Conflict and Cooperation: Communication and Deception

Sandeep Baliga and Tomas Sjöström
Northwestern and Rutgers

Can Non-Binding Negotiation Reduce Conflict?

Negotiations to reduce the build-up of arms sometimes seem to work in this context and sometimes fail either as someone reneges on an agreement or because they refuse to sign:

- When, in 1912, the British discovered that the Kaiser planned to purchase new Dreadnoughts (warships), but the British felt that "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, our emphasis). Emissaries were sent to Germany to propose that neither country should build more warships. In fact, Churchill proposed the following "Naval Holiday" for 1913: "[S]uppose that Germany were to build no more ships that year...[W]e would not begin our ships until Germany has started hers. The three ships that she did not build would therefore automatically wipe out no fewer than five British potential super-Dreadnoughts. That is more than they could hope to do in a brilliant naval action."

This offer was rejected by the Kaiser.

Examples

- Unlike the Kaiser, Hitler was very willing to sign treaties promising Germany would not arm and then renege (Britain and Poland). This was a quite deliberate policy as suggested by Goebbels (reported by Kissinger):

  "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!"

- The signing of a test ban treaty by the United States and Soviet Union in 1963 appears to have been a brake on the arms race.

Examples

- While on the one hand Ronald Reagan described the Soviet Union as an "evil empire" prepared "to commit any crime, to lie, to steal" to achieve its goals (Reagan Papers) he also told Brezhnev "there is absolutely no substance to charges that the United States is guilty of imperialism or attempts to impose its will on other countries, by use of force" (Reagan Papers).
- He was aware that escalating fear could cause an arms race and of the importance of communication. He reports in his diary the message he wants to communicate in a meeting with the Soviet Foreign Minister, Andrei 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" (our emphasis, Reagan's diary)

Main Question

We interpret the conflict game as an arms race game.

- We investigate whether negotiations can prevent an arms race and consider the cheap talk extension of the conflict game, where each state makes a non-binding statement of its of intentions to arm or not to arm before the arms race game is played. We show that if the probability that a player is a dominant strategy type is sufficiently small, then there exists an equilibrium of the cheap-talk extension of the arms race game where the probability that a state acquires new weapons is close to zero. We study a conflict game where the unique equilibrium without cheap talk has an arms race with probability one. Thus, when the dominant strategy types are sufficiently unlikely, cheap talk reverses the situation completely: without cheap talk, the probability of an arms race is one - with cheap talk it is close to zero.

Model of Arms Race

- Two players (states).
- Two actions: Build new weapons (B) or No new weapons (N). Cost of acquiring new weapons for player i is c_i and payoffs are as follows:

$$\begin{array}{ccc} & B & N \\ B & -c_i & \mu - c_i \\ N & -d & 0 \end{array} \quad (1)$$

  We assume $\mu > 0$ and $d > c_i$ for each $i \in \{1, 2\}$.
- If $c_i > \mu$ and all payoffs are common knowledge, then there are two pure strategy Nash equilibria: (B,B) and (N,N). In this case, the players may be able to coordinate on the Pareto dominant equilibrium (N,N).

Model of Arms Race c'td

- However, we will assume player i's true cost of building the weapons system, c_i, is his private information or 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 $F(0) = 0$, $F'(c) > 0$ for all $c \in [0, \bar{c}]$, and $F(\bar{c}) = 1$. Assume $\bar{c} < d$.
- A type whose cost of building weapons satisfies $c_i < \mu$ will be called a dominant strategy type, since he will have a strictly dominant strategy to build. The probability that player i is a dominant strategy type is $F(\mu)$, which is close to zero if μ is small. Typically, we will assume this is the case and set $F(\mu) = \varepsilon$.
- Since $-c_i \geq -\bar{c} > -d$, B is always a (strict) best response against B and there is a Bayesian Nash equilibrium where all types choose B with probability one.

Model of Arms Race c'td

Definition

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

- Intuitively, any equilibrium will have a cut-off property: if type c_i plays B, then any type $c_i' < c_i$ (whose cost of building weapons is strictly lower) will also play B. The dominant strategy types ($c_i < \mu$) will play their dominant strategy, B, for sure. Knowing that the dominant strategy types will play B, a type which is "almost" a dominant strategy type ($\mu - c_i < 0$ but close to zero) will also play B. In turn, this "infects" other types, with slightly higher c_i, who also decide to play B. And so on. Will this contagion ever stop? Consider type $c_i > 0$. If he thinks all types that have a lower cost of building than he has will play B, then he will strictly prefer B as long as

$$\begin{aligned} & F(c_i)(-c_i) + (1 - F(c_i))\,(\mu - c_i) \\ > \quad & F(c_i)(-d) + (1 - F(c_i)) \times 0 \end{aligned}$$

This is equivalent to $S(c_i) > 0$ where

$$S(c_i) \equiv F(c_i)d - c_i + (1 - F(c_i))\mu$$

If the multiplier condition is satisfied then $S(c_i) > 0$ for any $\mu > 0$ and any $c_i \geq 0$ and there exists a unique BNE (BB).

- Conversely, if the multiplier condition is violated, then there exists a type $c_i > 0$ such that for sufficiently small μ, $S(c_i) \leq 0$. Thus, if the multiplier condition is violated then for sufficiently small $\mu > 0$ there exists a Bayesian Nash equilibrium where N is played by some types - more precisely, by all types c_i such that $S(c_i) \leq 0$.

Theorem

(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 played by some types.

- Remarks: When the multiplier condition holds, the equilibrium can be found by iterated deletion of strictly dominated strategies.
- Remark: Had the players's types been common knowledge, then they would have been able to avoid the arms race (by coordinating on (NN)) with probability $(1-\varepsilon)^2$.

Cheap Talk Extension

- Since $\mu > 0$ and $d > 0$, no matter what player i plans to do he always strictly wants player j to choose N (Aumann).
- We consider a cheap talk extension of the arms race game. There are three stages:

Stage 0: Nature determines each player's type, and this becomes his private information.
Stage 1 (Cheap Talk stage): Messages are announced simultaneously and publicly. The two messages that will be sent in equilibrium will be labelled D and H. Message D will be a "dovish" (conciliatory) message. Message H will be a "hawkish" message. (Other messages could be allowed, but will not be sent in equilibrium.)
Stage 2 (Action Stage): Players play the arms race game.

Cheap Talk Extension: Main Result

- We will now construct an equilibrium of this game to prove the following result:

Theorem

Suppose the multiplier condition is satisfied. For any $\delta > 0$ *there is* $\bar{\mu} > 0$ *such that if* $0 < \mu \leq \bar{\mu}$ *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 construction of the equilibrium requires dividing each player's type space $[0, \bar{c}]$ into three regions by specifying two cut-off points c_L and c_H such that $0 < \mu < c_L < c_H < \bar{c}$. Player *i* is said to be 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* will send message Hawk (H) if he is fairly tough. Otherwise, he sends message Dove (D).

Step 1: Why are normal types willing to say Dove and then not build? The normal types do not like building at all, and only build new weapons in the absence of negotiations because everyone who has lower costs than them , including the fairly tough types, build. Now, they recognize a fairly tough type when they see one as he announces himself by being hawkish. They can then coordinate with him and play B rather than get exploited by playing N. Therefore, when the proportion of their opponent's normal types is high enough and the probability of facing very tough types is low, they prefer to be dovish and risk the chance of being exploited by a very tough type.
Step 2: Why are fairly tough types willing to say Hawk? They could after all say D, Build weapons and then exploit the normal types. To avoid this, hawkish types must coordinate on the "good" equilibrium when they meet each other.
Step 3: Why do the very tough types say Dove? The very tough types are going to build anyway, and they just want to maximize the probability that their opponent does not build. Therefore, if the probability that their opponent is normal is high enough, they will announce D.

-

	$c_2 < c_L$ (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

- One other relevant part of the strategy profile requires defining the type c^* who is indifferent between B and N when both players send the message Dove:

$$(1 - F(c_H))(\mu - c^*) + F(c_L)(-c^*) = F(c_L)(-d). \quad (2)$$

  If both players said Dove in stage 1, then player i chooses N if and only if $c_i \geq c^*$.

- Finally, c_L and c_H are one solution to the following pair of equations:

$$[F(c_H) - F(c_L)]\, c_L - (1 - F(c_H))\, \mu = 0 \; (H(c_L, c_H) = 0) (3)$$
$$[1 - (F(c_H) - F(c_L))]\, c_H - [(F(c_H) - F(c_L))\, c_H + F(c_L) d]$$
$$= \; 0 \;\; (G(c_L, c_H) = 0).$$

  Roughly speaking, the first condition implies that type c_L is indifferent between announcing Dove and Hawk and the second that type c_H is also indifferent between announcing Dove and Hawk.

Proof

- We claim that for small enough $\mu > 0$, these strategies form a perfect Bayesian equilibrium of the cheap talk extension of the arms race game. We need a few preliminary results.

Lemma

(3) implies

$$1 - F(c_H) > F(c_H) - F(c_L). \tag{4}$$

and $c_L > \mu$ *and* $c_H > c_L$.

Proof.

First, we claim that (3) implies $c_L > 0$ and therefore $F(c_L) > 0$. Otherwise, from $H(c_L, c_H) = 0$ and $\mu > 0$, we conclude that $c_H = \bar{c}$. But then, from $G(c_L, c_H) = 0$, we obtain $-\bar{c} = 0$, a contradiction.

Next, recall $d > c$ for all $c \in [0, \bar{c}]$. Therefore, $G(c_L, c_H) = 0$ implies

$$\begin{aligned} & (1 - F(c_H) - (F(c_H) - F(c_L)))(c_H) \\ = \; & F(c_L)(d - c_H) > 0 \end{aligned}$$

which implies (4). The rest will be shown diagrammatically later. □

Proof c'td

Lemma

(3) implies that $c_L < c^* < c_H$.

Proof Notice that, conditional on both players having announced Dove in stage 1, the right hand side of (2) is the expected payoff to type c^* from choosing the action N while the left hand side is the expected payoff from the action B. We show the expected payoff to B is greater than the expected payoff to the action N for type c_L but less than the expected payoff to the action N for type c_H. Therefore, there is some intermediate type, c^*, who is indifferent between the two actions.
First, we claim

$$(1 - F(c_H))(\mu - c_L) + F(c_L)(-c_L) > F(c_L)(-d). \quad (5)$$

Notice that $G(c_L, c_H) = 0$ can be rewritten as

$$(1 - F(c_H) - (F(c_H) - F(c_L)))(-c_H) + F(c_L)(d - c_H) = 0.$$

Using (4) and the fact that $c_L < c_H$, we get

$$(1 - F(c_H))(-c_L) + (F(c_H) - F(c_L))c_L + F(c_L)(d - c_L) > 0. \quad (6)$$

$H(c_L, c_H) = 0$ can be rewritten as

$$(1 - F(c_H))\mu = (F(c_H) - F(c_L))c_L$$

Proof of lemma c'td

and substituting this in (6), we obtain

$$(1 - F(c_H))\,(\mu - c_L) + F(c_L)\,(d - c_L) > 0$$

and therefore

$$(1 - F(c_H))\,(\mu - c_L) + F(c_L)\,(-c_L) > F\,(c_L)\,(-d)$$

as claimed.

Proof of lemma c'td Next, we claim

$$(1 - F(c_H))(\mu - c_H) + F(c_L)(-c_H) < F(c_L)(-d). \qquad (7)$$

Notice that $H(c_L, c_H) = 0$ implies

$$\begin{aligned}(1 - F(c_H))(-c_L) + F(c_L)(-c_L) = & \qquad (IC(c_L)) \\ (1 - F(c_H))(\mu - c_L) - [F(c_H) - F(c_L)]\, c_L - F(c_L)c_L. &\end{aligned}$$

As $c_H > c_L$ by Lemma 4, we have

$$\begin{aligned}(1 - F(c_H))(-c_H) + F(c_L)(-c_H) > & \qquad (8) \\ (1 - F(c_H))(\mu - c_H) - [F(c_H) - F(c_L)]\, c_H - F(c_L)c_H. &\end{aligned}$$

Combining $G(c_L, c_H) = 0$ and (8), we obtain

$$\begin{aligned}&[F(c_H) - F(c_L)](-c_H) + F(c_L)(-d) = \\ &(1 - F(c_H))(-c_H) + F(c_L)(-c_H) > \\ &(1 - F(c_H))(\mu - c_H) - [F(c_H) - F(c_L)]\, c_H - F(c_L)c_H\end{aligned}$$

which can be re-written as

$$(1 - F(c_H))(\mu - c_H) + F(c_L)(-c_H) < F(c_L)(-d)$$

as claimed. Combining (5) and (7), we find c^* as claimed.

Back to Main Proof!

Given these results, the strategies specified in the actions stage are sequentially rational.

We now turn to the cheap talk stage. Now, we make a series of claims.

Claim 1. Player i of type $c_i \leq \mu$ prefers to announce Dove rather than Hawk.

Proof. As player i of this type chooses B for certain, his objective is simply to maximize the probability of his opponent choosing N. If he announces Hawk, player j will choose N if and only if he himself announced Hawk, an event occurs with probability $F(c_H) - F(c_L)$. If he announces Dove, player j will choose N if and only if he himself announced Dove, an event occurs with probability $1 - F(c_H)$. Therefore, by (4) (i.e. $1 - F(c_H) > F(c_H) - F(c_L)$), it is optimal for player i of type c_i to announce Dove.

Proof c'td

Claim 2. Player i of type c_L is indifferent between announcing Hawk and Dove. Also, player i of type c_i where $c_L \leq c_i \leq c^*$ prefers to announce Hawk rather than Dove and player i of type c_i where $\mu \leq c_i \leq c_L$ prefers to announce Dove rather than Hawk

Proof. Player i, if he announces Hawk, chooses N in the action stage if and only the other player also announces Hawk. As this event occurs with probability $F(c_H) - F(c_L)$, the expected payoff of player i type c_i from announcing Hawk is

$$(1 - F(c_H))(-c_i) + F(c_L)(-c_i). \tag{9}$$

If he announces Dove, player i of type $c_i \leq c^*$, will Build at the action stage whatever player j announces. Player j of type c_j will choose N if and only if he himself has also announced Dove and $c_j \geq c_H$, an event that occurs with probability $1 - F(c_H)$. Therefore, player i's expected payoff from announcing Dove is

$$(1 - F(c_H))(\mu - c_i) + [F(c_H) - F(c_L)](-c_i) + F(c_L)(-c_i). \tag{10}$$

But, by $H(c_L, c_H) = 0$ (i.e. $F(c_H) - F(c_L)c_L - (1 - F(c_H))\mu = 0$) (9) is exactly equal to (10) if $c_i = c_L$ so player i of type c_L is indifferent between announcing Hawk and Dove. If $c_i \geq c_L$, (10) is smaller than (9) so player i of type c_i prefers to announce Hawk rather than Dove. If $\mu \leq c_i \leq c_L$, (10) is higher than (9) so player i of type c_i prefers to announce Dove rather than Hawk.

Proof c'td

Claim 3. Player i of type c_H is indifferent between announcing Hawk and Dove. Also, player i of type c_i where $c_H \leq c_i \leq \bar{c}$ prefers to announce Dove rather than Hawk and player i of type c_i where $c^* \leq c_i \leq c_H$ prefers to announce Hawk rather than Dove.

Proof. Player i, if he announces Hawk, will play N if and only the other player also announces Hawk. As this event occurs with probability $F(c_H) - F(c_L)$, the expected payoff of player i type c_i from announcing Hawk is

$$(1 - F(c_H))(-c_i) + F(c_L)(-c_i). \qquad (11)$$

If he announces Dove, player i of type c_i where $c_i \geq c^*$,will choose N at the action stage if and only if player j also announces Dove. Player j of type c_j will announce Dove either if $c_j \leq c_L$ or if $c_j \geq c_H$ and will choose Build if and only if both players announced Dove and $c_j \geq c_H$, an event that occurs with probability $1 - F(c_H)$. Therefore, player i's expected payoff from announcing Dove is

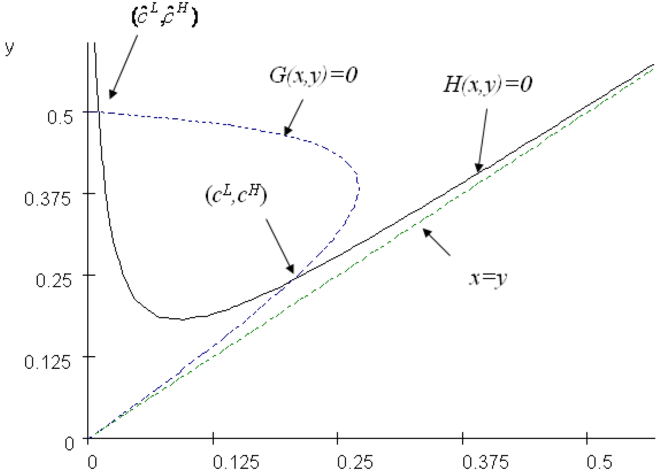
$$[F(c_H) - F(c_L)](-c_i) + F(c_L)(-d). \qquad (12)$$

But, by $G(c_L, c_H) = 0$, (11) is exactly equal to (12) if $c_i = c_H$ so player i of type c_H is indifferent between announcing Hawk and Dove. By (4), as $1 - F(c_H) > F(c_H) - F(c_L)$, if $c_i \geq c_H$, (11) is less than (12) so player i of type c_i prefers to announce Dove rather than Hawk. If $c \leq c_i \leq c_H$, (11) is higher than (12) so player i of type c_i prefers to announce Hawk rather than Dove.

Proof c'td

- These claims show that the announcement strategies in stage 1 are in equilibrium given the actions strategies in stage 2. As the latter were already shown to be sequentially rational, the strategy profile specified is a perfect Bayesian equilibrium of the cheap talk extension of the arms race game.
- We complete the analysis by (diagrammatically) studying the behavior of $(*)$ and $(**)$ for small μ

Figure 1



$(\hat{c}^L, \hat{c}^H)$
y
$G(x,y)=0$
$H(x,y)=0$
(c^L, c^H)
$x=y$
0.5
0.375
0.25
0.125
0
0
0.125
0.25
0.375
0.5

Conclusions

- Aumann intuition implies there is some deception in equilibrium.
- However, it is possible to generate enough information to play NN with positive probability.
- Intuition: Although all types want their opponent to play N, coordination types also want to learn the opponent's action to coordinate against it. "Strong" coordination types are willing to pay a price in terms of greater risk of B to learn information. This allows us to separate them out, prevent the spiral logic from fully taking hold and create coordination on NN.

Strategic Ambiguity and Arms Proliferation

- On the American side, there is a fear that arms will fall into the wrong hands but on the other side there is fear of American motives:

  "The U.S. is after an excuse. If we stop nuclear technology, they will find another excuse. They invaded Iraq even though there were no weapons (of mass destruction)" (Habibollah Hosseini, Iranian cleric).

  "In the contemporary world, it is obvious that having access to advanced weapons shall cause deterrence and therefore security, and will neutralize the evil wishes of great powers to attack other nations" (Jumhuri-ye Islami).

- One way to reduce fear is to create deterrence by revealing your strength.
- Traditional Cold War logic would suggest that they should reveal arms if they have them. For example, Dr. Strangelove was dismayed to discover that the Russians had not revealed the existence of their "Doomsday machine": "the whole point of the Doomsday Machine is lost if you keep it a secret. Why didn't you tell the world, eh?"
- But Israel and Iraq maintained "strategic ambiguity." How and why is ambiguity maintained and what are its consequences?

- In the 1960s, if Israel had revealed its nuclear program, it might have triggered an attack.
  - Nasser threatened to invade Israel should he discover it was developing nuclear weapons (NYT, Dec 4, 1960).
  - If Saddam *did have* WMDs and he had revealed them it might have triggered fear and attack from the other side - maybe Saddam will sell them to Osama, use them against Israel etc.

- If Israel had cancelled its nuclear program and revealed no WMDs, that would have eliminated “deterrence by ambiguity”
    - Israeli Prime Minister Eshkol’s view on publicly revealing absence of nuclear capability: “[I]t does not appear advisable to release President Nasser from any apprehension he may entertain as to Israel’s nuclear capability. Nasser loses no opportunity of publicly emphasizing that war with Israel is inevitable...” (Cohen *Israel and the Bomb*, p. 199)
    - If Saddam did *not* have WMDs and revealed this, he might have thought this would give the US or someone else the confidence to attack.
- Hence, *strategic ambiguity* may be forced upon a country- opponents do not know if you have WMDs or not.

- Should we force countries to reveal their arsenal? Does strategic ambiguity make the world more dangerous? Does it help America or Iran? Washington Post (May 11, 2003):

  "Does an American policy to deny unfriendly nation-states the policy option of creating ambiguity around WMD possession and the support of terrorism make the world a safer place? The Bush administration has made a game-theory-like calculation that it does.

  In fact, WMD ambiguity was at the core of Iraq's strategy...[I]f it ever became unambiguously clear that Iraq had major initiatives underway in nuclear or bio-weapons, America....might intervene militarily. If....it ever became obvious that Iraq lacked the unconventional weaponry essential to inspiring fear.., then the Kurds, Iranians and Saudis might lack appropriate respect for Hussein's imperial ambitions. Ambiguity thus kept the West at bay while keeping Hussein's neighbors and his people in line."

- This conventional wisdom that ambiguity reduces “world safety” also captured by Nuclear Non-Proliferation Treaty:

ARTICLE III

Each Non-nuclear-weapon State Party to the Treaty undertakes to accept safeguards...for the exclusive purpose of verification of the fulfilment of its obligations assumed under this Treaty with a view to preventing diversion of nuclear energy from peaceful uses to nuclear weapons or other nuclear explosive devices.

The Model

- Two players, A and B. A is deciding whether to be aggressive or not against B. A has a type that determines its level of aggression. B has a type that represents its motives for acquiring weapons. B's investment and weapons' status are unknown.
- If B invests, then he will acquire (rudimentary) nuclear weapons with probability $\sigma \in (0, 1)$. The cost of investing is $k > 0$.
- B's knows his nuclear capability and his type.
- His nuclear capability is *hard* information that can be verified by (perfectly reliable) weapons inspectors. But B's type is private information and is *soft* information.
- If there is an inspection, then B incurs a small cost $\varepsilon \in [0, \bar{\varepsilon}]$.

- A does not observe B's investment, the success of his weapons program or his type
- Finally, A decides whether or not to be aggressive against B. A derives a private benefit a if he is aggressive. a has a continuous distribution with support $[a_0, a_1]$, where $a_0 < 0 < a_1$. Density is f. B can be "crazy" type z with probability τ or a "normal" type n with probability $1 - \tau$. Other details of payoffs depend on whether B has weapons or not.
- Payoffs

	B has nukes	No nukes
A aggressive	$a - c, -\alpha + \gamma$	$a, \ -\alpha$
A passive	$-d_t, \ \delta_t$	$0, \ 0$

where $d_z > d_n > 0$ and $\delta_z > \delta_n = 0$. Define $\kappa \equiv k/\sigma\gamma$ as the *normalized cost of investing.*

- Payoffs

	B has nukes	No nukes
A aggressive	$a-c, -\alpha+\gamma$	$a, \ -\alpha$
A passive	$-d_t, \ \delta_t$	$0, \ 0$

- For B, weapons used for defense and can be used to counterattack. For A, aggression is costly if B has arms:

$$\alpha > \gamma > 0$$
$$c > 0$$

- The objective of our research is to understand how incomplete information about the opponent's motives and capabilities can trigger arms races and conflicts. Accordingly, we will assume that player A's optimal decision depends on his preferences (his type) and his beliefs.

- Payoffs

	B has nukes	No nukes
A aggressive	$a-c, -\alpha+\gamma$	$a, \ -\alpha$
A passive	$-d_t, \ \delta_t$	$0, \ 0$

- Player A is a *dove* if $a < 0$, an *opportunist* if $0 < a < c - (\tau d_z + (1-\tau) d_n)$ (strategic substitutes), and a *hawk* if $a > c - (\tau d_z + (1-\tau) d_n)$. The probability that A is a dove is $D \equiv F(0) > 0$. The probability that he is a hawk is $H \equiv 1 - F(c - (\tau d_z + (1-\tau) d_n)) > 0$.
- Since $a - c < -(\tau d_z + (1-\tau) d_n)$, the opportunist fears B's nuclear arsenal and can be deterred by it, but $a > 0$ implies that the opportunist is aggressive if he thinks B is unarmed. A's optimal decision depends on his beliefs.

Parametric Assumptions

$$\textbf{Assumption 1}: \tau d_z + (1-\tau)\, d_n < c < d_z.$$

- The first inequality in Assumption 1 says that the cost of being aggressive when B has rudimentary nuclear weapons exceeds the *ex ante* expected value of eradicating the threat. The second inequality implies that if A is *certain* that B is crazy then the cost of being aggressive is *smaller* than the expected value of wiping out the threat (strategic complements).
- Our second assumption is that the cost of inspections $\varepsilon \in [0, \bar{\varepsilon}]$ is small. And finally:

$$\textbf{Assumption 3}: \frac{k}{\sigma} < (1 - F(c - d_z))(-\alpha + \gamma) + (1 - F(0))\alpha - \bar{\varepsilon}$$

If Assumption 3 is violated, then the cost of investing is so high that there is an equilibrium where the normal type of B never invests and always allows arms inspections.

- In fact, as a crazy type gets a greater benefit from arming, using Assumption 3, we can show

Proposition 1 In any perfect Bayesian equilibrium, the crazy type of player B invests with probability one.

Timing

- Time 0: A privately learns a. B privately learns $t \in \{z, n\}$
- Time 1: Communication stage.
- Time 2: B decides whether or not to invest in a weapons program. If he invests, then B privately observes the success or failure of the program. A cannot observe anything that happens at time 2.
- Time 3: B decides whether or not to allow inspections. If inspections take place, the inspectors publicly announce whether or not nuclear capability exists.
- Time 4: A decides whether or not to be aggressive.

No informative communication: Full Disclosure

- B always allows inspections or allows inspections if and only if he is armed.
- Cheap-talk cannot be effective: All of A's types want to minimize the probability that B invests in arms. Hence, the probability of arming cannot depend on the message they send.
- As communication is uninformative, under Assumption 3, the normal type of B invests with probability 1 under full disclosure.

Proposition 2 There is an equilibrium with full disclosure. Full disclosure implies B invests with probability 1 and there is no informative cheap-talk.

Full ambiguity

- B never allows inspections.
- Again, for the same reason, cheap-talk cannot be ineffective

Proposition 3 There is an equilibrium with full ambiguity. Full ambiguity implies B invests with probability $\tilde{x} \in (0,1)$ if

$$\kappa > 1 - F(\sigma(c - \tau d_z - (1-\tau)d_n)), \qquad (1)$$

and probability 1 otherwise. Cheap-talk cannot reduce the probability that B invests.

- This equilibrium has a cutoff property where A is aggressive if and only if $a \geq \tilde{a}$ where $\tilde{a}$ is indifferent between his two actions:

$$\tilde{a} - \sigma(\tau + (1-\tau)\tilde{x})c = -\sigma(\tau d_z + (1-\tau)\tilde{x} d_n) \text{ or}$$
$$\tilde{a} = \sigma\tau(c - d_z) + \sigma(1-\tau)\tilde{x}(c - d_n).$$

- Condition (1) guarantees that $\tilde{x}$ is interior and $c - \tau d_z - (1-\tau)d_n > \tilde{a} > 0$.

Welfare: Full Ambiguity vs. Full Disclosure.

- A's welfare depends on the normalized cost of investing κ and the preferences of type $\tilde{a}$ who is most interested in information. He prefers full ambiguity if and only if

$$(1-\sigma)\tilde{a}-\sigma\left(\tau d_z+(1-\tau)\,d_n\right)<\tilde{a}-\sigma\left(\tau+(1-\tau)\,\tilde{x}\right)c \quad (*)$$

- Note $\tilde{x}$ is lower the higher is κ. Hence if κ is high (Case 1) , all of A's types who are opportunists prefer full ambiguity as it reduces the chances of B arming. Doves and Hawks just want to minimize the chances B invests and hence all of A's types prefer full ambiguity to full disclosure.
- (Case 2) If κ is low, there is conflict between A's types as full ambiguity does not reduce the chances of B arming significantly.

- With full ambiguity, A makes mistakes: sometimes opportunists are aggressive even when B is armed, and sometimes they are not aggressive when he is not armed. B prefers full ambiguity if the benefit to deterring opportunists when B is unarmed is greater than the cost of being attacked by some opportunists who do not observe that B is armed. I.e. B prefers full ambiguity to full disclosure if and only if the following expression is negative:

$$\sigma(\alpha - \gamma)\left[F(c - \tau d_z - (1-\tau)d_n) - F(\tilde{a})\right] - (1-\sigma)\,\alpha\left[F(\tilde{a}) - F(0\right.$$

Proposition 4 Full ambiguity dominates full disclosure for A and the normal type of B if and only if $(*)$ holds and $(**)$ is negative.

Equilibrium with partial disclosure

- Inspections generate information about B's nuclear capability. Hawks and Doves do not act on information generated by invests inspections but opportunistic types do.
- This allows us to construct an equilibrium with two informative messages. The intermediate types send a "tough" message that leads to inspections, but also induces B to invest (the message proves to B that A is an opportunist). The extreme types send a "conciliatory" message. After hearing the conciliatory message, B is less likely to invest, but there will be no inspection.
- The communication equilibrium exists if two conditions are satisfied. First, the normalized cost of investing must be low (Case 2 above). Otherwise, all of A's types would prefer ambiguity, and no-one would send the "tough" message. Second, the prior probability that A is a hawk must be small. Otherwise, B would invest for sure after the conciliatory message, since it does not distinguish hawks from doves.

Proposition 5 Suppose κ is low and so is $\frac{H}{H+D}$. There is a communication equilibrium where, for some a' and a'', player A sends a "tough" message if $a \in (a', a'')$ and a "conciliatory" message otherwise. Player B invest with probability 1 and allows inspections if he is armed following the tough message. After the conciliatory message, B invests with positive probability and always refuses inspections.

- If A sends the tough message, B knows A is an opportunist. Therefore, B invests.
- If A sends the conciliatory message, B refrains from investing with probability 1 if A is unlikely to be a hawk. However, B cannot invest probability zero, because then also the opportunists would prefer the tough message. Hence, the cost of going nuclear has to be low enough.

Welfare: Ambiguity vs. Disclosure

- If the partial disclosure equilibrium exists then all of A's types prefer partial to full disclosure (by revealed preference). If the partial disclosure equilibrium doesn't exist, but κ is high, all of player A's types prefer full ambiguity.
- Since $0 < a' < a'' < c - \tau d_z - (1-\tau) d_n$, some opportunistic types send conciliatory message and then are aggressive even though B is armed (there is no inspection). B would have deterred aggression by full disclosure. On the other hand, partial ambiguity protects an unarmed B from aggression from opportunists, who are deterred by the chance that B is armed.

- Hence, B's prefers partial to full disclosure if and only if

$$\sigma(\alpha-\gamma)[F(c-\tau d_z-(1-\tau)d_n)-F(a'')]-(1-\sigma)\,\alpha[F(a')-F(0)$$

is negative.

- There was a similar trade-off with full ambiguity vs. full disclosure. *Thus, B may or may not want ambiguity.*

- Finally, we consider whether other communication equilibria exist.
- We study equilibria which are robust to a small random cost of inspections $\varepsilon \in [0, \bar{\varepsilon}]$ to B.
- An equilibrium has *effective cheap-talk* if A's type affects the probability that B invests.

Proposition 6 All equilibria with effective cheap-talk can be replicated by a two message equilibrium where, for some a' and a'', player A sends a "tough" message if $a \in (a', a'')$ and a "conciliatory" message otherwise. After the tough message, player B invest with probability 1 and allows inspections iff he is armed. After the conciliatory message, the normal type of B invests with positive probability and refuses inspections with positive probability if he is armed.

- Let M be some message space and fix an equilibrium of the game. Let M^c be the set of messages that minimize the probability that the normal type of B invests and let $M^t = M \backslash M^c$.
- First, it must be the case that hawks and strong opportunists, who are always aggressive in equilibrium, and doves and weak opportunists, who are never aggressive, send messages in M^c to minimize the probability that (normal) B invests.

- Second, if any other message $m^t \in M^t$ is sent in equilibrium, it reveals the sender is an opportunist who values information. In particular, the opportunist is aggressive if B is unarmed or refuses inspections and is not aggressive if inspections reveal B is armed. This maximizes B's incentive to arm and reveal his invests if he is successful. As all such messages lead to the same outcome, we can assume there is just one such tough message m^t.
- Third, B must arm with positive probability in response to any message $m^c \in M^c$. If he invests with probability 1, the equilibrium does not have effective cheap-talk. If he invests with probability zero, there is no incentive for an opportunistic type of A to send the message m^t and the equilibrium does not have effective cheap-talk. Also, the armed normal type must refuse inspections with positive probability in response to $m^c \in M^c$. Suppose not. The armed crazy type always has a greater benefit to inspections and hence both types will allow inspections. But then there is again no incentive to send message m^t.

- Finally, all messages in M^c lead to B refusing inspections with positive probability when he is armed and arming with the same probability. Hence, we can assume there is only one such conciliatory message m^c.

Conclusion

- We suggest that the welfare of a country facing a potential invests proliferator may be increased by allowing the latter to maintain ambiguity. The welfare of the proliferator may also go up.
- The aggressor faces face trade-off between better information and higher probability that the opponent invests. This generates equilibria where types who value information (opportunists) demand inspections and types who do not (Hawks and Doves) allow ambiguity.
- Extensions: Taiwan-China-U.S., Obama?