Brinkmanship and Nuclear Deterrence: The Neutrality of Escalation

Barry Nalebuff

Princeton University

1. Introduction*

The defense of the United States and the defense of Europe ultimately rely on a threat to use nuclear weapons if all else fails. This type of deterrence cannot be based on a cold rational calculation that nuclear retaliation for any attack is assured. Nuclear war is irrational; it would destroy that which we are trying to save. To threaten nuclear war, therefore, is equally irrational. Nuclear deterrence becomes credible only when there exists the possibility for any conventional conflict to escalate out of control. The threat is not a certainty but rather a probability of mutual destruction.

Nuclear deterrence involves a fundamental trade-off. There is a value in being able to make the threat of mutual destruction. The nuclear age has been forty years without world war. But, there is the cost of leaving our fate to chance. Nuclear deterrence requires accepting the risk of mutual destruction. Much of the current debate about nuclear deterrence centers on this risk. What can we do to lower the probability of nuclear war without losing the value of deterrence?

*I very much appreciate helpful comments from Larry Hilibrand, Mancur Olson, Ken Oye, John Riley and Andrei Shleifer.

This paper develops a stylized model of nuclear brinkmanship to study the effects of threats that leave something to chance. The formal model provides new insight into how changes in conventional military position and posture, changes in command and control, and changes in military technology affect the probability of inadvertent escalation to nuclear war. The main result is that the risk of inadvertent escalation is independent of position, posture, command, control, and technology; the danger of nuclear escalation cannot be reduced by changing the rules of conflict.

Section 2 provides an intuitive description of brinkmanship in nuclear deterrence. The formal model is presented in section 3. The proofs and mathematical details for all propositions are contained in the appendix.

2. Decision Making at the Brink

We begin our exploration into the relationship between nuclear deterrence and brinkmanship by reviewing the early models of deterrence from Ellsberg (1960) and Russett (1963). In these models, deterrence is based on a single static calculation. Credible deterrence requires that the defender believes "the prospective gains from a successful policy of firmness must be greater, when weighted by the probability of success and discounted by the cost and probability of war, than the losses from retreat. Formally, the defender will pursue a firm policy only if, in his calculation: $V_f \cdot s + V_w \cdot (1\text{-}s) > V_r$ where V_f is the value of successful firmness (deterrence without war), V_w is the value (usually negative) of the failure of firmness (war), V_r is the value (usually negative) of retreat, and s is the probability that firmness will be successful." [Russett (1963)].

This use of a cost-benefit calculation has stood the test of time. Yet, these calculations must be further developed in order to incorporate realistic dynamics. Where does the probability of success, s, come from? Is a country's decision to act firmly made only once and irrevocably?

Credible deterrence is a dynamic problem. In the face of a conflict, a country must maintain its firm position until either the other side backs down or until the conflict escalates into war. If the danger of war becomes too great, then it becomes in both sides' interests to back down and seek compromise. Firmness is not an all or nothing decision: it is a matter of degree. How long does a country maintain a firm position?

A problem with the static model is that the chance of success, s, will be continually changing as the conflict develops. If war appears immiment, then s will be near zero and neither side will wish to remain firm. This is especially true in the case of nuclear deterrence. In any nuclear confrontation both sides must expect to lose; no country could rationally decide to remain firm as the probability of escalation to nuclear weaponry approaches certainty.

The dynamics of brinkmanship tell a different story. The deterrence threat of nuclear weaponry is that they will be used *inadvertently*. As a conflict escalates, the probability of a chain of events leading to a nuclear confrontation increases. Eventually, the probability of war will be sufficiently high that one side will want to back down. But, the wheels of war set in motion have a momentum and the concession may come too late. All parties to the conflict may be desperately trying to prevent themselves from using nuclear weaponry. However, unanticipated, inadvertent, perhaps accidental or irrational actions beyond the leaders' control will provide the path of escalation to nuclear weapons.

> Escalation has generally been conceived of as either a rational policy choice, in which the leadership decides to preempt or to escalate in the face of conventional defeat, or as an accident, the result of a mechanical failure, unauthorized use, or insanity. But escalation arising out of the normal conduct of intense conventional conflict falls between these two categories: it is neither a purposeful act of policy nor an accident. What might be called 'inadvertent escalation' is rather the *unintended* consequence of a decision to fight a *conventional* war.
>
> —Barry Posen (1982)

Posen persuasively argues the difficulty in preventing nuclear escalation in the event of a conventional war between the superpowers. He describes three possible ladders that may lead to the inadvertent use of nuclear weapons: escalation may occur because of the inherent difficulty in planning for and controlling unforeseen contingencies; it may occur because of a misreading where defensive actions are seen as offensive; it may occur because of a breakdown in communications, command, and control arising in the fog of war. In addition to these inadvertent possibilities, there is also the possibility of a mistake. In peacetime, many levels of safeguards prevent the launching of nuclear weapons. But, in the atmosphere of edginess and distrust that goes hand-in-hand with war, defense forces will be on alert and many safeguards will be removed. As a result, there is a much greater danger of a mistake, mechanical failure, unauthorized use, or insanity.

The common denominator between all these routes of escalation is that they are unintentional. As Schelling (1960) first emphasized, nuclear deterrence is based on a "threat that leaves something to chance." It is this possibility of probabilistic escalation that allows the static models of deterrence to be translated into a dynamic story.

In a conventional war between the superpowers, as long as each side holds firm there is a risk of escalation. The risk is both from accidents and from the unintended consequences of fighting a conventional war. This risk continues until the conflict is resolved: either one side backs down or the probability of unintentional nuclear war turns into a reality. This nuclear game of "chicken" is called brinkmanship.

> Brinkmanship is the deliberate creation of a recognizable risk, a risk that one does not completely control. It is the tactic of deliberately letting the situation get somewhat out of hand, just because its being out of hand may be intolerable to the other party and force his accommodation. It means intimidating an adversary and exposing him to a shared risk, or deterring him by showing that if he makes a contrary move he may disturb us so that we slip over the brink whether we want to or not, carrying him with us.
>
> —Thomas Schelling (1960) in *The Strategy of Conflict*

Schelling's interpretation of brinkmanship can be translated into a formal mathematical model. The advantage of examining brinkmanship under the modeller's microscope is that it casts new insight into the debate over the risks of inadvertent escalation. On one side, former U.S. Navy Secretary Lehman argues that an aggressive U.S. naval position is needed to deter potential Soviet aggression against Norway.[1] On the other side, Barry Posen responds that the offensive strategy Lehman espouses would risk igniting a nuclear war. Whether by design or by accident, the US Navy's conventional war plan would threaten and possibly destroy Soviety nuclear missile submarines, as they are indistinguishable from conventional attack submarines. This might be read by the Soviet Union as the opening gambit of a nuclear attack.[2]

Each of these viewpoints focuses on only one of the two countervailing forces that arise when the United States takes a more aggressive military position. On the positive side, the greater risk of escalation that goes along with a more aggressive posture means that the Soviets will act less aggressively; the chance of conventional war is diminished. On the negative side, if deterrence fails and a conventional war breaks out, then the more aggressive posture increases the likelihood of inadvertent escalation. To calculate the *net* effect, we must multiply the chance of conventional conflict times the probability of nuclear escalation conditional on a conventional conflict. *A priori* there is no way to judge which effect will be more important.

The main result of the mathematical model below is that these two effects *exactly* cancel: the lower chance of conventional war is just offset by the increase in the conditional probability of escalation in the event of a conventional war. For this reason, changes in military policy and posture are impotent at reducing the risk of inadvertent nuclear war. The risk of inadvertent nuclear war is independent of the rules of the game.

This neutrality result for escalation has a predecessor in the field of finance.[3] The Modigliani-Miller theorem (1958) shows that stock valuation should be independent of the firm's ratio of debt to equity. The reason is that shareholders can readjust their own portfolio to undo any changes in the way a firm is levered. In a very similar fashion, countries too can adjust their portfolio of military actions to neutralize the effect of any changes in the leverage of escalation. Countries care not about words nor even deeds but rather the consequences of their deeds. The threat in nuclear brinkmanship is the risk of inadvertent escalation. When military position, posture, or technology changes these risks, countries can adjust their strategy so that the consequences of their actions remain unchanged. This intuitive explanation for the neutrality of escalation is formalized in the next section.

3. Mathematical Model

Consider the following highly stylized model of conflict between the superpowers. A conflict begins if one side challenges the status quo. To resolve the conflict peacefully, one side must make some concession to the other. Either the aggressor must back down or the challenged country must give in. The country that makes the concession pays a cost, C; the other country then receives a benefit from winning, B. If neither side offers a concession, the conflict escalates. The risk of escalation is that it may lead to inadvertent nuclear war, the costs of which are summarized by W.[4] The level of escalation is measured by a continuous variable, t. If the conflict reaches level t, then the chance of unintentional nuclear war is given by a damage function D(t).[5]

At the start of the conflict, an immediate concession eliminates the danger of war, $D(O)=0$. If the conflict is allowed to continually escalate unresolved, the probability of war becomes a certainty, $D(\infty)=1$.

Initially, neither of the two superpowers is certain about the other's exact costs and benefits (C_i, B_i, W_i). As the conflict escalates, both sides will revise their beliefs about the other's payoffs. The United States' strategy depends on its belief over the distribution of the Soviet's payoffs, denoted by C_2, B_2, and W_2. The Soviet Union's strategy depends on its belief over the distribution of U.S. payoffs, denoted by C_1, B_1, and W_1.

The relevant strategic variable is the probability distribution of how long the other side will wait before offering a concession. Country j's belief that the probability country i will offer a concession by time t is $H_i(t)$. In any fulfilled (rational) expectations equilibrium, country j's expectations will be correct.

What is the value to country i of letting the escalation rise until level t before offering a concession? There is a chance $H_j(t)$ that the other side will have given in already. If by level t, no concessions have been made, the momentum leading up to a confrontation has built up to the point where making a concession (at cost C_i) can stop the threats from being carried out only with probability 1-D(t). Thus, expected utility is

$$(1) \qquad E[U_i(t)] = \{-C_i[1-D(t)] - W_iD(t)\}[1-H_j(t)] + \int_0^t \{B_i[1-D(\tau)] - W_iD(\tau)\}H'_j(\tau)d\tau.$$

The choice of t that maximizes E[U] remains the same for any monotonic transformation of the utility function. We use this fact to observe that equation (1) can be rewritten so that the optimal t depends on B_i, C_i, and W_i only through a cost/benefit ratio.

Proposition 1: The optimal choice of t_i depends on B_i, C_i, and W_i only through

$$x_i \equiv (B_i + C_i)/(B_i + W_i).$$

The gain from offering a concession is W_i-C_i; the country saves the cost of war but still suffers the cost of defeat. The gain to holding out and waiting for the other side to offer a concession is $B_i + W_i$; the country saves the cost of war and reaps the benefit of victory. The ratio of these two payoffs, $(1\text{-}x_i)$, is the relevant strategic variable for both countries.[6]

All of the relevant information about the two sides engaged in brinkmanship is captured in the distribution functions $G_i(x_i)$ and $G_j(x_j)$. Neither side knows the exact value of the other's cost/benefit ratio but both are informed as to the distributions. We restrict attention to distributions of x_i that are supported on [0,1]. A country with $x_i > 1$ will never offer a concession and act identically to a country with $x_i = 1$. At the other extreme, if $x_i < 0$, there is a rush to offer a concession and the country acts as if $x_i = 0$.

Proposition 2: The escalation level is a monotonically increasing function of the cost/benefit ration x_i.

The implication of Proposition 2 is that there exists some monotonic function $t_i(x_i)$ which yields the optimal level of escalation for a country with cost/benefit factor x_i.[7] Let $x_i(t)$, $x_j(t)$ be the inverse functions for $t_i(x_i)$ and $t_j(x_j)$. The advantage of this reformulation is that we may replace the probability of conceding first, $H_j(t)$, by the probability distribution $G_j(x_j(t))$: by construction $G_j(x_j(t)) = H_j(t)$. Since escalation levels are monotonic with respect to x_i, the chance that country i will make a concession by level t is equal to the probability that its cost/benefit ratio is lower than $x_i(t)$, the cost/benefit ratio which concedes exactly at time t.

Proposition 3: For any monotonic differentiable damage function D(t) with range 0 to 1, the equilibrium level of damage is *independent* of the functional form of D(t).

The solution for $D(t_i(x_i)$ is $D(t_i^*(x_i)) = 1 - e^{-\int_{x_i(0)}^{x_i} \frac{\chi G_j'(x_j^*(\chi))\, x_j^{*\prime}(\chi)}{(1-\chi)(1-G_j(x_j^*(\chi)))} d\chi}$

which is cleary independent of D(t) since it only depends on G_1 and G_2.

The intuition behind this surprising result is that countries adjust their level of escalation to offset any changes in the damage function. At some level, the functional form of the damage function is a semantic question. Neither side is interested in the level of escalation per se. They are both concerned with the implications of reaching a certain level of escalation. If the damage function changes so that escalating the conflict has a higher probability of war, then both sides will choose lower levels of escalation in their equilibrium strategies. The ability to offset any change in the functional form of D(t) of course requires that the new damage function have the same range as the old.

This simple observation has direct policy implications. Consider the debate over equipping European battlefield commanders with tactical nuclear weapons. The presence of nuclear weapons on the battlefields of Europe is thought to facilitate escalation of nuclear war. It is argued that moving the weapons off shore would reduce the probability of their use. Without the first step of tactical nuclear weapons to break the focal equilibrium of no use of nuclear weapons, it would be more difficult to rationally escalate to a nuclear war.

In the brinkmanship framework, proving tactical nuclear weapons to battlefield commanders should have no effect on equilibrium. It raises the damage function for any level of escalation. Hence, the Soviets and the Americans respond by reducing their willingness to go to the brink. In equilibrium, use is no more or less likely.[8]

This neutrality result does not imply that the probability of nuclear war is some type of immutable physical constant. There are three ways in which it is possible to change the equilibrium. First, the distribution of the cost/benefit ratios, x_i's, can be changed. A policy that makes low cost/benefit ratios more likely leads to higher expected utility for both sides.

Proposition 4: In a symmetric equilibrium, a first-order stochastic dominance shift in the distribution of x towards lower costs results in higher expected utility for both countries.

Along these lines, there are other ways in which the distribution of costs and benefits can be changed. For example, better information might lead to less dispersion in the distribution of beliefs. A decrease in the uncertainty about the opponent's payoffs will change the equilibrium probability of inadvertent war needed to maintain deterrence. But, the direction of this effect is waiting to be discovered.[9]

A third policy change which will affect the equilibrium is to change the range of possible damages. This is the direction taken by many of the anti-nuclear movements. By working to banish nuclear weapons, they hope to eliminate the paths that escalate to nuclear war. The benefit of this approach is that it eliminates the random game of nuclear brinkmanship. If no paths lead to the ultimate use of nuclear weapons, then any threat to employ nuclear weapons is vacuous. The cost of this approach is equally apparent. Nuclear deterrence no longer exists. The defense of Europe can not be based on any ultimate nuclear threat.

In conclusion, nuclear deterrence is based on creating a dangerous game that nobody should want to play. The mathematical model helps illustrate why much of the debate over changing the technology of escalation may be misdirected. The rules of brinkmanship care only about the ends not the means. It is not possible to make a safer world by simply putting in safeguards to prevent nuclear escalation. In the event of a superpower conflict, both sides will just act more recklessly. To achieve a safer world, it is necessary to change the incentives and not just the rules of the game. If the distribution of costs and benefits can be altered so that low concession costs (relative to the value of winning) are more likely, then the dangers of brinkmanship will be reduced for both sides. The effect of changing the distribution via improving information remains an important topic for future research.

Notes

[1] "At the outbreak of war, the Navy would move aggressively into the Norwegian Sea, first with submarines and then with several aircraft carriers. They would roll back the Soviet fleet and attack its home base stations, striking ports and any bastions within reach of the carriers' attach planes"—Lehman (New York Times, 1985).

[2] "To threaten Soviet nuclear missile submarines is to wage nuclear war. It is very escalatory."—Posen (New York Times, 1985).

[3] It is tempting to draw a further analogy with the neutrality or revenue equivalence results in auction theory [see Riley and Samuelson (1981) and Milgrom and Weber (1982)]. However, this analogy would be false. The revenue equivalence results depend on the equilibrium being symmetric. The recent work of Maskin and Riley (1986) demonstrates that changing the format of the auction may change the expected revenue when the bidders valuations are asymmetrically distributed. The brinkmanship problem allows for asymmetric distributions. This neutrality result, therefore, depends on the specific and limited type of transformation being considered.

[4] The variables C, B, and W correspond to Russett's V_r, V_f, and V_w respectively.

[5] The level of escalation may be interpreted as the alert level status. As the conflict level rises, both sides move to a higher level alert. At each succeedingly higher alert level, there are fewer and fewer safeguards against the use of nuclear weapons. The possibility of inadvertent nuclear war becomes increasingly likely.

[6] From *a priori* reasoning both sides' decision must only depend on a ratio. Otherwise, a change in units would affect the results. Yet there is no natural measurement to measure the spoils and tragedy of war.

[7] We restrict attention to solutions which are differentiable.

[8] The reader might be tempted to conclude from this argument that providing battlefield commanders with tactical nuclear weapons is a good idea. The probability of nuclear war is unchanged while the probability of conventional war has been reduced. This conclusion is unwarranted. The present model does not take into account the costs and benefits from a conventional war. Were these costs (and benefits) to be included, the neutrality result would apply to the expected damage of war, conventional and nuclear combined.

[9] These issues are particularly relevant in regard to the U.S. Navy's strategy in the Northern Atlantic. Once the U.S. attacks Soviet submarines, both sides' payoffs become less certain. The U.S. will be uncertain about whether the attack submarine carried nuclear weapons; the Soviets will be uncertain about whether the Americans know the true consequences of their actions.

References

Ellsberg, D. (1960), "The Crude Analysis of Strategic Choices," reprinted in *Approaches to Measurement in International Relations*, John Mueller (ed.) New York: Appleton-Century-Crofts (1969).

Keller, B. (1985), "The Navy's Brash Leader," *New York Times Magazine*, December 15, pp. 31-69.

Maskin, E. and J. Riley (1986), "Asymmetric Auctions," UCLA Working Paper #254, revised.

Maynard Smith, J. (1974), "The Theory of Games and the Evolution of Animal Conflict," *Journal of Theoretical Biology*, 47, pp. 209-48.

Milgrom, P.R. and R.J. Weber (1982), "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50, pp. 1089-122.

Modigliani, F. and M.H. Miller (1958), "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review*, 48 pp. 261-97.

Nalebuff, B. and J. Riley (1985), "Asymmetric Equilibria in the War of Attrition," *Journal of Theoretical Biology*, 113 pp. 517-27.

Posen, B.R. (1982), "Inadvertent Nuclear War?," *International Security*, 7 pp. 28-54.

Riley, J. and W.F. Samuelson (1981) "Optimal Auctions," *American Economic Review*, 71, pp. 381-93.

Russett, B.M. (1963) "The Calculus of Deterrence," *Journal of Conflict Resolution*, 7 pp. 97-109.

Schelling, T. (1960), *Strategy of Conflict*, Cambridge: Harvard University Press, Ma.

Appendix

This appendix provides the proofs for Propositions 1 through 4.

Proposition 1: The optimal choice of t_i depends on B_i, C_i, and W_i only through

$$x_i \equiv (B_i + C_i)/(B_i + W_i).$$

Proof: Equation (1) can be rewritten as follows,

$$E[U_i(t)] = -W_i + [B_i + W_i]\{(1-x_i)[1-D(t)][1-H_j(t)] + \int_0^t [1-D(\tau)]H'_j(\tau)d\tau\}. \tag{2}$$

In this representation of expected utility, W_i is an additive constant and $[B_i + W_i]$ is a multiplicative factor. The optimization problem is then equivalent to choosing t to maximizing $U_i^*(t)$, an affine transformation of U_i,

$$E[U_i^*(t)] = (1-x_i)[1-D(t)][1-H_j(t)] + \int_0^t [1-D(\tau)]H'_j(\tau)d\tau. \quad \| \tag{3}$$

Proposition 2: The escalation level is a monotonically increasing function of the cost/benefit ration x_i.

Proof: Monotonicity follows from the fact that $\partial^2 E[U_i^*]/\partial t \partial x$ is always positive. This implies that the gain from waiting ($\partial E[U_i]/\partial t$) is bigger for a country which has a higher x_i,

$$\partial^2 E[U_i^*]/\partial t \partial x = D'(t)][1-H_j(t)] + [1-D(x)]H'_j(x) > 0. \quad \| \tag{4}$$

Proposition 3: For any monotonic differentiable damage function D(t) with range 0 to 1, the equilibrium level of damage is *independent* of the functional form of D(t).
Proof: The proof is by construction. A differential equation that defines an equilibrium can be discovered by equating the first-order conditions defined below in equations (5) and (5') for countries 1 and 2. After a series of manipulations, this leads to the neutrality result in equation (11).

Let $x_1(t)$, $x_2(t)$ be the inverse functions for $t_1(x_1)$ and $t_2(x_2)$. The advantage of this reformulation is that we may replace the probability of conceding first, $H_i(t)$, by the probability distribution $G_i(x_i(t))$: by construction $G_i(x_i(t)) = H_i(t)$. Since escalation levels are monotonic with respect to x_i, the chance that country i will make a concession by level t is equal to the probability that its cost/benefit ratio is lower than $x_i(t)$, the cost/benefit ratio which would lead to concession exactly at time t. Substituting this reformulation into the expected utility functions yields

$$(1'') \quad E[U_i^*(t)] = (1-x_i)[1-D(t)][1-G_j(x_j(t))] + \int_0^t [1-D(x_j(\tau))]G'_j(x_j(\tau))x_j'(\tau)d\tau.$$ [10]

At time t, the first-order condition is satisfied for country 1 with $x = x_1(t)$,

$$(5) \qquad -(1-x_1(t))D'(t)[1-G_2(x_2(t))] - x_1(t)[1-D(t)]G_2'(x_2(t))x_2'(t) = 0.$$

For country 2 at $x = x_2(t)$,

$$(5') \qquad -(1-x_2(t))D'(t)[1-G_1(x_1(t))] - x_2(t)[1-D(t)]G_1'(x_1(t))x_1'(t) = 0.$$ [11]

Take the ratio of equation (5) to (5 ') yields

$$(6) \qquad \frac{[1-x_1(t)]G_1'(x_1(t))x_1'(t)}{x_1(t)[1-G_1(x_1(t))]} = \frac{[1-x_2(t)]G_2'(x_2(t))x_2'(t)}{x_2(t)[1-G_2(x_2(t))]}.$$

Since $x_1(t)$ and $x_2(t)$ are both increasing functions, equation (6) implicity defines a first-order differential equation for x_2 as a function of x_1. Rearranging equation (6) leads to

$$(7) \qquad \frac{(1-x_2)G_2'(x_2)}{x_2(1-G_2(x_2))}\frac{dx_2}{dx_1} = \frac{(1-x_1)G_1'(x_1)}{x_1(1-G_1(x_1))}.$$

Define

$$H_i(x) \equiv \int_\beta^x \frac{(1-y)G_i'(y)}{y[1-G_i(y)]}dy$$

and consequently,

$$H_i'(x) = \frac{(1-x)G_i'(x)}{x[1-G_i(x)]}.$$

We can now rewrite equation (7) as

$$(7') \qquad \frac{d}{dx_2}H_2'(x_2)\frac{dx_2}{dx_1} = \frac{d}{dx_1}H_1(x_1).$$

If this is integrated,

$$H_2(x_2) = H_1(x_1) + k, \tag{8}$$

where k is a constant of integration. Since both H_1 and H_2 are strictly increasing functions, x_2 can be defined as an increasing function of x_1,

$$x_2 = H_2^{-1}(H_1(x_1) + k). \tag{9}$$

Equation (9) gives the solution for $x_2(x_1)$ as a function of the cumulative densities, $G_1(x_1)$ and $G_2(x_2)$, and the initial conditions, $x_1(0)$ and $x_2(0)$. The damage function, D(t), does *not* enter into the calculation of how x_2 is a function of x_1 and vice versa. Writing x_2 as a function of x_1 and rearranging the first-order conditions, we have in equilibrium

$$\left.\frac{d\log[1 - D(t)]}{dt}\right|_{t = t_1(x_1)} = \frac{x_1\, G_2'(x_2(x_1))\, x_2'(x_1)}{(1 - x_1)\,(1 - G_2(x_2(x_1)))}\,. \tag{10}$$

To solve for $D(t_1(x_1))$, we must integrate equation (10). It is simple to calculate the boundary value. At $x_i = 0$, the cost of concession is zero. Hence, concession will be made immediately so as to avoid any probability of the threat being carried out, $t_i(0) = 0$ and $D(t_i(0)) = 0$. Integrating forward shows

$$D(t_i^*(x_i)) = 1 - e^{-\int_{x_i(0)}^{x_i} \frac{\chi G_j'(x_j^*(\chi))\, x_j^{*\prime}(\chi)}{(1-\chi)(1 - G_j(x_j^*(\chi)))} d\chi} \tag{11}$$

The reservation level damage for country i can be written as a function of its cost-benefit ratio, x_i, and the other country's response function, $x_j(x_i)$. ||

Proposition 4: In a symmetric equilibrium, a first-order stochastic dominance shift in the distribution of x towards lower costs results in higher expected utility for both countries.

Proof: We are looking at the symmetric solution to a symmetric problem and hence we drop the subscripts. A first-order stochastic dominance shift in the distribution towards smaller costs is represented by an increase in ζ such that $dG_i(x,\zeta)/d\zeta \geq 0$ for all x. For ease of notation, the ζ argument is suppressed until it is needed in equation (14).

$$\begin{aligned} E[U^*(x(t), t)] &= (1-x)[1-D(t)][1-G(x(t))] + \int_0^t [1-D(x(\tau))]G'(x(\tau))d\tau \\ &= U^*(x(0), 0) + \int_0^t \left\{ \frac{\partial U(x(t), \tau)}{\partial x(\tau)} x'(\tau) + \frac{\partial U^*(x(\tau), \tau)}{\partial \tau} \right\} d\tau \\ &= 1 - \int_0^t [1-D(\tau)][1-G(x(\tau))]x'(\tau)d\tau \\ &= 1 - \int_0^{x(t)} [1-D(t(y))][1-G(y)]dy. \end{aligned} \tag{12}$$

The simplification in the third step follows from the fact that t is always optimally chosen for the cost/benefit ratio x(t). The final step is a result of changing the variable of integration. Note that in any symmetric equilibrium, $x_1(0) = x_2(0) = 0$.

Similarly, the equilibrium level of damage function can be transformed using integration by parts:

$$\text{(13)} \qquad D(t(x)) = 1 - e^{-\int_0^{t(x)} \frac{x(\tau)G'(x(\tau))x'(\tau)}{(1-x(\tau))(1-G(x(\tau)))} d\tau}$$

$$= 1 - e^{-\int_0^{x} \frac{y\,G'(y)}{(1-y)(1-G(y))} dy}$$

$$= 1 - e^{-\int_0^{x} \frac{\ln[(1-G(y))/(1-G(x))]}{(1-y)^2} dy} .$$

A shift in the distribution towards smaller cost/benefit ratios, represented by an increase in ζ, raises expected utility as

$$\text{(14)} \qquad \frac{dEU^*(x(t), t)}{d\xi} = \int_0^x \{[1 - D(t(y))]G_\zeta(y, \xi) + [1 - G(y, \xi)]D_\zeta(t(y))\}dy$$

$$= \int_0^x [1 - D(t(y))] \left\{ \frac{G_\zeta(y, \xi)}{(1-y)} - \int_0^y \frac{[1 - G(y, \xi)]\, G_\xi(\omega, \xi)]}{(1-\omega)^2 [1 - G(\omega, \xi)]} d\omega \right\} dy$$

$$= \int_0^x G_\zeta(y, \xi) \left\{ \frac{[1 - D(t(y))]}{(1-y)} - \int_y^x \frac{[1 - D(t(\omega))]\,[1 - G(\omega, \xi)]}{(1-y)^2 (1 - G(y, \xi))} d\omega \right\} dy$$

$$\geq \int_0^x G_\zeta(y, \xi) \left\{ \frac{[1 - D(t(y))]}{(1-y)} - \int_y^x \frac{[1 - D(t(y))]\,[1 - G(y, \xi)]}{(1-y)^2 (1 - G(y, \xi))} d\omega \right\} dy$$

$$= \int_0^x G_\zeta(y, \xi) \left\{ \frac{[1 - D(t(y))]}{(1-y)^2} \right\} \{1 - x\} dy$$

$$\geq 0. \quad \|$$

[10] The decision to concede will not be made at time zero but rather at time t. Therefore, the expected utility should be conditioned on the fact that escalation has not yet occurred. The consequence of this updating is to divide everything by $[1 - D(\tau)][1 - G_j(x_j(\tau))]$ and to integrate from τ not 0: this does not affect the derivative nor any other calculations that follow.

[11] Note that in the equilibrium where the two countries use strategies $\{t_1(x_1), t_2(x_2)\}$, the second-order conditions for a maximum will be satisfied,

$$\text{sign}\ \{dE[U^*_i(t)]/dt = \text{sign}[x_i(t) - x_i].$$

If $t < t_i(x_i)$ then $x_i(t) < x_i$ and it follows that $dE[U^*_i(t)]/dt > 0$ as $x_i(t)$ is a monotonically increasing function of t; conversely, if $t > t_i(x_i)$ then $dE[U^*_i(t)]/dt < 0$.