A DECISION-THEORETIC ANALYSIS OF THE NUCLEAR PRE-EMPTION DECISION

Stephen Shenfield
CREES, University of Birmingham

April 1984

Introduction

The idea for this paper arose in the course of research into Soviet nuclear policy. Traditional Soviet military thought considered victory in nuclear war possible, and also emphasised the great advantage accruing in war to the side which succeeds in taking the enemy by surprise in a first strike. At the same time it was desired to avoid war altogether if possible. There was as a result at least a certain tendency to think in terms of launching a pre-emptive first strike in the event of the leadership coming to the conclusion that a NATO first strike was imminent.

Over time the Soviet elite have become more aware of the devastating likely consequences of nuclear war. Many commentators, including some military men, have questioned the possibility of victory or survival in nuclear war. Official statements by political leaders allow for some uncertainty on the matter (eg "the future of humanity may be at stake"). The decision has, however, been made to renounce first use of nuclear weapons, ceding the "advantage of surprise" to the enemy in the hope of averting escalation to nuclear war. This measure is regarded by traditionalists as "unilateral disarmament in the face of the enemy".

It seemed useful to make a decision-theoretic analysis of the dilemma whether or not to strike first when one has reason to fear in a crisis that an enemy first strike is imminent. The decision-maker is assumed to believe that there is some quite small but non-zero chance of surviving a nuclear war (probability up to 0·4) and that this chance is maximised by striking first (with striking first increasing the probability of survival by up to 0·2). A decision to strike first is nevertheless not pre-determined by these assumptions, because a measure of uncertainty is ascribed

to the decision-maker's expectation of imminent enemy attack. He knows that the alarm may be false (and that, even if the alarm is true, the enemy may still change his mind), so that it may be possible to avert nuclear war.

In a very simple model of this problem there are a number of variables the values of which influence the decision: the perceived probability that the expectation of enemy attack is erroneous (P_E), the perceived probability of surviving a nuclear war if one strikes first (P_1), the perceived probability of surviving a nuclear war if one strikes second (P_2), and the relative utilities of the three possible outcomes - averting nuclear war, surviving nuclear war, and not surviving nuclear war. We are interested in the relationships between these variables and the propensity of the decision-maker to pre-empt. Our purpose is to determine the ways in which it is most urgent to influence the perceptions of potential decision-makers in order to reduce their propensity to pre-empt, for our own position is $P_1 = P_2 = 0$, which rules out the possibility of pre-emption being "rational" under any circumstances.

The objection may be raised against this exercise that decision-makers under enormous pressure in a crisis will not act on the basis of cool decision-theoretic calculations. While this is doubtless true, decision theory remains the most obvious tool for exploring approximately how the judgements and values of decision-makers may affect their decisions. Such an analysis at a minimum provides a baseline for considering likely departures of decision-makers from decision-theoretic rationality. The results are not in conflict with commonsense.

Analysis

The credibility assigned to an alarm of imminent enemy attack ($1 - P_E$) depends on many factors - perceptions of enemy intentions and of the evolving situation, degree of confidence in the reliability of one's military intelligence etc. It is obvious that the greater the credibility of the alarm, the greater the propensity to pre-empt. Given the values of other variables, there will be a certain threshold level of the perceived probability of error below which the decision will be to pre-empt. This error probability threshold (P_E^{min}) is a convenient indicator of pre-emption propensity, measuring

what degree of doubt in the alarm will suffice to dissuade the decision-maker from launching a first strike.

The three possible outcomes of the decision, in order of desirability, are:

A no nuclear war - decision not to strike and alarm proves false;

B surviving nuclear war;

and C not surviving nuclear war.

If the decision is not to strike, all three outcomes are possible - A if the alarm proves false, B or C if it proves true. If the decision is to strike, outcomes B or C are possible. It is convenient to set the utility of B, U_B, at zero. The question then is: what is the ratio of the positive (or at least non-negative) utility of A, U_A, to the negative utility of C, U_C?

If the decision-maker is concerned only to maximise the probability of mere survival, he will in effect conflate outcomes A and B: $U_A=U_B=0$. The desirability of surviving is so great relative to not surviving that the difference between the desirabilities of the two modes of surviving (with and without nuclear war) shrinks by comparison to zero. This we shall refer to as Case 1 - "pure survival utilities".

At the opposite extreme one may take the view that "the survivors of nuclear war will envy the dead", and regard the difference between the desirabilities of surviving and not surviving a nuclear war as insignificant by comparison with the difference between the desirability of nuclear war (irrespective of the outcome) and that of no nuclear war. The aim is then to maximise the probability of peace: $U_B=U_C=0$. We do not need to analyse the case of "pure peace utilities" because we could be sure that a decision-maker holding them would never pre-empt. If all the relevant decision-makers held them, nuclear war could happen only by accident.

A more realistic intermediate case is that of "mixed survival-peace utilities" in which $U_A > 0$ and $U_C < 0$. As our Case 2 we shall take $U_A = -U_C$, with the desirability of surviving a nuclear war equidistant from the desirability of peace on one side and from the undesirability of not surviving a nuclear war on the other. The utilities of real decision-makers perhaps lie in the range between Case 1 and Case 2.

Omitting a little algebra, we obtain:

Case 1 $$P_E^{min} = 1 - \frac{1 - P_1}{1 - P_2} = \frac{P_1 - P_2}{1 - P_2}$$

Case 2 $$P_E^{min} = 1 - \frac{2 - P_1}{2 - P_2} = \frac{P_1 - P_2}{2 - P_2}$$

It will be more enlightening, however, to use instead of P_1 and P_2:

$$\Delta P = P_1 - P_2 \quad \text{and} \quad \bar{P} = \frac{P_1 + P_2}{2} .$$

ΔP represents the "survival probability differential" - that is, the perceived advantage in terms of probability of survival of striking first over striking second, while $\bar{P}$ is an "averaged" probability of surviving nuclear war, expressing the decision-maker's perceptions of nuclear war "as such" in abstraction from the question of who strikes first. We then obtain:

Case 1 $$P_E^{min} = \frac{2\Delta P}{2 - 2\bar{P} + \Delta P}$$

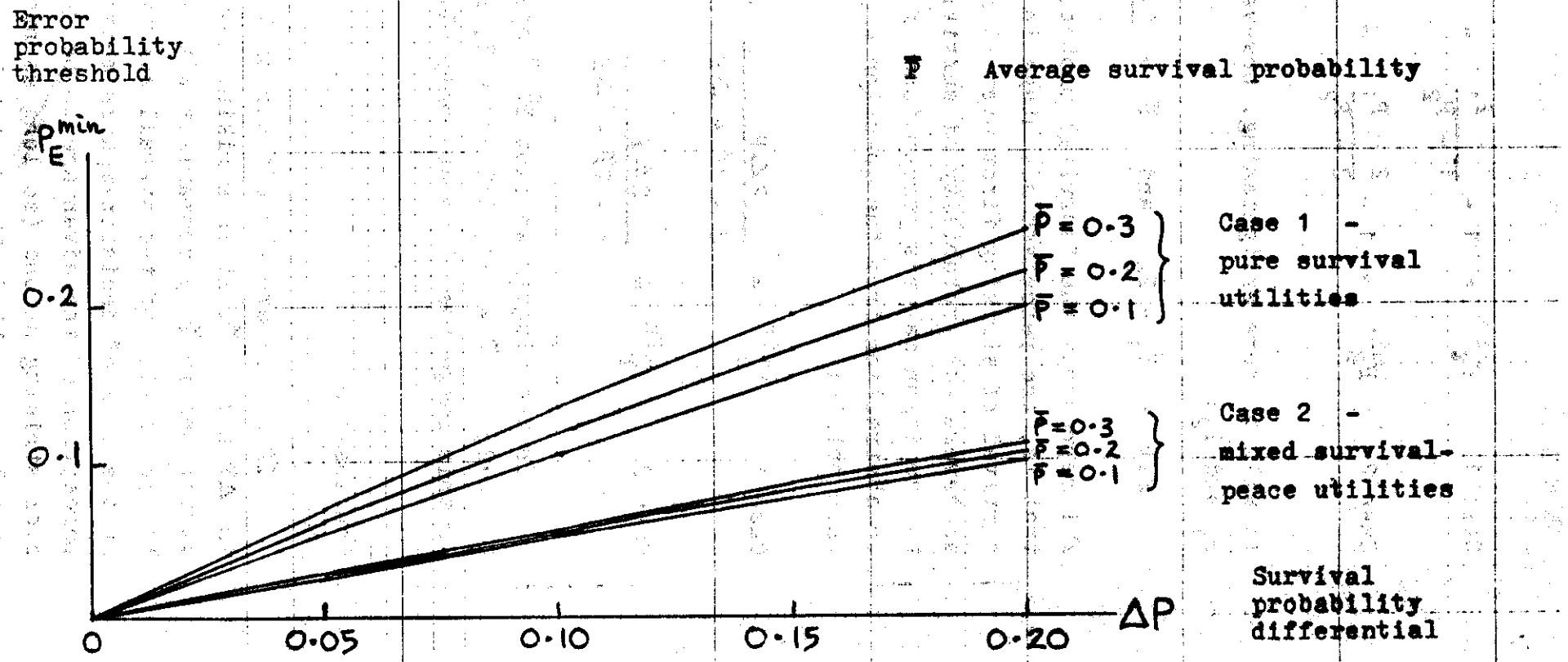
Case 2 $$P_E^{min} = \frac{2\Delta P}{4 - 2\bar{P} + \Delta P}$$

In the diagram we plot P_E^{min} against ΔP (in the range 0 - 0·2) for $\bar{P}$ = 0·1, 0·2 and 0·3, for Case 1 and Case 2. In this range the lines are only slightly non-linear.

Conclusions

If we compare the effects on P_E^{min} of equal increments in ΔP and in $\bar{P}$, it is striking how small is the impact of variations in $\bar{P}$ by comparison with that of variations in ΔP. The perception of the overall survivability of nuclear war in the eyes of decision-makers matters much less than their perception of the advantage to be gained by striking first over striking second (so long as nuclear war is regarded as at all survivable). The destabilising effect of the arms race towards perceived first-strike capability can be offset only by a very strong conviction of the impossibility of surviving nuclear war by striking first (as a result of the nuclear winter produced by one's own strike etc).

DIAGRAM - **DEPENDENCE OF ERROR PROBABILITY THRESHOLD (AS AN INDICATOR OF PRE-EMPTION PROPENSITY) ON SURVIVAL PROBABILITY DIFFERENTIAL, AVERAGE SURVIVAL PROBABILITY AND UTILITY STRUCTURE**



The structure of the decision-maker's utilities also has a very substantial impact on his propensity to pre-empt. It follows that the domination of the nuclear debate by the simple idea of ensuring some sort of human survival may have counter-productive effects by legitimising the "pure survival utilities" which (so long as nuclear war is perceived as at all survivable) raise the propensity of decision-makers to pre-empt. The risk of nuclear war might be somewhat reduced by broad acceptance of the attitude that, even if nuclear war were to be survivable, life would not be worth living in the post-war world. Therefore the call not for mere physical survival but for a life worthy of human beings may actually increase the chance of physical survival.