Arms Races and Negotiations[*]

Sandeep Baliga

M.E.D.S., Kellogg Graduate School of Management and
Institute for Advanced Study, Princeton

Tomas Sjöström

Department of Economics, Penn State University

March 13, 2001

Abstract

A state which does not desire an arms race may nevertheless acquire new weapons if it believes another state will acquire them. If each state assigns some arbitrarily small probability to the event that the other state has a dominant strategy to acquire more weapons, then a *multiplier effect* appears, and the unique Bayesian Nash equilibrium involves an arms race with probability one. However, if the prior probability that a player is a dominant strategy type is sufficiently small, then there is an equilibrium of the *cheap-talk extension* of the arms race game where the probability of an arms race is close to zero.

"Pakistan does not intend to aggress...[W]e are the victim of (Indian) aggressions."
Foreign Minister Gohar Ayub Khan as reported by the Pakistan News Service, June
1999.

"In India, one often hears that 'Pakistan understands' that India has no hostile
designs on it..In Pakistan, however, there is strong sense that the nation's survival is
potentially at risk in the event of a major Indian attack. Without a clearer under-
standing of India's defence doctrine, this could generate a catastrophic miscalculation,"
CSIS South Asia Monitor, February 1, 1999

"Whatever happens in India, they blame Pakistan. Whatever happens in Pakistan,
we blame India...[N]either Pakistan nor India has gained anything from the conflicts
and tensions of the past 25 years." Nawaz Sharif, then Prime Minister of Pakistan,
Washington Post, Feb. 22, 1999.

1 Introduction

Two states have to decide whether or not to invest in a new weapons system. Each
state thinks the *best* possible outcome is for *neither* side to invest in new weapons, but
the *worst* possible situation is to be unarmed when the other state is armed. So no
state desires an arms race, but each state will acquire new weapons if it believes the
other state will acquire them. If these preferences are common knowledge, then two
pure strategy Nash equilibria exist: an "arms race equilibrium" in which both states
acquire new weapons, and a "detente equilibrium" in which neither state acquires
new weapons. Rational players should be able to coordinate on the Pareto dominant
detente equilibrium, perhaps using communication (O'Neill [12]). However, suppose
each state assigns some very small probability to the event that the opponent is a
truly aggressive type for whom acquiring arms is a dominant strategy. For example,
the fact that India actually has no desire to attack Pakistan may not be known for
certain in Pakistan. This gives Pakistan a reason to arm in self-defense. But it is
usually hard to distinguish between offensive and defensive weapons (Schelling [17],
Jervis [7]), so the suspicion that Pakistan may arm, for whatever purpose, makes
India more likely to arm. Anticipating this, Pakistan has an even stronger reason to
arm, and so on. So as Schelling [17] pointed out, a *multiplier effect* appears, creating
an escalating cycle of pessimistic expectations toward mutual armaments.

We formalize the situation as follows. Each state has a *type* which parameterizes
the propensity to arm. The state's true type is its private information, and types are
independently drawn from a continuous distribution. For all types, the worst possible
outcome is to be unarmed while the opponent arms. However, the type determines
whether or not the state prefers to arm when it is unsure about the actions of the
opponent. At one end of the distribution are the aggressive "dominant strategy"
types: they prefer to arm regardless of the opponent's actions. At the other end of
the distribution are the peaceful types who prefer to arm only if they are virtually
sure that the opponent will arm. Let the fraction of dominant strategy types be some

small $\varepsilon > 0$. These types will certainly arm, but this triggers a multiplier effect. Some fraction $\delta > 0$ of all types are not dominant strategy types but prefer to arm when the opponent arms with at least probability ε. These "almost dominant strategy types" must arm in equilibrium. But then, all types that prefer to arm when the opponent arms with at least probability $\varepsilon + \delta$ must arm, etc. The spiral of ever more pessimistic expectations causes more and more types to arm. Even though each state thinks it is *extremely unlikely* that the opponent is the dominant strategy type, the *unique* Bayesian-Nash equilibrium may involve an arms race with probability one.

What can be done to escape this logic ? World leaders, aware that escalating fear may cause arms races and conflict, seem to believe in the importance of communication. Ronald Reagan reported in his diary the message he wanted to communicate to Soviet foreign minister Gromyko: "I have a feeling we'll get nowhere with arms reductions while they are as suspicious of our motives as we are of theirs. I believe we need a meeting to see if we can't make them understand we have no designs on them but think they have designs on us" (Reagan [15]). Arms control talks eventually did lead to a 50% reduction in strategic forces and the elimination of intermediate and medium-range ballistic missiles in Europe. However, in the beginning of his administration Reagan had used aggressive rhetoric, declaring the Soviet Union to be an "evil empire" prepared "to commit any crime, to lie, to steal" to achieve its goals (Reagan [14]). The precise role of communication in signalling a "type" may be subtle. This motivates our study.

We consider a cheap-talk extension of the arms race game. Before making the decision to arm, each state sends a hawkish (aggressive) or a dovish (conciliatory) message to the opponent. The messages are pure cheap-talk: a state is free to arm itself regardless of what messages were sent. Since all types are better off if the opponent does not arm (whatever they themselves decide to do) one might suspect that all types would send whatever message is most likely to persuade the opponent not to arm. If this were the case, then cheap talk would be uninformative, and we would be back in the original arms race spiral. However, cheap talk *can* be informative. The reason is that for most types it is not *only* the probability that the opponent arms that matters - it is also important to be able to coordinate with the opponent. The main result of this paper is that if the dominant strategy types are sufficiently rare, then there exists an equilibrium of the cheap-talk extension where the probability of an arms race is close to zero. Thus, communication can *expand the set of equilibria*. Indeed, it can have the dramatic effect of reducing the probability of an arms race from one to almost zero![1]

The equilibrium works as follows. Most types send the dovish message. The exception is a small number of *fairly tough* types who have a high propensity to arm, but not so high as to be dominant strategy types. These fairly tough types send

---

[1]In any cheap talk game, there are "babbling" equilibria where the parties simply disregard the messages. However, throughout the paper we assume that if there are multiple Pareto ranked equilibria, the players manage to coordinate on an efficient one.

the hawkish message. If the two states sent different messages, the result is an arms race: both states acquire new weapons. If both states sent hawkish messages then neither state acquires any weapons. (The hawkish messages reveal that both states are fairly tough, but not tough enough to be dominant strategy types). Finally, if both states sent dovish messages, then the *very tough* types that have the highest propensity to arm (including the dominant strategy types) acquire new weapons, while the remaining types (who may be called *normal*) do not. If the number of dominant strategy types is very small, then the number of normal types is close to one, so the probability is close to one that both states are dovish and refrain from acquiring weapons.

With a very small probability, a normal type will face a very tough type who appears dovish at first but then arms unilaterally ("a wolf in sheep's clothing"). In that case, the normal type's realized payoff will be the lowest possible. Still, *ex ante* the normal types are better off trusting an opponent who appears dovish, as long as the probability that the opponent is very tough is sufficiently low. The fairly tough types, however, are not willing to take this gamble. Their propensity to arm is high enough that they prefer to appear hawkish. Of course, since messages are *non-binding*, the fairly tough types could send a dovish message and then go ahead and arm anyway. However, with such a strategy they would *always* end up arming. This would be worse for them than what they get in equilibrium, since in equilibrium they avoid the arms race whenever they meet another fairly tough type. Being able to coordinate in this way is valuable to the fairly tough types (recall that they are not dominant strategy types). This separation of very tough types (who send the dovish message) from fairly tough types (who send the hawkish message) prevents the multiplier effect. Notice that the very tough types prefer to appear dovish because they are not interested in coordinating with the opponent. They know that they themselves will surely arm, all they want to do is to encourage the opponent not to arm, and this is done most efficiently by appearing dovish. Indeed, *any* cheap-talk equilibrium which does not involve an arms race with probability one must have the property that dominant strategy types sometimes arm unilaterally against peaceful opponents. For the only way to prevent such unilateral arms build-up would be to get the dominant strategy type to reveal his true nature, alerting the opponent that he should arm too. But this would not be incentive compatible since the dominant strategy type does not want his opponent to arm.

The paper is organized as follows. Section 2 presents the basic model without communication. A *multiplier condition* on the distribution of types is shown to be necessary and sufficient for an arms race to occur with probability one even if the dominant strategy types are very rare. Section 3 shows how cheap talk reduces the probability of an arms race to almost zero when the dominant strategy types are very rare. Section 4 discusses related literature and Section 5 concludes. Technical calculations are contained in the appendix.

2 The Arms Race Game

Two players must simultaneously and independently decide whether or not to invest in a new weapons program. The possible choices are *Build new weapons* (B) or *No new weapons* (N). We normalize the payoff to be zero for each player if both choose N. A player who chooses N while the other player chooses B suffers a loss of $d > 0$ units of utility. This loss represents the insecurity that a player suffers when the other player has a more advanced weapons system than he has. (Psychological factors such as "loss of prestige" could influence d). A player who builds the new weapons system does not have to suffer this insecurity, as he will always be at least as strong as his opponent. On the other hand, he has to pay the cost of the new weapons. Let player i's cost of acquiring new weapons be denoted $c_i \geq 0$. A player who builds the new weapons system while his opponent does not obtains a gain of $\mu > 0$ units of utility. (Perhaps he can extract some resources from his weaker opponent, or gains "prestige" etc.). We shall be mainly interested in the case where μ is small, so that the temptation to build new weapons is not too big. Player i's payoffs can be represented in a payoff matrix as follows (player i chooses a row, player j a column):

$$
\begin{array}{ccc}
 & \text{B} & \text{N} \\
\text{B} & -c_i & \mu - c_i \\
\text{N} & -d & 0
\end{array}
\tag{1}
$$

If $d > c_i > \mu$ for each $i \in \{1, 2\}$, and if all payoffs are common knowledge, then there are two pure strategy Nash equilibria: (B,B) and (N,N). In this case, rational players should be able to coordinate on the Pareto dominant equilibrium (N,N). However, we will assume c_i is player i's *private information*.[2] We refer to c_i as player i's *type*. Each player i knows his own type c_i, but not the other player's type c_j. Everything except the true c_1 and c_2 is common knowledge. Each c_i is independently drawn from the same distribution, with cumulative distribution function denoted F. F has support $[0, \bar{c}]$ with $F(0) = 0$, $F'(c) > 0$ whenever $0 < c < \bar{c}$, and $F(\bar{c}) = 1$. Assume $\bar{c} < d$. We will make the mild regularity assumption that F has a power series representation. That is, there are coefficients $a_0, a_1, a_2...$ such that for all $c \in [0, \bar{c}]$,

$$
F(c) = \sum_{j=0}^{\infty} a_j c^j
$$

---

[2] The crucial point is that some privately observed parameter influences a player's propensity to arm. For convenience we consider the *cost of acquiring new weapons*. But the private information could equally well relate to, for example, *the benefit from being armed when the opponent is unarmed*. Nevertheless, in many real world cases incomplete information about cost may be paramount. For example, a non-nuclear state which has access to fissile material (plutonium and enriched uranium) and technical expertise from the former Soviet Union may find it relatively cheap to develop nuclear weapons - it would be very expensive otherwise. Whether or not a state has access to such material and expertise may be hard to verify for an opponent.

where the series converges uniformly on $[0, \bar{c}]$.

Since $-c_i \geq -\bar{c} > -d$, B is always a (strict) best response against B. Therefore there is a Bayesian Nash equilibrium where all types choose B with probability one. Is there any other Bayesian Nash equilibrium? Notice that N is a best response against N for player i if and only if $c_i \geq \mu$. If $c_i < \mu$ then player i is a *dominant strategy type*: B is a strictly dominant strategy for him. The probability that player i is a dominant strategy type is $F(\mu)$, which is close to zero if μ is small. The existence of dominant strategy types may have a large effect on the set of equilibria even as $\mu \to 0$. Indeed, if the following condition is satisfied then even for arbitrarily small $\mu > 0$ the unique Bayesian Nash equilibrium is for all types to play B.

Definition 1 *The distribution satisfies the* multiplier condition *if $F(c)d \geq c$ for all $c \in [0, \bar{c}]$.*

For example, the uniform distribution $F(c) = c/\bar{c}$ satisfies the multiplier condition because $d > \bar{c}$ implies $cd/\bar{c} \geq c$ for all $c \geq 0$.[3] More generally, the multiplier condition is satisfied if F is concave (because $F(c) \geq c/\bar{c} \geq c/d$ for all $c \geq 0$ when F is concave).[4]

Any equilibrium will have a cut-off property: if type c_i builds new weapons, then any type $c_i' < c_i$ will also build. Knowing that the dominant strategy types will play B, a type which is "almost" a dominant strategy type ($\mu - c_i$ negative but close to zero) will also play B. This "infects" other types with slightly higher c_i, who also decide to play B, and so on. If this contagion stops before all types have been infected, then there must be some "cut-off type" $c_i^* > 0$ such that all types with a lower cost than him play B and all types with a higher cost play N. Type c_i^* himself must be indifferent. Suppose for the moment that the equilibrium is symmetric, so $c_1^* = c_2^* = c^*$. The condition that player i's type c^* is indifferent between B and N when player j is expected to choose B with probability $F(c^*)$ is $S(c^*) = 0$, where

$$S(c) \equiv F(c)(d - c) + (1 - F(c))(\mu - c) \tag{2}$$

However, if the multiplier condition is satisfied then $S(c) > 0$ for any $\mu > 0$ and any $c \geq 0$, so there can be no symmetric Bayesian Nash equilibrium where N is chosen with positive probability. The proof of Theorem 1 shows that there are no asymmetric equilibria where N is chosen with positive probability either. Thus, if the multiplier condition holds, then the fact that the dominant strategy types will definitely choose B triggers an arms race with probability one. Conversely, if the multiplier condition is violated then for sufficiently small μ there exists $c^* < \bar{c}$ such that $S(c^*) = 0$. That

---

[3]For any F, $\bar{c}/F(\bar{c}) = \bar{c} < d$, so a *sufficient* condition for the multiplier condition to hold is for $c/F(c)$ to be an increasing function. This is true in the uniform case.

[4]If F is sufficiently convex then the multiplier condition will be violated. In this case, low cost types are relatively rare compared to high cost types, and the contagion to play B will not necessarily infect the whole population.

means type c^* is just indifferent between building and not building if he thinks his opponent plays B with probability $F(c^*)$. Therefore, there exists a Bayesian Nash equilibrium where N is chosen with positive probability.

Theorem 1 *(i) If the multiplier condition is satisfied, then for any $\mu > 0$ there is a unique Bayesian Nash equilibrium. In this equilibrium all players choose B, regardless of type. (ii) If the multiplier condition is violated, then for sufficiently small $\mu > 0$ there exists a Bayesian Nash equilibrium where N is chosen with strictly positive probability.*

Proof. If $\mu \geq d$ then B is a dominant strategy for all types, so the analysis is trivial. Suppose instead that $0 < \mu < d$. First, we establish the cut-off property: if B is a weak best response for type c_i then it is a strict best response for type $c_i' < c_i$. Indeed, if player i thinks player j will choose B with probability p_j, the payoff to player i from B is

$$p_j(-c_i) + (1 - p_j)(\mu - c_i) = (1 - p_j)\mu - c_i$$

while the payoff from N is $p_j(-d) + (1 - p_j) \times 0$. Type c_i weakly prefers B if and only if

$$c_i \leq (1 - p_j)\mu + p_j d \tag{3}$$

Notice that all of player i's types have the same beliefs about player j, since types are assumed to be uncorrelated. If type c_i weakly prefers B, then inequality (3) is strict for type $c_i' < c_i$ so type c_i' strictly prefers B. Now we can prove the two parts of the theorem.

(i) If player j chooses B with probability one, then all of player i's types will choose B since $\bar{c} < d$. Therefore, there is always an equilibrium where all players choose B, regardless of type. Suppose in addition there is an equilibrium where N is played with positive probability. We claim the multiplier condition is violated. If a player chooses N then he must expect the opponent to choose N with strictly positive probability, hence if N is chosen with positive probability by one player then *both* players must choose N with positive probability. For $i \in \{1, 2\}$ let c_i^* be such that B is a weak best response for player i at the equilibrium if and only if his type satisfies $c_i \leq c_i^*$. By hypothesis, $c_i^* < \bar{c}$, for otherwise player i chooses N with probability zero. The probability that player $i \in \{1, 2\}$ chooses B is $p_i = F(c_i^*)$. Since type c_i^* must be indifferent between B and N,

$$c_i^* = (1 - p_j)\mu + p_j d = (1 - F(c_j^*))\mu + F(c_j^*)d$$

Without loss of generality, suppose $c_1^* \leq c_2^*$. Then

$$c_1^* = (1 - F(c_2^*))\mu + F(c_2^*)d \geq (1 - F(c_1^*))\mu + F(c_1^*)d > F(c_1^*)d$$

7

since $0 < \mu < d$ and $F(c_1^*) \le F(c_2^*) < 1$. Thus, the multiplier condition is violated.

(ii) Suppose the multiplier condition is violated. Then, there exists c' such that $c' > dF(c')$. For sufficiently small $\mu > 0$, we have $S(c') \le 0$ where S is defined by (2). Also, $S(\bar{c}) = d - \bar{c} > 0$. By continuity, there is $c^* < \bar{c}$ such that $S(c^*) = 0$. Let each player i choose B if and only if $c_i \le c^*$. Since $S(c^*) = 0$, by the construction of S it follows that type c^* is indifferent between B and N. Type $c_i < c^*$ strictly prefers B and type $c_i > c_i^*$ strictly prefers N. Thus, these strategies form a Bayesian Nash equilibrium. ∎

Remark 1. *The proof of Theorem 1 shows that when the multiplier condition holds, the game is interim dominance solvable. After iterated elimination of strongly (interim) dominated strategies only the "arms race" outcome remains.*

Remark 2. *The arm race outcome is inefficient because all types prefer NN to BB.*

Remark 3. *The arms race is caused by mutual distrust. Suppose, contrary to our assumptions, that c_2 (but not c_1) becomes common knowledge as soon as the types are determined by nature. Then there would exist a Bayesian Nash equilibrium where player 2 as well as all the non-dominant strategy types of player 1 choose N whenever $c_2 \ge (1 - F(\mu))\mu + F(\mu)d$, which happens with probability close to 1 for μ small enough. If both c_1 and c_2 become common knowledge as soon as they are determined by nature, then there is an equilibrium where both players choose N whenever $c_1 \ge \mu$ and $c_2 \ge \mu$, which again happens with probability close to 1 for μ small enough.*

3   Cheap Talk

In this section we will assume the multiplier condition holds. Without cheap talk, there is a discontinuity in the equilibrium correspondence: if μ were zero then there would be an equilibrium where all types choose N, but for any $\mu > 0$, a "contagion" is started by the fraction $F(\mu) > 0$ of dominant strategy types. The multiplier condition guarantees that each player who is not a dominant strategy type will choose B whenever he thinks all types that have lower cost than him will choose B. Thus, the game unravels and everybody plays B, as shown in Theorem 1. We now show that adding cheap talk restores continuity in the sense that for small enough $\mu > 0$ there exists an equilibrium where almost all types choose N.

Since we are assuming $\mu > 0$ and $d > 0$, no matter what player i plans to do he always strictly prefers player j to choose N. This makes credible communication difficult but not impossible to achieve. We will divide the types into three groups: "very tough", "fairly tough", and "normal." The very tough types have the lowest cost and will always play B. The fairly tough types have a slightly higher cost, and are willing to play N as long as they are assured that they are not facing a very tough type (who would play B against them). Thus, very tough types and fairly tough types must send different messages, making it possible for fairly tough types to coordinate

on N with other fairly tough types while playing B against very tough types. This is the way we break the contagion. Of course, the very tough types must be given some incentive to separate themselves from the fairly tough types in this way. This is done by letting them pool with the normal types, allowing the very tough types to arm unilaterally against the normal types. The normal type is willing to be taken advantage of in this way as long as it happens sufficiently infrequently. Most of the time, normal types will meet other normal types and coexist peacefully. We now formalize these arguments.

In the cheap talk extension of the arms race game there are three stages. In stage zero, nature determines c_1 and c_2, and c_i becomes player i's private information. In stage one, messages are announced simultaneously and publicly. The two messages that will be sent in equilibrium will be labelled Dove and Hawk. "Dove" is interpreted as a conciliatory message and "Hawk" is interpreted as an aggressive message. (Other messages can be allowed, but will not be sent in equilibrium.) In stage two, the players simultaneously choose either B or N, and player i's payoff is determined by his payoff matrix (1). The messages sent in stage one do not influence the payoffs directly, but they may convey information about what the players plan to do in the future. Our main theorem states that arms races can be avoided almost surely if the fraction of dominant strategy types $F(\mu)$ is sufficiently small, that is, if $\mu > 0$ is sufficiently small.

Theorem 2 *Suppose the multiplier condition is satisfied. For any $\delta > 0$ there is $\bar{\mu} > 0$ such that if $0 < \mu < \bar{\mu}$ then there is a perfect Bayesian equilibrium of the cheap-talk extension of the arms race game where N is played with at least probability $1 - \delta$.*

The remainder of this section consists of the proof of this theorem. We need the following lemma, which is proved in the Appendix.

Lemma 1 *For sufficiently small $\mu > 0$, there exists a triple (c^L, c^*, c^H) such that*

$$\mu < c^L < c^* < c^H < \bar{c} \tag{4}$$

$$\left[F(c^H) - F(c^L)\right] c^L = \left(1 - F(c^H)\right) \mu \tag{5}$$

$$\left[1 - 2\left(F(c^H) - F(c^L)\right)\right] c^H = F(c^L)d \tag{6}$$

$$\left(1 - F(c^H)\right)(\mu - c^*) + F(c^L)(-c^*) = F\left(c^L\right)(-d) \tag{7}$$

If $\mu \to 0$ then $c^H \to 0$.

Now let us (c^L, c^*, c^H) be as defined by Lemma 1 and consider the following strategies in the cheap talk extension of the arms race game. Player i is *normal* if $c_i > c^H$, *fairly tough* if $c^L \leq c_i \leq c^H$, and *very tough* if $c_i < c^L$. In stage 1, the cheap-talk stage, player i says Hawk if he is fairly tough. Otherwise, he says Dove. In stage 2, the arms race stage, player i behaves as follows. If $c_i \leq \mu$ then he chooses B no matter what announcements were made in stage one. If $c_i > \mu$ then player i plays as follows: (i) if both players said Hawk then player i chooses N; (ii) if one player said Dove and the other said Hawk then player i chooses B; (iii) if both players said Dove then player i chooses N if and only if $c_i \geq c^*$; (iv) If any message except Dove or Hawk was sent by any player then player i chooses B.

We describe the *equilibrium* announcements made in stage 1, and the actions played *on the equilibrium path* in stage 2, in the table below. For example, if player 1 is very tough $(c_1 < c^L)$ and player 2 is fairly tough $(c^L \leq c_2 \leq c^H)$, then player 1 says Dove and player 2 says Hawk. Both players proceed to choose B, i.e., there is an arms race.

	$c_2 < c^L$ (Dove)	$c^L \leq c_2 \leq c^H$ (Hawk)	$c_2 > c^H$ (Dove)
$c_1 < c^L$ (Dove)	BB	BB	BN
$c^L \leq c_1 \leq c^H$ (Hawk)	BB	NN	BB
$c_1 > c^H$ (Dove)	NB	BB	NN

We claim that for small enough $\mu > 0$, these strategies form a perfect Bayesian equilibrium[5] of the cheap talk extension of the arms race game. Notice for future reference that (5) and the fact that $\mu < c^L$ implies that there are more normal types than fairly tough types:

$$1 - F(c^H) > F(c^H) - F(c^L) \tag{8}$$

In fact the right hand side of (8) will be close to zero and the left hand side close to one for μ small (for then both c^L and c^H will be close to zero).

Lemma 2 *The strategies specified above are sequentially rational in the action stage for all types, following all messages.*

Proof. If $c_i \leq \mu$, then it is clearly in player i's interest to choose B no matter what happened in stage 1. Suppose $c_i > \mu$. If both players announced Hawk in stage

---

[5] See Fudenberg and Tirole [6] for a formal definition.

10

1, then the opponent is expected to choose N in stage 2, and N is a best response against N. If one player said Dove and the other Hawk, or someone said something else than "Hawk" or "Dove", then the opponent is expected to choose B in stage 2, and B is a best response against B. Finally, suppose both players said Dove in stage 1. In this case, player i thinks his opponent is either a normal type who will choose N, or a very tough type who will choose B (recall that fairly tough types are the only ones who say Hawk in equilibrium). Now there are $1 - F(c^H)$ normal types and $F(c^L)$ very tough types. Then (7) implies that player i is indifferent between B and N if he is of type $c_i = c^*$. Clearly it is a best response for player i to choose B if $c_i < c^*$ and N if $c_i \geq c^*$. ∎

We now turn to the cheap talk stage. Notice that for any type the expected utility from following the strategies specified above is greater than what he gets from playing BB for sure. Hence, no type has an incentive to send any message other than Hawk or Dove.

Lemma 3 *Player i prefers to say Dove if $c_i \leq \mu$.*

Proof. The dominant strategy type will go on to choose B for sure, so his objective is simply to maximize the probability of his opponent choosing N. If player i says Hawk, his opponent will choose N if and only if the opponent is a fairly tough type (who says Hawk according to his equilibrium strategy), an event which occurs with probability $F(c^H) - F(c^L)$. If player i says Dove, his opponent will choose N if and only if the opponent is a normal type (who says Dove according to his equilibrium strategy), an event which occurs with probability $1 - F(c^H)$. By (8), the dominant strategy type prefers to say Dove. ∎

Lemma 4 *Player i prefers to say Dove if $\mu < c_i < c^L$ and Hawk if $c^L \leq c_i < c^*$.*

Proof. Suppose $\mu < c_i < c^*$. First, suppose player i says Hawk. Then if the other player also says Hawk both will choose N. This event occurs with probability $F(c^H) - F(c^L)$. Otherwise, both choose B. The expected payoff to player i is

$$\left[1 - (F(c^H) - F(c^L))\right](-c_i) \tag{9}$$

Suppose instead player i says Dove. Since $c_i < c^*$ player i will then choose B at the action stage whatever his opponent (player j) has said. Player j will choose N if and only if he is a normal type (who says Dove according to his equilibrium strategy), an event which occurs with probability $1 - F(c^H)$. Therefore, player i's expected payoff from saying Dove is

$$\left(1 - F(c^H)\right)(\mu - c_i) + F(c^H)(-c_i) \tag{10}$$

But, equation (5) implies that (9) equals (10) if $c_i = c^L$ so player i of type c^L is indifferent between Dove and Hawk. If $c_i > c^L$, (10) is smaller than (9) so player i prefers to say Hawk. If $\mu < c_i < c^L$, (10) is bigger than (9) so player i prefers to say Dove in this case. ∎

Lemma 5 *Player i prefers to say Hawk if $c^* \leq c_i \leq c^H$ and Dove if $c_i > c^H$.*

Proof. Suppose $c_i \geq c^*$. First, suppose player i says Hawk. If the opponent is a fairly tough type (who says Hawk according to his equilibrium strategy) then both will choose N in the action stage. This event occurs with probability $F(c^H) - F(c^L)$. Otherwise, both will choose B. Therefore, player i's expected payoff from saying Hawk is

$$\left[1 - \left(F(c^H) - F(c^L)\right)\right](-c_i) \tag{11}$$

Suppose instead player i says Dove. If the opponent is a fairly tough type (who says Hawk according to his equilibrium strategy) then both will choose B at the action stage. Otherwise, the opponent will say Dove, player i will choose N at the action stage, and the opponent will choose B if he is very tough and N if he is normal. Therefore, player i's expected payoff from saying Dove is

$$\left[F(c^H) - F(c^L)\right](-c_i) + F(c^L)(-d). \tag{12}$$

But, equation (6) implies that (11) equals (12) if $c_i = c^H$ so player i of type c^H is indifferent between saying Dove and Hawk. By (8), if $c_i > c^H$ then (11) is smaller than (12) so player i prefers to say Dove. If $c^* \leq c_i < c^H$, (11) is bigger than (12) so player i prefers to say Hawk. ∎

These results show that the strategy profile specified is a perfect Bayesian equilibrium of the cheap talk extension of the arms race game. Since $c^L \to 0$ and $c^H \to 0$ as $\mu \to 0$, it is evident from the construction of the strategies that the fraction of types that play B in equilibrium goes to zero as μ goes to zero.

4 Related Literature

The idea that fear and mutual distrust, sparked by uncertainty about the opponent's motives, may be the cause of conflict has a rich history. Thucydides [19] (Book I, 23) argued that "the growth of Athenian power and the fear which this caused in Sparta" made the Peloponnesian War inevitable.[6] Rousseau (quoted in Jervis [7]) argued that "It is quite true that it would be much better for all men to remain always at peace. But so long as there is no security for this, everyone, having no guarantee that he can avoid war, is anxious to begin it at the moment which suits his own interest

---

[6]A famous passage describes how the Spartans are spurred on by the Corinthians: "You Spartans are the only people in Hellas who wait calmly on events, relying on your defense not on action but on making people think you will act. You alone do nothing in the early stages to prevent an enemy's expansion; you wait till the enemy has doubled his strength. Certainly you used to have the reputation of being safe and sure enough; now one wonders if this reputation was deserved.......The Athenians...live close to you, yet you still do not appear to notice them; instead of going out to meet them, you prefer to stand still and wait till you are attacked, thus hazarding everything by fighting with opponents who have grown far stronger than they were originally" (Book I, 69, [19]).

and so forestall a neighbor, who would not fail to forestall the attack in turn at any moment favorable to himself, so that many wars, even offensive wars, are rather in the nature of unjust precautions for the protection of the assailant's own possessions than a device for seizing those of others." Pigou (1941, Chapter 2) argued that "in view of the enormous expense of modern military and naval operations, and of the chance that a war begun on a small scale may draw in other Powers, it is extremely *improbable* that there will be to any country in the end any balance of economic gain...The fear of war itself forces governments to adopt policies that make war more likely...Nursed by the fear of war, they themselves make war more likely and are the cause of further fears." It is often argued that the arms race spiral that preceded World War I was caused by Britain's and Germany's mutual distrust of each other, rather than any nation's desire to fight a war (Wainstein [20] and Sontag [18]). Many authors have argued that similar mechanisms operated during the cold war (see the quote by Ronald Reagan in the Introduction).

Jervis [7] describes the *security dilemma* as a situation where each state thinks the best possible outcome is for nobody to invest in new weapons, but the fear that the opponent may arm causes all states to buy weapons. When this initial arms build-up becomes public information, it triggers subsequent rounds of arms build-up, which Jervis calls the *spiralling model.* Jervis did not provide any formal model and in fact argued that *irrationality* is a crucial component of arms races and conflict: "if the spiral theory is correct, it is so partly because the actors do not understand it or follow its prescriptions" (Jervis [7], page 81). Schelling [17] provides a formal model where *fully rational* players may attack each other inadvertently because of a "false alarm." Knowing that the other may inadvertently attack, each will be more likely to mount an inadvertent attack, hence there is a multiplier effect, just as in our model. However, if the underlying problem is an imperfect warning system then there is no private information that can be transmitted via cheap talk. Our results show that cheap talk can be highly useful when the underlying problem is incomplete information about preferences. Like Schelling, we assume all decision makers are perfectly rational.

Game theoretic articles related to the Schelling multiplier effect include Rubinstein [16], Carlsson and van Damme [3] and Morris and Shin [10] (see Morris and Shin [11] for a survey). However, these authors assume that players' types are correlated in a way which would be unnatural in our context. They do not consider the role of cheap talk. We obtain a multiplier effect without any correlation of types, and our focus is on the cheap talk extension. Unlike most articles in the cheap-talk literature, we assume both players have private information, send messages, and take actions. The exceptions include a recent article by Baliga and Morris [1] which contains an example of cheap talk with two-sided incomplete information, an article on double auctions by Matthews and Postlewaite [9], and an article on the battle-of-the-sexes game with two sided incomplete information by Banks and Calvert [2].

In the battle-of-the-sexes game discussed by Banks and Calvert [2], each player

has his own favorite outcome, but the intensity of his preference is determined by his type ("high" or "low"). There is no dominant strategy type and no multiplier effect. Without communication there are efficient asymmetric Bayesian Nash equilibria where one player always gets his favorite outcome. However, the only *symmetric* equilibrium involves a randomization which is inefficient. Banks and Calvert show how communication can improve by achieving a symmetric equilibrium where coordination occurs more frequently, and player i's favorite action is more likely to be chosen when the intensity of his preference is high. In the battle-of-the sexes game, player i always wants player j to take the same action as player i, while in our game player i always wants player j to disarm regardless of what player i does, so the nature of communication is somewhat different. The paper by Banks and Calvert [2] is particularly interesting because they show that a *mediator* may be necessary for efficiency. We can approximate first-best efficiency without a mediator when the dominant strategy types are rare. The case where the dominant strategy types are not rare is left for future work.

5   Conclusion

This paper makes two contributions. First, we provide a formalization of Schelling's [17] multiplier effect using incomplete information about preferences rather than an imperfect warning system. Our model is very close to Schelling's *informal* argument: "If I go downstairs to investigate a noise at night, with a gun in my hand, and find myself face to face with a burglar who has a gun in his hand, there is a danger of an outcome that neither of us desires. Even if he prefers to leave quietly, and I wish him to, there is a danger that he may *think* I want to shoot, and shoot first. Worse, there is danger that he may think that *I* think *he* wants to shoot. Or he may think that *I* think *he* thinks *I* want to shoot. And so on." (Schelling [17], page 207). Our second contribution is to show how cheap talk can resolve Schelling's dilemma. The equilibrium has several interesting properties. First, there is a positive probability that a player sends a hawkish message. If both players do this, there is no arms race, but the combination of a hawkish and a dovish message triggers an arms race. Second, there is a positive probability that a player sends a dovish message but then embarks on a unilateral arms build up. In such a case, the opponent will be at a disadvantage, but from an ex ante perspective the gamble was worth taking. In fact, the introduction of cheap talk raises *all* types' *expected payoffs*, since without cheap talk the unique equilibrium involves an arms race with probability one.

In the real world, non-binding agreements seem to be important. For example, the signing of a test ban treaty by the United States and the Soviet Union in 1963 appears to have been a brake on the arms race, even though the treaty was essentially non-enforceable in the absence of a "world government". But occasionally, a ruthless leader will break the promises he has made. In 1935 the British and the Germans exchanged dovish messages, resulting in a naval accord which limited the German

14

fleet to 35 % of the British. Hitler, of course, did not intend to respect any treaty, but the British did not know that.[7] The British realized that the accord was "cheap talk" and they worried about the fact that Hitler's true military strength was difficult to assess, but they felt that trusting Hitler was a chance worth taking (Kissinger [8], pp. 295-296). In 1940, Goebbels described Hitler's success at taking advantage of the credulity of the western powers: "up to now we have succeeded in leaving the enemy in the dark concerning Germany's real goals... They left us alone and let us slip through the risky zone, and we were able to sail around all dangerous reefs. *And when we were done, and well armed, better than they, then they started the war!*" (Kissinger [8], p. 295). In contrast, the messages sent by the British and the Germans in 1912 may be identified as dovish and hawkish, respectively. The major new weapons technology was the Dreadnought warship. With an arms race looming on the horizon the British felt "it might be possible by friendly, sincere and intimate conversation to avert this perilous development" and that "surely something could be done to break the chain of blind causation" (Churchill [4], p. 75). Churchill proposed a "naval holiday" for 1913, where each nation would promise not to build any new Dreadnoughts, but the messages coming from the Kaiser were not at all conciliatory. An arms race followed (Churchill [4], pp. 79-81). Perhaps some of the logic of a cheap-talk equilibrium can be seen in these examples.

6 Appendix

Before giving the proof of Lemma 1 we need a technical result.

Lemma 6 *If the multiplier condition holds, then there exists $\gamma > 0$ such that $F'(c)d > 1$ for all $c \in (0, \gamma)$.*

  Proof. Since F has a power series representation by assumption, the function $\lambda(c) \equiv F(c)d - c$ also has a power series representation

$$\lambda(c) = \sum_{j=0}^{\infty} k_j c^j \tag{13}$$

The multiplier condition says that $\lambda(c) \geq 0$ for all $c \geq 0$. Moreover, $k_0 = 0$ since $\lambda(0) = 0$. Also, as $\lambda(\bar{c}) = d - \bar{c} > 0$ by assumption, there is j such that $k_j \neq 0$. Let $n \geq 1$ be the *smallest* integer such that $k_n \neq 0$. For small enough $c > 0$ the expression in (13) will be dominated by the term $k_n c^n$. Hence, we must have $k_n > 0$ for $\lambda(c) \geq 0$ to be true for c close to zero. The derivative of $\lambda(c)$ is

$$\lambda'(c) = \sum_{j=1}^{\infty} j k_j c^{j-1}$$

---

[7]Hitler's abrogation of the naval accord is described by Craig [5] pages 686 and 710. Other dovish messages sent by Hitler included the signing of the German-Polish non-aggression pact.

which for small enough $c > 0$ is dominated by the term $nk_n c^{n-1} > 0$. Hence, $\lambda'(c) = F'(c)d - 1 > 0$ for $c > 0$ close enough to zero. ∎

Proof of Lemma 1.

Define two functions H and G as follows:

$$H(x,y) \equiv [F(y) - F(x)]\, x - (1 - F(y))\, \mu \tag{14}$$

$$G(x,y) \equiv [1 - 2(F(y) - F(x))]\, y - F(x)d \tag{15}$$

Then, equations (5) and (6) are equivalent to the statement that $(x,y) = (c^L, c^H)$ solves the equation system

$$\begin{aligned} H(x,y) &= 0 \\ G(x,y) &= 0 \end{aligned} \tag{16}$$

To analyze this system consider the shape of the two curves defined by (16), restricting our attention to x and y in $[0, \bar{c}]$. We have

$$H(0,y) = -(1 - F(y))\, \mu$$

Therefore, there is a unique $y \in [0, \bar{c}]$, namely, $y = \bar{c}$, such that $H(0,y) = 0$. Similarly,

$$H(\bar{c}, y) = -(1 - F(y))(\mu + \bar{c})$$

so that there is a unique $y \in [0, \bar{c}]$, namely, $y = \bar{c}$, such that $H(\bar{c}, y) = 0$. For $0 < x < \bar{c}$ we notice that

$$H(x,0) = -F(x)x - \mu < 0$$

$$H(x,\bar{c}) \equiv [1 - F(x)]\, x > 0$$

and

$$\frac{\partial H(x,y)}{\partial y} = F'(y)(x + \mu) > 0$$

Therefore, there is a unique $y \in (0, \bar{c})$ that satisfies $H(x,y) = 0$. We may write $y = \phi(x)$, where $H(x, \phi(x)) \equiv 0$ for all $x \in [0, \bar{c}]$. Notice that for all $x > 0$, $\mu \to 0$ implies $\phi(x) \to x$.

Next, we turn to the G function. We have

$$G(x,0) \equiv -F(x)d$$

so that there is a unique $x \in [0, \bar{c}]$, namely, $x = 0$, such that $G(x, 0) = 0$. Let c^{med} denote the median type $(F(c^{med}) = 1/2)$. We have

$$G(x, c^{med}) \equiv \left[1 - 2 \left(\frac{1}{2} - F(x) \right) \right] c^{med} - F(x)d = F(x) \left(2c^{med} - d \right)$$

Notice that $2c^{med} = c^{med}/F(c^{med}) \leq d$ by the multiplier assumption. If $c^{med}/F(c^{med}) = d$ then $G(x, c^{med}) = 0$ for all $x \in [0, \bar{c}]$. However, if $c^{med}/F(c^{med}) < d$, then there is a unique $x \in [0, \bar{c}]$, namely $x = 0$, such that $G(x, c^{med}) = 0$.

Now suppose $0 < y < c^{med}$. Then

$$G(0, y) \equiv [1 - 2F(y)] y > 0$$

and

$$G(y, y) \equiv y - F(y)d \leq 0$$

by assumption. Moreover, for $y < c^{med}$

$$\frac{\partial G(x, y)}{\partial x} = F'(x)(2y - d) = F'(x) \left(\frac{y}{F(c^{med})} - d \right) < F'(x) \left(\frac{y}{F(y)} - d \right) \leq 0 \quad (17)$$

using the multiplier condition. Hence, for each $y \in (0, c^{med})$, there is a unique $x \in (0, y]$ such that $G(x, y) = 0$. We may write $x = \theta(y)$, where $G(\theta(y), y) \equiv 0$ for all $y \in (0, c^{med})$.

Claim. *If $0 < x \leq y$ and x and y are sufficiently close to zero, then*

$$0 < \frac{d\theta(y)}{dy} < 1$$

Proof. Totally differentiating $G(\theta(y), y) \equiv 0$, we obtain

$$\frac{d\theta(y)}{dy} \frac{\partial G(x, y)}{\partial x} + \frac{\partial G(x, y)}{\partial y} = 0$$

We calculate

$$\frac{\partial G(x, y)}{\partial x} \equiv F'(x)(2y - d)$$

and

$$\frac{\partial G(x, y)}{\partial y} \equiv 1 - 2(F(y) - F(x)) - 2F'(y)y$$

Therefore,

$$\frac{d\theta(y)}{dy} = -\frac{\partial G(x, y)/\partial y}{\partial G(x, y)/\partial x} = \frac{1 - 2(F(y) - F(x)) - 2F'(y)y}{F'(x)(d - 2y)} \quad (18)$$

17

For small enough x and y, (18) is strictly positive since both numerator and denominator are strictly positive. To show that (18) is strictly smaller than 1, it suffices to show that

$$\frac{1 - 2\left(F(y) - F(x)\right) - 2F'(y)y}{F'(x)\left(d - 2y\right)} < \frac{1}{F'(x)d} \tag{19}$$

since, for small enough x, $F'(x)d > 1$ by Lemma 6. But (19) is equivalent to

$$\left(F'(y)d - 1\right)y + \left(F(y) - F(x)\right)d > 0 \tag{20}$$

The first term in (20) is strictly positive for small enough y, by Lemma 6, while the second term is non-negative since $y \geq x$. Thus, (20) is satisfied. ∎

Figure 1 is a typical depiction of (16). Notice that the G function does not involve μ, hence the function θ does not involve μ either. However, for all $x > 0$, $\mu \to 0$ implies $\phi(x) \to x$. Since $0 < d\theta(y)/dy < 1$, for small enough $\mu > 0$ the two curves $y = \phi(x)$ and $x = \theta(y)$ must have an intersection arbitrarily close to the point $(0,0)$ in the positive quadrant, and with $y > x > 0$ as depicted in the Figure. This point is denoted $(x, y) = (c^L, c^H)$. Notice that $c^H > c^L$.

It remains only to show that $\mu < c^L < c^* < c^H$. Notice that for (c^L, c^H) close to zero we are guaranteed that

$$1 - F(c^H) > F(c^H) - F(c^L) \tag{21}$$

The equation $H(c^L, c^H) = 0$ implies

$$\left[F(c^H) - F(c^L)\right]c^L = \left(1 - F(c^H)\right)\mu \tag{22}$$

Since $c^L > 0$ and $\mu > 0$, (21) and (22) imply $\mu < c^L$.

Finally, let c^* solve (7). That is, let c^* satisfy

$$\left(1 - F(c^H) + F(c^L)\right)c^* = \left(1 - F(c^H)\right)\mu + F\left(c^L\right)d \tag{23}$$

Clearly c^* exists, because $1 - F(c^H) + F(c^L) > 0$. We claim that $c^L < c^* < c^H$. The equation $G(c^L, c^H) = 0$ implies

$$\left[1 - F(c^H) - \left(F(c^H) - F(c^L)\right)\right]c^H = F(c^L)\left(d - c^L\right) \tag{24}$$

Since $c^L < c^H$, (24) implies

$$\left(1 - F(c^H) - \left(F(c^H) - F(c^L)\right)\right)c^L < F(c^L)\left(d - c^L\right) \tag{25}$$

and also

$$\left(1 - F(c^H)\right)c^H - \left(F(c^H) - F(c^L)\right)c^L > F(c^L)\left(d - c^H\right) \tag{26}$$

Now substitute from (22) into (25) and (26) to get

$$\left(1 - F(c^H) + F(c^L)\right)c^L < \left(1 - F(c^H)\right)\mu + F(c^L)d$$

and

$$\left(1 - F(c^H) + F(c^L)\right)c^H > \left(1 - F(c^H)\right)\mu + F(c^L)d$$

But these two inequalities and equation (23) imply $c^L < c^* < c^H$.

References

[1] Sandeep Baliga and Stephen Morris (2000), "Coordination, Spillovers and Cheap-Talk," mimeo, Northwestern University.

[2] Banks, J. and Calvert, R. (1992), "A Battle-of-the-Sexes Game with Incomplete Information", *Games and Economic Behavior* **4**: 347-372

[3] H. Carlsson and Eric van Damme (1993), "Global Games and Equilibrium Selection," *Econometrica,* **61**: 989-1018.

[4] Winston Churchill (1931). *The World Crisis (abr. and revised).* New York: C. Scribner.

[5] Craig, G. (1978) *Germany 1866-1945*, Oxford University Press, New York

[6] Drew Fudenberg and Jean Tirole (1991), "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory, 53*: 236-260.

[7] Robert Jervis (1976). *Perception and Misperception in International Politics.* Princeton: Princeton University Press.

[8] Kissinger, H. (1994). *Diplomacy.* Touchstone, New York

[9] Matthews, S. and A. Postlewaite (1989) "Pre-play Communication in Two-Person Sealed-Bid Double Auctions", *Journal of Economic Theory* **48**: 238-263

[10] Stephen Morris and Hyun Shin (1998), "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review,* **88**: 587-597.

[11] Stephen Morris and Hyun Shin (1998), "Global Games: Theory and Applications," mimeo, Yale University.

[12] Barry O'Neill. *Honor, Symbols and War.* Ann Arbor: University of Michigan Press.

[13] Pigou, A.C. (1941), *The Political Economy of War.* Revised edition, MacMillan Company, New York.

[14] Ronald Reagan (1983). *Public Papers of the Presidents of the United States, Ronald Reagan, book 1.* Washington Dove.C.: U.S. Government Printing Office.

[15] Ronald Reagan (1990). *An American Life.* New York: Simon and Schuster.

[16] Ariel Rubinstein (1989), "The Electronic Mail Game: Strategic Behavior under Almost Common Knowledge," *American Economic Review,* 79, 385-391.

[17] Thomas C. Schelling (1960,1980). *The Strategy of Conflict.* Cambridge: Harvard University Press.

[18] Raymond Sontag (1933). *European Diplomatic History, 1871-1932.* New York: Appleton-Century-Crofts.

[19] Thucydides (1972). *The History of the Peloponnesian War.* London: Penguin Classics.

[20] Leonard Wainstein (1971), "The Dreadnought Gap," in Robert Art and Kenneth Waltz, eds., *The Use of Force.* Boston: Little Brown.