



WORKING PAPER SERIES

2008/01

A Brinkmanship Game Theory Model of Terrorism

Dr. Francois Melese
Defense Resources Management Institute
Naval Postgraduate School

Defense Resources Management Institute
School of International Graduate Studies
Naval Postgraduate School
Building 234
699 Dyer Road
Monterey, CA 93943-5138
831-656-2306
www.nps.edu/drmi

The views herein expressed reflect the views of the author and are not the policy of DRMI, the Naval Postgraduate School, or the U.S. Government.



Defense Resources Management Institute Naval Postgraduate School

The Defense Resources Management Institute (DRMI) is an educational institution sponsored and supervised by the Secretary of Defense and located at the Naval Postgraduate School in Monterey, California. Since 1965 the Institute has conducted professional education programs in analytical decision making and resources management for military officers of all services and for senior civilian officials of the United States and 162 other countries.

The mission of the Defense Resources Management Institute is to educate civilian and military government officials in the principles and concepts of defense resources management.

The unique course offerings create an interdisciplinary, interactive learning environment that facilitates the exchange of ideas among U.S. and international participants in Monterey, California, and locations worldwide. These programs are presented on a regularly scheduled basis at DRMI in Monterey and by specific arrangement at other locations in the United States and in other countries.

The success of DRMI reflects the breadth and depth of the in-house technical expertise that DRMI can draw upon. The DRMI faculty, a community of scholars and practitioners, are experts in defense resources management and has published in major academic and technical journals. The faculty has extensive experience in providing technical assistance and educational programs worldwide over the past 40 years. Our educational strategy is one of collaboration, tailored to the specific environment of the participant.

The Defense Resources Management Institute specializes in four broad educational areas:

- Economic analysis of the public sector
- Operations research of public sector decision problems
- Public budgeting and fiscal management
- Defense organizations and management methods

For more information about our educational and technical assistance activities, please visit our website at <http://www.nps.edu/drmi> or email us at drmiadmin@nps.edu.

A Brinkmanship Game Theory Model of Terrorism

Dr. Francois Melese
Professor of Economics
Defense Resources Management Institute (DRMI)
Naval Postgraduate School
Monterey, CA
fmelese@nps.edu

Brinkmanship: “The practice, especially in international politics, of seeking advantage by creating the impression that one is willing and able to push a highly dangerous situation to the limit rather than concede.”

American Heritage Dictionary (4th Ed. 2000)

I. Introduction

The nightmare scenario in the war on terror is that terrorists unleash chemical, biological or nuclear weapons against innocent civilians. This paper examines conditions under which a world leader such as the Secretary General (or Security Council) of the United Nations (UN) might credibly issue a brinkmanship threat of preemptive action to deter sovereign states or transnational terrorist organizations from acquiring weapons of mass destruction (WMD).¹

A historical precedent and classic illustration of a brinkmanship game is the Cuban Missile Crisis (see Blight & Welch 1989; Dixit & Skeath 1998; Kagan 1995; Fursenko & Naftali 1997; May & Selikow 1997). In 1962 the USSR began to install nuclear ballistic missiles in Cuba pointed at the US. Swift reaction by the US nearly led to an all out nuclear war. The brinkmanship strategy played (explicitly or implicitly) by President Kennedy involved exposing the USSR, and the US, to a gradually increasing risk of mutual harm (Mutually Assured Destruction or “MAD”), partly outside the President’s control.

¹ The generic term “terrorist” will be used for both state and non-state actors that acquire WMD.

As soon as the missile sites were discovered, one of Kennedy's top military advisors (Air Force General Curtis Lemay) recommended bombing Cuba. But Kennedy held back. Instead, the President decided to blockade Cuba to prevent any further shipments of arms and equipment from the USSR. He simultaneously issued a credible threat to Premier Khrushchev. If the USSR did not withdraw its missiles from Cuba, the President indicated there was a serious risk he would face such intense political pressure from his military advisors and the U.S. public that he would be forced to bomb the Cuban missile sites. Understanding the threat of nuclear war implied in the brinkmanship game being played by the U.S.—combined with a face-saving gesture from the US to withdraw its missiles from Turkey—the USSR eventually backed down. Although risking a nuclear holocaust, Kennedy's brinkmanship strategy was ultimately successful in deterring the USSR.²

Today the world faces a different threat: that transnational terrorist organizations or state actors will acquire WMD and deploy them against innocent civilians and opportunistic military targets. Much like the brinkmanship game played by President Kennedy during the Cuban Missile Crisis, this paper considers the possibility the UN might issue a credible, probabilistic threat to deter the use of WMD by state and non-state (terrorist) actors.³

While President Kennedy leveraged the aggressive stance of General Curtis Lemay (his "loose cannon") to establish credibility in issuing his threat to the USSR, the

² It was later learned Cuba had an agreement with the USSR that provided for retaliation in the event the US attacked Cuba.

³ Suppose an outright threat to eliminate terrorists that acquire WMD is too dangerous to be tolerable to the world community, say due to the likelihood of significant collateral damage to innocent civilians. Then this threat can be reduced, yet still remain credible, by creating a probability rather than a certainty terrorists will suffer dire consequences if they do not comply.

UN Secretary General could point to the credible threat of U.S. retaliation and its policy of preemption in the war against terror. The UN Secretary General could issue the following warning: “Acquire WMD and all bets are off.” Specifically, the UN Secretary General could warn that if it is discovered that a terrorist organization has acquired WMD the Security Council will pass a binding resolution condemning that action and open the way for a “coalition of the willing” to engage in swift and harsh “preemptive retaliation”—regardless of how “justified” the terrorists’ cause might be.⁴

What is clear is that any attempted preemptive move against elusive terrorist targets is risky (especially against those with newly acquired WMD capability). A credible threat of preemptive retaliation against terrorists having acquired WMD is likely to weigh heavily on terrorist organizations. Preemptive actions will have very bad consequences (negative payoffs) on the terrorist organization, but those retaliating are also clearly at risk, as are innocent civilians. In fact, in the event of substantial collateral damage to innocent civilians the blame could ultimately shift onto those engaged in retaliation (for example, onto the United States and its “coalition of the willing”). Nobody is likely to escape harm.

A Brinkmanship strategy always involves a probabilistic strategic action (preemptive retaliation in this case) that has a mutually harmful outcome. Retaliation might be bad for the terrorists, but it is also bad for those that carry out the threat. The objective is to make the brinkmanship threat a sufficiently credible and unpleasant option

⁴ The term “preemptive retaliation” will occasionally be reduced to simply “retaliation” in what follows. The term is meant to indicate retaliation against the acquisition of WMD and preemption to prevent the use of WMD. A further refinement of the model might include all of the many steps involved in issuing and implementing a credible, internationally sanctioned threat of preemptive retaliation. For instance, faced with evidence a terrorist group has acquired WMD, what is the probability the Security Council passes a resolution (i.e. avoids a veto) that sanctions preemptive retaliation?

that it deters terrorists from using WMD, while not being so catastrophic and repugnant to those that would have to carry it out that they would refuse to do so. The key is to identify conditions under which issuing a brinkmanship threat is “incentive compatible”...in the sense that it aligns interests of a Terrorist (Agent) with the UN (Principal)—namely to renounce any attempts to acquire and use WMD.

Given the probabilistic nature (say $0 \leq q \leq 1$) of the threat of retaliation, there is still a small probability (1-q) the UN Security Council might not pass a resolution condemning the WMD attack and/or that the US and its allies would not engage in a worldwide crackdown on the perpetrators.⁵ To complete the brinkmanship strategy, the UN Secretary General needs to guarantee that if a terrorist organization complies and renounces the use of WMD, world condemnation of their actions would not be as severe, and (say for example in the case of Palestinian organizations) that the UN would continue to play a role in addressing and negotiating reasonable grievances that underlie their actions. According to Bruce Bueno de Mesquita of Stanford’s Hoover Institute:

“[I]t is a mistake to view all terrorist organizations as being alike. I see two types of terrorists: One type I refer to as “true believers” [**HARD**] and the other as “relatively reasonable terrorists” [**SOFT**]...The first steadfastly pursues a goal and has no interest in compromises or concessions...Relatively reasonable terrorists, in contrast, are interested in bringing their real or perceived plight to the attention of others; they seek greater understanding and hope for concessions but may accept a resolution...that is far short of all-out victory.”

www.hooverdigest.org/021/bdm.html 6/27/06)

A key challenge in brinkmanship is to know whether we are engaging a HARD or SOFT terrorist type. The basic outline of the brinkmanship model is described in Table 1.

⁵ This could be modeled more explicitly, but would add complexity and thus will be saved as a future refinement of the model. For example, including the probability the UN Security Council passes a resolution that sanctions preemption might be combined with the probability member countries can build a “coalition of the willing” to enforce that resolution.

TABLE 1: The Brinkmanship Game

- Sequential (Extensive Form) Two Player Game
 1. Terrorists (agent)
 - Unknown Type
 - Nature determines: **Hard** (probability= p) or **Soft** (probability= $1-p$)
 2. UN (principal)
 - Brinkmanship Decision: Threaten Preemption or Not Threaten
 3. Terrorists (agent)
 - WMD Decision: Defy (Pursue WMD) or Comply (Not Pursue WMD)
 4. UN (principal)
 - Brinkmanship: If UN does Threaten Preemption and Terrorists Defy => Preemption will occur with probability= q , or not with probability= $1-q$

The next section describes the model in detail and illustrates the extensive form of the brinkmanship game. Section 3 derives two incentive compatibility or “credibility constraints” that must be satisfied for the UN to adopt a brinkmanship strategy. The first “Effectiveness Constraint” relates to the type of terrorist organization one is up against (Hard or Soft), and the second “Acceptability Constraint” relates to risks the World might be willing to take with a policy of preemption. Section 4 offers a graphical solution and interpretation of the results. This reveals when a brinkmanship threat of preemption is credible (a set of incentive compatible Nash equilibrium solutions), and when it is not. We conclude with some policy guidance and recommendations for future research.

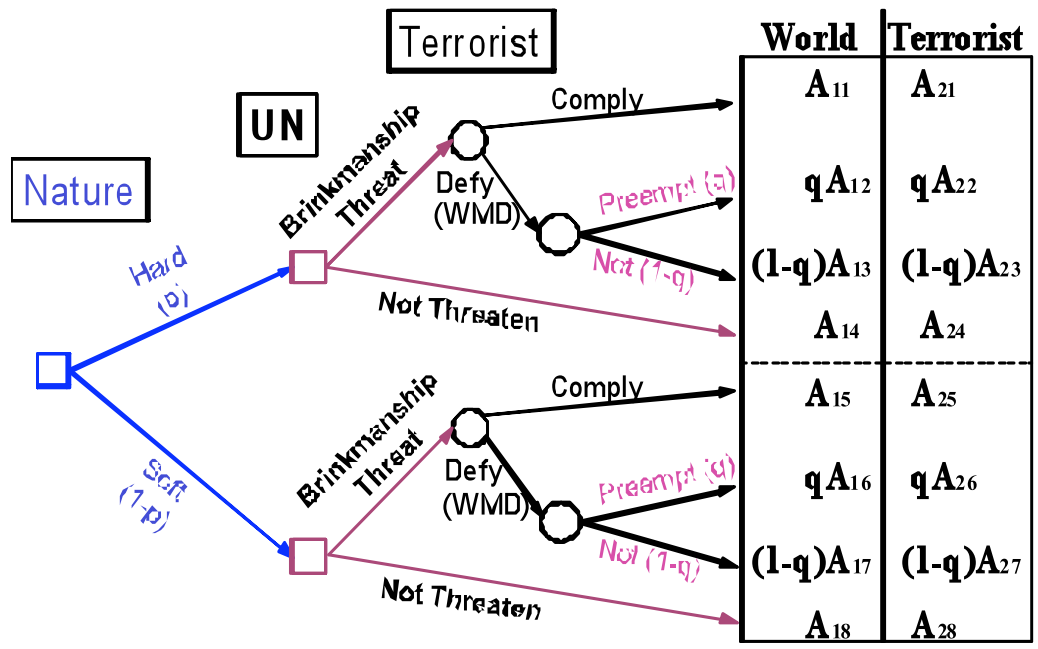
II. The Extensive Form of the Brinkmanship Game

Consider two players: 1) the World as represented by the UN (the “Principal” in this game), and 2) a representative Terrorist organization or rogue state (the “Agent”). The UN’s objective is to identify conditions under which a brinkmanship threat (much

like an incentive-based contract) is incentive compatible, in the sense that it aligns the interest of the Terrorist (Agent) with the rest of the World (the Principal)—namely, to renounce the use of WMD. In reality, the UN does not know the type of terrorist organization it faces. Following Harsanyi (1967), the UN views the terrorist’s type as randomly determined by “Nature,” drawn from a distribution of types that are common knowledge. For ease of exposition we assume two types: “Hard” and “Soft.” Since the UN/World does not know which player it is up against, based upon its best information, it assigns a probability, $0 \leq p \leq 1$, it is playing against a “Hard” terrorist and the probability, $(1-p)$, it is playing against a “Soft” terrorist. Figure 1 illustrates the payoffs for each terrorist type (Hard and Soft), and the corresponding payoffs for the World community. This represents the extensive form of a non-cooperative brinkmanship game under incomplete information (see Friedman 1986 and Rasmusen 1995).

Figure 1

[Brinkmanship Game Tree: Probabilities & Payoffs]



The distinction between Hard and Soft terrorists arises in terms of how each organization assesses its payoffs for various outcomes of the game. Hard terrorists are assumed to have a more optimistic view of their expected return from defying a brinkmanship threat than Soft terrorists. Consequently, a Hard terrorist group is more difficult to deter.

Given that “Nature” determines the type of terrorist it faces, the UN must decide whether or not to adopt a brinkmanship strategy. (See Figure 1) The key challenge for the UN in issuing a brinkmanship threat is that it must be credible. As Schelling (2002) emphasizes: “Talk is not a substitute for moves... Talk can be cheap when moves are not.” (p.117)

A threat aimed at deterrence will fail if it is not credible. Two “credibility constraints” define the equilibrium solution in this game: an “Effectiveness constraint” and an “Acceptability constraint.” (In a different context, see Dixit & Skeath 1998)

To be “effective” the threatened risk of retaliation, $0 \leq q \leq 1$, must be high enough to at least deter the Soft terrorist from acquiring WMD, and yet still small enough (since retaliation involves a mutually harmful outcome) to be “acceptable” to the World community. Derived in Section 3, these two credibility (or incentive compatibility) constraints define a unique set of Nash equilibria—a restricted set of conditions under which the UN could credibly issue a brinkmanship threat of preemptive action to deter a terrorist organization from acquiring WMD.

If the UN adopts a brinkmanship strategy, then from Figure 1 terrorists will look ahead knowing that if they acquire WMD they face a probability, q , of preemption.

Expected payoffs for the terrorists depend partly on their choices to Comply or Defy, and partly on whether they are Hard or Soft (and only they know their own type).

Deriving a “Rollback Nash Equilibrium” solution of the game through backwards induction requires the UN to first focus on the Terrorist group’s payoffs for each possible outcome. These are given by the far column in Figure 1. After the unpleasantness of placing itself in the terrorist’s shoes and establishing the each terrorist type’s best response to a brinkmanship threat (“Comply” or “Defy”), then the UN/World must evaluate its own payoffs.

Suppose the UN issues a probabilistic brinkmanship threat of preemptive action against a terrorist group that acquires WMD. Then a Hard terrorist must decide whether to “Comply” (not acquire WMD) with its associated payoff of A_{21} , or to “Defy” and acquire WMD. (See the top half of Figure 1) If the Hard terrorist under a brinkmanship threat goes ahead and acquires WMD, then there is a probability, $0 \leq q \leq 1$, the UN will engage in preemptive action, and $(1-q)$ it will not. This leads to a payoff of $[A_{21}]$ if the Hard terrorist group Complies, and an expected payoff of $[qA_{22} + (1-q)A_{23}]$ if they Defy.

From the perspective of the Hard terrorist group, if the UN decides not to issue a brinkmanship threat, their payoff is A_{24} (and the corresponding threat to the world is given by A_{14}).⁶ Assuming the Hard terrorist type prefers to operate in the absence of a

⁶ Payoffs in the game must reflect the preferences (utility functions) of the decision makers involved to represent real world terrorism. The Nash equilibrium brinkmanship game solution derived in Section 4 mostly depends on ordinal not cardinal payoff values, and thus is largely independent of differences in magnitudes of payoffs across players. The magnitudes of the payoffs are not as important in determining the solution. Instead, the solution is sensitive to the order of a player’s payoffs across the various possible outcomes. Some simplifying assumptions regarding relative magnitudes serve to increase tractability. This initial set of assumptions is meant to reasonably represent the order and (where possible) comparative magnitudes of the preferences of the UN and the terrorists. Section 5 explores the sensitivity of the solution to changes in the magnitudes and ordering of the payoffs, and in values of the probabilities q and p .

brinkmanship threat (**A24**), rather than to Defy the threat and “get lucky” and not suffer any preemptive actions (**A23**), the ordinal ranking of payoffs for the Hard terrorist type is **A24**≥**A23**>**A22**>**A21**.

The situation is analogous for a Soft terrorist type. Their associated payoffs appear in the bottom half of Figure 1. Given a brinkmanship threat, Soft can choose to “Comply” (not acquire WMD) with a payoff of **A25**, or to “Defy” and acquire WMD. If the Soft terrorist type acquires WMD, there is a probability, $0 \leq q \leq 1$, the World will engage in preemptive actions and $(1-q)$ it will not. This leaves the Soft terrorist with an expected payoff of $[q\mathbf{A26} + (1-q)\mathbf{A27}]$ if they Defy a brinkmanship threat. Of course, the best outcome is not to face a brinkmanship threat of preemption (**A28**).

What distinguishes the Soft terrorist type from the Hard terrorist type is a reversal in how they rank their two worst outcomes—facing down a brinkmanship threat of preemption. A Hard terrorist would rather Defy the brinkmanship threat of preemption (pursue WMD) and suffer the possibility (with probability $0 \leq q \leq 1$) of Preemption, than to Comply and agree to renounce WMD (i.e. **A22**>**A21**). The reverse is true for the Soft terrorist. A Soft terrorist would rather Comply under a Brinkmanship Threat (i.e. not pursue WMD), than to Defy the brinkmanship threat and suffer the possibility of Preemption (**A25**>**A26**). (See Figure 1)

Assuming the Soft terrorist prefers to operate in the absence of a UN threat (**A28**), rather than to Defy the threat and “get lucky” and not suffer any preemptive actions (**A27**), the ordinal ranking of payoffs for the Soft terrorist type is **A28**≥**A27**>**A25**>**A26**. Thus the ordinal preference ranking for the Hard terrorist type’s payoffs is

$$\mathbf{A24} > \mathbf{A23} > \mathbf{A22} > \mathbf{A21}, \quad (1)$$

whereas the Soft terrorist type's ordinal preference ranking is

$$A_{28} > A_{27} > A_{25} > A_{26}. \quad (2)$$

Conversely, from the perspective of the World community, the best outcome is for terrorists to Comply with a brinkmanship threat and renounce WMD ($A_{11}; A_{15}$). (see Figure 1) The next worse outcome for the World community is for terrorists to retain their option to acquire WMD in the absence of a brinkmanship threat ($A_{14} < A_{11}; A_{18} < A_{15}$), followed by terrorists acquiring WMD without Preemption ($A_{13} < A_{14} < A_{11}; A_{17} < A_{18} < A_{15}$). The worst outcome is for terrorists to acquire WMD with preemption (possibly inducing terrorists to use WMD). From best to worst, the World's ordinal preference ranking of outcomes against a Hard terrorist type is

$$A_{11} > A_{14} \geq A_{13} > A_{12}, \quad (3)$$

and similarly against a Soft terrorist type is

$$A_{15} > A_{18} \geq A_{17} > A_{16}. \quad (4)$$

III. Incentive Compatibility (“Credibility”) Constraints

The challenge in issuing a brinkmanship threat is that it must satisfy two “credibility constraints”—an “Effectiveness constraint” and an “Acceptability constraint.” The goal is to make the probabilistic brinkmanship threat of preemption sufficiently credible and unpleasant that it deters terrorists from obtaining WMD (*Effectiveness constraint*), but not so catastrophic and repugnant to those that need to carry it out that they would refuse to exercise that option (*Acceptability constraint*).

To be “effective” the threatened risk of preemption, $0 \leq q \leq 1$, must be high enough to at least deter the Soft terrorist from acquiring WMD, and yet still small enough (since preemption involves a mutually harmful outcome) to be “acceptable” to the World

community. These two incentive compatibility constraints define a set of Nash equilibrium solutions under which the UN could credibly issue a brinkmanship threat to deter terrorists from using WMD. Alternatively, it establishes the region in which a brinkmanship strategy is not a credible option, and other actions need to be considered.

A) The Effectiveness Constraint

In order for a probabilistic brinkmanship threat to be “Effective,” it must at least deter Soft Terrorists from acquiring WMD. In other words, the probability of preemption needs to be sufficiently high that it at least motivates Soft Terrorists to Comply. The first step is to derive the minimum threat probability, q^* , required to deter a Soft terrorist type. This is the lower bound of $q \in [0,1]$ for a threat to be effective. A brinkmanship threat of preemption with probability $1 \geq q \geq q^* > 0$ of being carried out will induce a Soft terrorist type to comply (and renounce attempts to acquire WMD) if and only if their expected value of Complying is greater than Defying, or iff:

$$E[\text{Soft Comply}] \geq E[\text{Soft Defy}] \Leftrightarrow A25 \geq qA26 + (1-q)A27 \quad (5)$$

From Figure 1, the rankings in (2), and equation (5), the set of “Effective” brinkmanship preemption threat probabilities that would deter a Soft terrorist type is given by:

$$1 \geq q \geq (A25 - A27) / (A26 - A27) = q^* > 0. \quad (6)$$

For each possible player type (Soft or Hard), the expected return from defying the brinkmanship threat rather than complying and not acquiring WMD, is jointly determined by the threat probability and the payoffs. The inequality in (6) represents the range of threat probabilities of preemption that would be effective in deterring a Soft terrorist type. The value q^* represents the lower bound of preemption probabilities for a brinkmanship threat to be effective.

Uncovering the “Rollback Nash Equilibrium” conditions for the UN to issue a probabilistic brinkmanship threat involves simple backwards induction. First the UN must consider the Soft Terrorist type’s decision whether to Comply or Defy faced with a brinkmanship threat given by (6). It is clear from Figure 1, the ordinal payoff rankings given by (2), and condition (6), that the expected value of Complying for Soft is greater than Defying. Conversely, given the ordinal payoff rankings (1), Hard will Defy since

$$\mathbf{A21} < q\mathbf{A22} + (1-q)\mathbf{A23} \Leftrightarrow E[\text{Hard Comply}] < E[\text{Hard Defy}] \quad . \quad (7)$$

(Note that, from (1) and (7), if $q = 1$, $\mathbf{A21} < \mathbf{A22}$ and if $q=0$, $\mathbf{A21} < \mathbf{A23}$ so that $E[\text{Hard Comply}] < E[\text{Hard Defy}]$ for all $\mathbf{1} \geq q \geq (\mathbf{A25}-\mathbf{A27})/(\mathbf{A26}-\mathbf{A27}) = q^* > \mathbf{0}$)

The result from the perspective of the UN/World Community is that, given these payoffs, any probabilistic brinkmanship threat sufficient to deter a Soft terrorist type, will not deter the Hard terrorist type. So in Figure 1, although a necessary condition for the UN to issue a brinkmanship threat of preemption is that the probability it is implemented is in the range $\mathbf{1} \geq q \geq q^* > \mathbf{0}$ (given by (6)), this is not sufficient. While Soft will Comply, Hard will Defy. So the decision that faces the UN of whether or not to issue a brinkmanship threat also rests on the probability the World is facing a Hard terrorist. Intuitively, a brinkmanship threat of preemption, with probability $\mathbf{1} \geq q \geq q^* > \mathbf{0}$ of being executed, will only be issued if there is a sufficiently low risk the World is facing a Hard terrorist type.

Given the best information available, the UN estimates it is facing a Hard terrorist type with probability, $p \in [0,1]$, and Soft with probability, $(1-p)$. The UN/World is now in a position to evaluate its expected payoffs if it issues a brinkmanship threat, $E_{UN}[\text{Threat}]$,

compared to its expected payoff if it does not, $E_{UN}[\text{No Threat}]$. A brinkmanship threat will only be issued as long as $E_{UN}[\text{Threat}] \geq E_{UN}[\text{No Threat}]$.

It was previously established that, for any brinkmanship threat probability in the range given by (6), Soft will Comply but Hard will Defy. Therefore, a UN brinkmanship threat is expected to result in World payoffs,

$$\begin{aligned} E_{UN}[\text{Threat}] &= p E_{UN}[\text{Hard Defy}] + (1-p) E_{UN}[\text{Soft Comply}] \\ &= p [qA12 + (1-q)A13] + (1-p) A15. \end{aligned} \tag{8}$$

Meanwhile, the No Threat option yields expected World payoffs of

$$E_{UN}[\text{No Threat}] = p A14 + (1-p) A18. \tag{9}$$

The condition under which the UN would issue a brinkmanship threat is if $E_{UN}[\text{Threat}] \geq E_{UN}[\text{No Threat}]$. To issue a credible threat, the expected value of issuing the threat, (8), must be greater than the expected value of not issuing the threat, (9).

The question is how high the UN can set a brinkmanship threat probability of preemption (q) that is still tolerable to the World community? Since issuing a brinkmanship threat involves a very real risk of mutual harm, the higher the probability the World is facing a Hard terrorist type, p , the lower the acceptable threat probability, q . Conversely, the higher the probability of a Soft terrorist, $(1-p)$, the higher the acceptable threat probability, q . This suggests a relationship between the probability of preemption, q , and the probability of playing against a Hard terrorist type, p , or $q=f(p)$. This relationship defines the second incentive compatibility constraint under which the UN could credibly use a brinkmanship strategy to deter terrorists from using WMD—the “Acceptability Constraint.”

B) The Acceptability Constraint

A brinkmanship threat is not credible if it is unacceptable to the party making the threat. The question is how high the brinkmanship threat probability of preemption with its possibly grim consequences can be set and still remain tolerable to the World community. Of course this depends to a large extent on the type of terrorist faced by the UN. Thus the “Acceptability constraint” involves deriving the relationship, $q=f(p)$, such that the expected value to the UN of issuing a brinkmanship threat of preemption, $E_{UN}[\text{Threat}]$, is greater than not to threaten brinkmanship, $E_{UN}[\text{No Threat}]$, or from (8) and (9),

$$p [qA_{12} + (1-q)A_{13}] + (1-p) A_{15} \geq pA_{14} + (1-p) A_{18} . \quad (10)$$

Assuming the cost to the World if a Hard terrorist type defies a brinkmanship threat and preemption does not occur, A_{13} , is the same as if no threat were issued, A_{14} (i.e. $A_{13} = A_{14}$), then expression (10) yields the Acceptability constraint,

$$q \leq f(p) = [(1-p)/p]z , \text{ where: } 1 > z = (A_{18} - A_{15}) / (A_{12} - A_{13}) > 0; \quad (11)$$

$$\text{and where } q=f(p): dq/dp < 0; d^2q/dp^2 > 0.$$

Combined with the Effectiveness constraint ((5) and (6)), the Acceptability constraint function given by (11) reveals conditions under which the UN can credibly issue a probabilistic brinkmanship threat of preemption—i.e. such that $E_{UN}[\text{Threat}] \geq E_{UN}[\text{No Threat}]$.

IV. Equilibrium Solution and Interpretation of the Results

If the UN is certain it is dealing with a Hard terrorist (i.e. $p=1$), then from (11), no probabilistic brinkmanship threat is acceptable (i.e. $q=0$). A brinkmanship threat in the face of a Hard terrorist is not a viable option. Figure 2 offers a graphical illustration and

interpretation of the results. In graphing the solution with the probability of a Hard terrorist group (p) on the abscissa (x-axis), and the brinkmanship threat probability of preemption (q) on the ordinate (y-axis), the point (1,0) defines the x-intercept of the Acceptability constraint.

To anchor the top limit of the Acceptability constraint requires deriving from (11) how low the probability of facing a Hard terrorist needs to be ($p=p^*$) in order for the UN to issue a dire threat (i.e. with certainty, $q=1$, that it will be carried out). Letting $q=1$ in (11) and solving for p yields,

$$p^* = 1/(1+z); \text{ where: } 1 > z = (A18 - A15)/(A12 - A13) > 0 \quad (12)$$

The point ($p^*, 1$) anchors the top limit of the Acceptability constraint which, from (11), slopes down at a decreasing rate as illustrated in Figure 2.

The Effectiveness constraint ((5) and (6)) requires a minimum threshold brinkmanship probability of preemption, $q^* = (A25-A27)/(A26-A27)$, to deter a Soft terrorist type. Substituting q^* into (11) yields the maximum associated probability of facing a Hard terrorist compatible with this threshold brinkmanship threat, or

$$p' = 1/(q^*+z) > p^*; \text{ where: } 1 > z = (A18 - A15)/(A12 - A13) > 0. \quad (13)$$

The Nash Equilibrium UN strategy set of brinkmanship threat probabilities of preemption, q , are reported here and interpreted graphically in Figure 2 in terms of the escalating probability, p , the World is facing a Hard terrorist type:

$$\text{For all } p \in [0, p^*], \quad q = 1 \quad (14)$$

$$\text{For all } p \in [p^*, p'], \quad q^* \leq q \leq f(p) = [(1-p)/p][(A18 - A15)/(A12 - A13)] \leq 1 \quad (15)$$

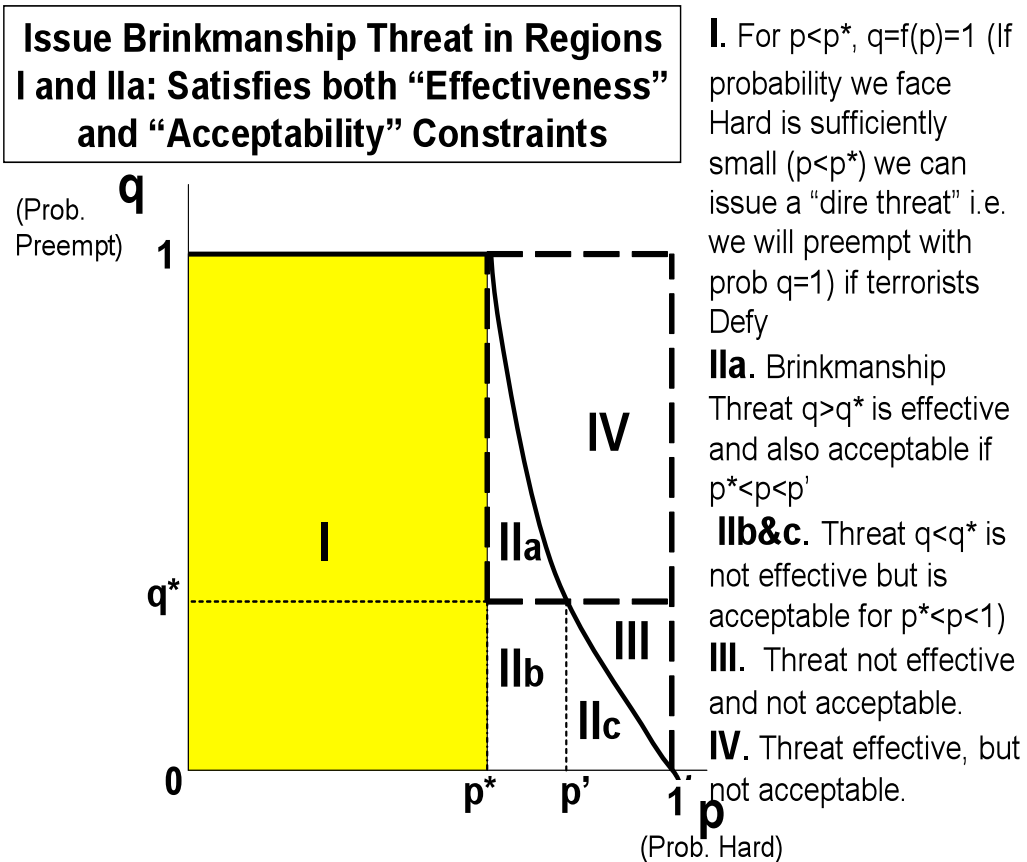
These results reveal conditions under which a brinkmanship threat of preemption is credible—i.e. a set of incentive compatible Nash equilibrium solutions that satisfy both

the “Effectiveness” and “Acceptability” constraints. The actual region in which the UN operates is driven by its assessment of the probability the World is facing a Hard terrorist type.

The Nash Equilibrium strategy set expressed in (14) and (15) is illustrated as Regions I and IIa in Figure 2. From (14), given a sufficiently low risk it faces a Hard terrorist type, $0 < p < p^*$, the UN can issue a “dire threat.” That is, it can promise preemption with certainty ($q=1$) if a terrorist group acquires WMD. This is illustrated as Region I in Figure 2. However, as the risk of facing a Hard terrorist grows beyond probability, p^* , the equilibrium threat strategy changes.

Figure 2

[Brinkmanship Game Solution Regions]



Given the increased risk, $p^* < p < p'$, of facing a Hard terrorist type, the UN can credibly issue a brinkmanship threat, but not a “dire threat,” since it would be too dangerous and not acceptable to the World community. Instead, the equilibrium brinkmanship probability of preemption is given by (15), and is inversely correlated with the probability of facing a Hard terrorist type. But, in order to be Effective, the probability of preemption cannot drop below q^* . This is illustrated as Region IIa in Figure 2. Regions I and IIa represent conditions under which the UN could credibly adopt a brinkmanship strategy of preemption to deter terrorists from acquiring WMD.

Brinkmanship threats would not be credible in Regions IIb and IIc in Figure 2, since while Acceptable, they would not be Effective. From (15), any risk of facing a Hard terrorist type in the range $p^* < p < p'$ that defines Region IIb, corresponds to a brinkmanship probability of preemption below the minimum effective level, q^* . Similarly, any risk of facing a Hard terrorist type in the range $p' < p < 1$ that defines Region IIc, corresponds to a brinkmanship probability of preemption below q^* . (See Figure 2)

A brinkmanship threat would also not be credible if the UN found itself in Regions III and IV of Figure 2. In Region III a brinkmanship threat is neither Acceptable nor Effective, while in Region IV, it is Effective, but not Acceptable.

In summary, Regions I and IIa represent the only two sets of conditions under which the UN could credibly issue a brinkmanship threat of preemption to deter terrorists from acquiring WMD. Outside these two regions, the UN must adopt other strategies to deter terrorist from acquiring WMD since a brinkmanship threat is either too weak to be effective, or unacceptably dangerous to the World community.

V. Conclusion

This paper examines conditions under which a world leader such as the Secretary General (or Security Council) of the United Nations (UN) might credibly issue a brinkmanship threat of preemptive action to deter sovereign states or transnational terrorist organizations from acquiring weapons of mass destruction (WMD). The model consists of two players, the United Nations (UN or World) “Principal,” and a representative terrorist organization “Agent.” The goal is to examine conditions under which the Principal might be able to structure a brinkmanship threat of preemption that aligns the interests of the agent with that of the principal—notably to Comply with a ban on the acquisition of WMD.

The UN (Principal) does not know which terrorist type it faces: Hard or Soft. Therefore, in choosing its brinkmanship strategy it must make a subjective assessment of the probability it is playing against a Hard terrorist type, say p , and the probability it is playing against a Soft terrorist type, $(1-p)$. In game theory these two states of nature are captured conceptually as Nature making the first move “choosing” whether the Principal is playing against a Hard or Soft terrorist organization. In the next stage the Principal (UN) must decide whether or not to issue a Brinkmanship threat. If the Principal issues a brinkmanship threat, then the Agent (Terrorist organization) must decide whether to “Comply” and refrain from using WMD, or to “Defy” and acquire WMD. A straightforward extensive form game tree is developed to illustrate the possible outcomes.

Except in the case of a “dire threat,” Brinkmanship always leaves something to chance. A Brinkmanship strategy always involves a probabilistic strategic action (preemptive retaliation in this case) that has a mutually harmful outcome. Preemption

might be bad for the terrorists, but it is also bad for those that carry out the threat. The goal is to make the brinkmanship threat sufficiently effective and unpleasant that it deters terrorists from using WMD, while not being so catastrophic and repugnant to those that must carry it out that they would refuse to do so.

Two incentive compatibility or “credibility constraints” are derived that must be satisfied for the UN to adopt a brinkmanship strategy. The first “Effectiveness Constraint” relates to the type of terrorist the World is up against (Hard or Soft), and the second “Acceptability Constraint” relates to risks the World is willing to take with a brinkmanship policy of preemption. The equilibrium solution and graphical interpretation of the results reveal when a brinkmanship threat of preemption is credible (a set of incentive compatible Nash equilibrium solutions), and when it is not.

Two sets of conditions are derived under which the UN could credibly issue a brinkmanship threat of preemption to deter terrorists from acquiring WMD. Outside these two regions, the UN must adopt other strategies to deter terrorist from acquiring WMD, since a brinkmanship threat would either be too weak to be effective, or unacceptably dangerous to the World community.

Hazy information and imprecise control can generate large risks in this model. Future research might examine the uncertainty that arises in identifying the probability the World faces a Hard terrorist type, p , and the uncertainty associated with the UN being able to stick to a brinkmanship probability of preemption, q . Since, in reality, p and q are uncertain, a Monte Carlo simulation might be constructed in which the actual probability that a terrorist group is Hard, and that the actual chosen probability of preemption can be assured, are drawn from probability distributions. For example, it is possible that using

Beta distributions as an illustration, and specifying cardinal payoffs, a Monte Carlo simulation could reveal how likely it is the Brinkmanship strategy would succeed under different sets of assumptions about probability distributions and payoffs.

Brinkmanship involves the strategic use of a probabilistic threat. This strategy involves exposing your rival and yourself to a gradually increasing risk of mutual harm that is not entirely within your control. The challenge is to generate a threat with a risk that is large enough to be effective and yet small enough to be acceptable. Future research could reveal the risk tolerance of each player by modeling a gradual escalation of the risk of mutual harm. In this case the UN would gradually escalate the brinkmanship risk of preemption to its equilibrium level, adjusting as it learns more about the probability it faces a Hard terrorist type from signals it receives from the terrorist organization. (For example, see Arce and Sandler 2007)

REFERENCES

- Arce, D. and T. Sandler 2007 "Terrorist Signaling and the Value of Intelligence," *British Journal of Political Science* 37.
- Blight, J. and D. Welch 1989 On the Brink: Americans and Soviets Reexamine the Cuban Missile Crisis (New York: Hill and Wang)
- Dixit, A. and S. Skeath 1998 Games of Strategy (New York: Norton)
- Friedman, J., 1986 Game Theory with Applications to Economics (New York: Oxford University Press)
- Fursenko, A. and T. Naftali, 1997 One Hell of a Gamble: The Secret History of the Cuban Missile Crisis (New York: Norton)
- Harsanyi, J., 1967 "Games with Incomplete Information Played by "Bayesian" Players, I-III. Part I. The Basic Model", *Management Science*, 14/3.
- Kagan, D. 1995 On the Origins of War and the Preservation of Peace (New York: Doubleday)
- May, E. and P. Zelikow eds., 1997 The Kennedy Tapes: Inside the White House During the Cuban Missile Crisis (Cambridge, Mass.: Harvard University Press)
- Melese, F. and D. Angelis 2004 "Deterring Terrorists from Using WMD: A Brinkmanship Strategy for the United Nations," *Defense & Security Analysis* 20.
- Rasmusen, E., 1995 Games and Information (Cambridge Mass.: Blackwell)
- Schelling, T., 2002 The Strategy and Conflict (Cambridge: Harvard University Press)