Sabotaging the Peace: The Politics of Extremist Violence," *International Organization* 56(B):263–296.

[36] Andrew Kydd and Barbara F. Walter (2006): "The Strategies of Terrorism." *International Security*, 31 (A): 49-79.

[37] Steven Matthews (1989): "Veto Threats: Rhetoric in a Bargaining Game," *Quarterly Journal of Economics*, 104.

[38] Steven Matthews and Andrew Postlewaite (1989): "Pre-play communication in two-person sealed-bid double auctions," *Journal of Economic Theory* 48, 238-263.

[39] Joseph Nye (2007): *Understanding International Conflict (6th Edition). Longman Classics in Political Science.* Longman: New York City

[40] Peter Ordershook and Thomas Palfrey (1988): "Agendas, Strategic Voting, and Signaling with Incomplete Information," *American Journal of Political Science*, v. 32, #2, pp. 441-66.

[41] Robert Pape (1996): "The Strategic Logic of Suicide Terrorism," *American Political Science Review*, 97, 343-361.

[42] Robert Powell (2007b): "Allocating Defensive Resources with Private Information about Vulnerability," *American Political Science Review* 101(4):799–809.

[43] Robert Powell (2007c) "Defending Against Terrorist Attacks with Limited Resources," *American Political Science Review* 101(3):527–541.

[44] Nigel Rees (2002): *Mark my words: Great Quotations and the stories behind them*, Barnes and Noble.

[45] Thomas Schelling (1960): *The Strategy of Conflict*. Cambridge: Harvard University Press.

[46] A. Michael Spence (1973): "Job Market Signaling," *Quarterly Journal of Economics* 87(3): 355-374.

[47] Xavier Vives (2001): *Oligopoly Pricing: Old Ideas and New Tools*, Cambridge: MIT Press.

Figure 1. Strategic Complements:
Communication-Free Equilibrium

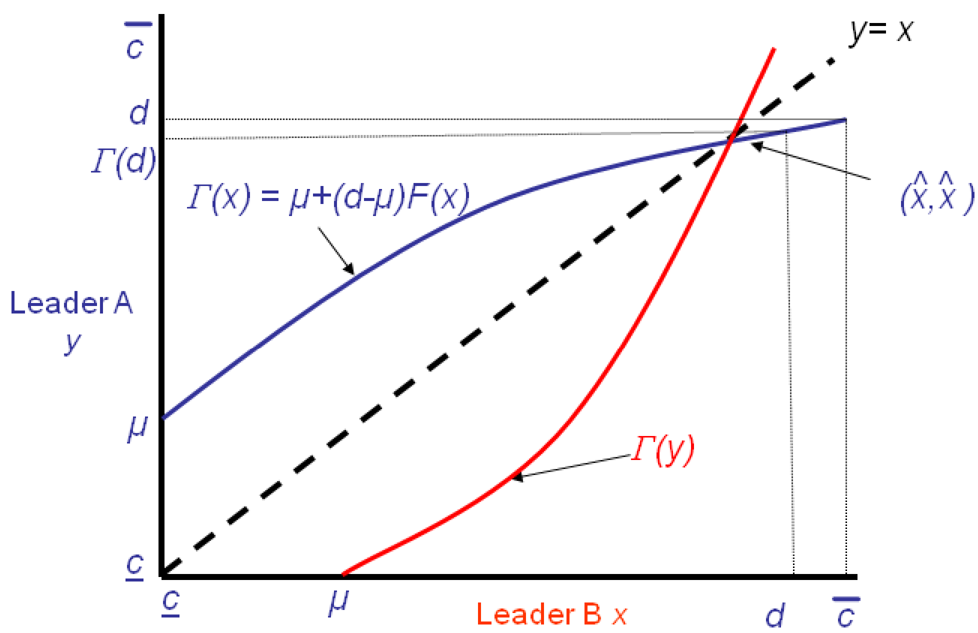

Figure 2. Strategic Complements: Lemma 2

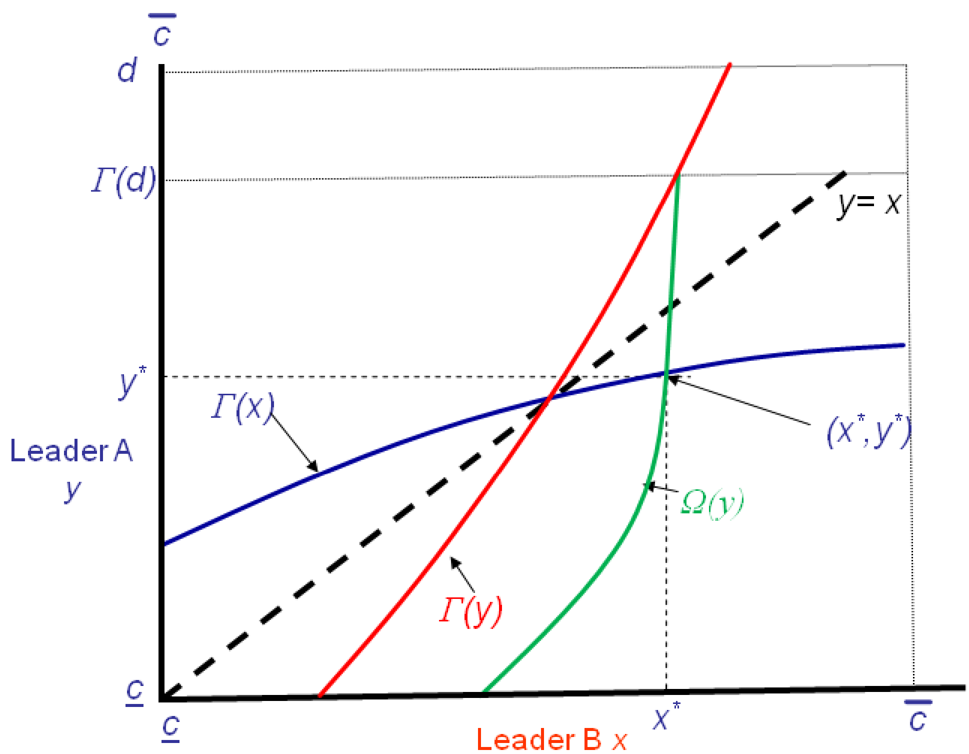

Figure 3. Strategic Substitutes: Lemma 3