Volume Title: Issues in Defense Economics

Volume Author/Editor: Roland N. McKean, editor

Volume Publisher: NBER

Volume ISBN: 0-87014-490-1

Volume URL: http://www.nber.org/books/mcke67-1

Publication Date: 1967

Chapter Title: Strategy, Arms Control, and the Deployment of Defensive Weapons Systems

Chapter Author: Jerome Rothenberg

Chapter URL: http://www.nber.org/chapters/c5160

Chapter pages in book: (p. 65 - 104)

PART II

Strategy and Resource Allocation

Strategy, Arms Control, and the Deployment of Defensive Weapons Systems

JEROME ROTHENBERG

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

I. Introduction

The purpose of this paper is to examine the two-way relationship between the size and composition of military forces on the one hand, and appropriate military strategies for employing them (which includes not using them) on the other. The rationale of the whole work is that such a relationship does exist. Given the assumption of a certain degree of hostility or tension between two nations, one particular pair of force configurations will render certain strategies inappropriate for either or both parties, a different pair of configurations will make a different set of strategies inappropriate. Similarly, given certain strategies and initial force configurations, some decisions about future force configurations will be more appropriate than others. The terms "appropriate" and "inappropriate" are not meant to convey normative or even prudential weight. They denote agreement or lack of agreement with a system of certain implicitly assumed goals and a set of technical relations between alternatives means and ends. In other words, they refer to efficient and inefficient alternatives as presumably perceived by the participants.

The form of the investigation is to begin by characterizing the present nuclear confrontation between the U.S. and the Soviet Union. This provides a point of departure for considering a variety of hypothetical

NOTE: This study was supported by the Bendix Systems Division, the Bendix Corporation. Richard Cady of the Bendix Systems Division provided valuable assistance. But the author assumes full responsibility for the facts, analyses and opinions contained herein.

modifications in force configuration. For each modification, a set of likely consequences on nuclear strategy and further force structure is predicted. I consider two basic types of modification: first, some arms control measures; second, the nontrivial deployment of several types of defensive weapons systems not currently deployed in any significant measure. Among the arms control measures considered are some general proposals which have figured in public discussions—a nuclear weapons freeze, stage 1 of the American plan for General and Complete Disarmament, and the proposal suggested by Roswell Gilpatrick in a well-known article in *Foreign Affairs* in 1964. Other arms control measures illustrate certain abstract features which seem of outstanding importance to the attainment and viability of arms control agreements.

The bulk of this paper is devoted to the strategic implications of the deployment of defensive weapons systems on a significant scale. I discuss effects on the balance of nuclear deterrence, the stability of the resulting system of deterrence, and the possibility of superseding a deterrence system for maintaining peace by a system of mutual defense. In my treatment of defensive deployments I utilize the results of the previous analysis of arms control measures.

Through all the analysis I tacitly assume sets of national goals for the U.S. and the Soviet Union, which, while not necessarily aggressive, do result in conflicts between them and, as a result of consequent disagreeable historical experience, in a nontrivial level of mutual suspicion, distrust and tension. I do not assume that this tension level is constant or that it cannot be reduced by constructive human endeavor. But variations in this dimension are deemed to be outside the scope of the present paper, despite the recognition that arms races and military confrontations may play a crucial role in an over-all dynamic process of emotional-political-military interaction between nations.

II. Relationship of U.S. Arms Control and Disarmament (AC and D) Policies to U.S. Nuclear Strategies

A. CURRENT PROGRAMS

Under its current military programs the U.S. is developing a posture which is designed to be employed with a damage-limiting nuclear strategy. The U.S. has a multiple of (and considerably more total mega-

tonnage than) the Soviet Union's strategic missiles—more hardened, less concentrated, and more mobile. Thus, a strategy of controlled response, eschewing countervalue targets for counterforce targets in early exchanges of a central war, while holding back a secure, invulnerable portion of its strategic attack forces for bargaining purposes, has been both attractive and feasible for the U.S. Such a policy holds open a number of options in the actual fighting of a central war, the choice among which is dictated by the desire to limit damage to the U.S. and its allies, and to terminate the war rapidly while being able to negotiate from a superior position. When used first, retaliatory damage to the U.S. is decreased by its counterforce nature. When used as a second-strike thrust in retaliation for a counterforce attack by the enemy, it serves to reduce the potential population damage of a second, and countercity, attack by the enemy, by being directed against unexpended offensive forces still possessed by the enemy.

The present strategic posture of the U.S. vis-à-vis the Soviet Union is not so advantageous that a true first-strike option exists, where an initiating counterforce first strike could weaken the Soviet Union's retaliatory forces so that only "acceptable" damage could be done in retaliation. The Soviet Union's force level is high enough, and secure enough, so that great retaliatory damage would be produced in the U.S. Thus, the Soviet strategic posture is strong enough to deter an American surprise attack. Besides this, the "partial damage-limiting" quality of the American strategy presumes for its operational success that a Soviet first strike would be directed against U.S. strategic weapons and not populations. A Soviet massive countercity blow would, of course, abort any subsequent strategic dialogue wherein U.S. superiority could "win" the war. Since the relative invulnerability and plenitude of the forces reserved by the U.S. for countervalue targets (in comparison with the over-all size of the Soviet attack force) is deemed more than sufficient to deter a rational Soviet first strike of any kind—counterforce or countervalue—this type of eventuality is not to be expected. Yet the Soviet Union insists publicly that, whether as a first or second strike, they will hit American (and European) cities. They insist that no self-imposed, or enemy-induced, scale or targeting restraints can be counted on to prevail in a central nuclear war. Escalation would be inevitable—uncontrolled and total. Whether they truly believe this or not is not clear; but what is much clearer is that they have a rational incentive to

assert this publicly whether they believe it or not. Having a considerably smaller strategic force, specialized in very large weapons that possess far more countercity than counterforce effectiveness, and which are—and very likely will remain—somewhat more vulnerable than our own, the Soviets have apparently adopted a basic strategy of finite deterrence. They can threaten only countervalue retaliation, but retaliation heavy enough to make any first strike of ours unprofitable. They must therefore establish the credibility of this deterrent by insisting on its countervalue-targeting. To hint at its operational use for an end for which it is too small and ill-adapted, is to degrade its deterrent power.

The present over-all balance, then is that of mutual deterrence against a first strike, neither side having a first strike capability, but the U.S. possessing a more flexible posture in case of nuclear war—a variety of options of controlled response to influence the character of the sequence of strategic interplay during such a war. The availability of these multiple options during the course of a war, or indeed, the prewar expectation that such options will be available, is not, however, absolute. It is conditional on the Soviet Union's willingness to have its operational strategy influenced by American structuring of its rewards and punishments, or on the prewar American belief that the Soviet Union can be so influenced. But the apparent asymmetry in American and Russian military decision-making may be real, and American strategists had better be more interested in making sure that their strategic model is empirically relevant than that it is "rational." Thus, the margin of superiority of the American over the Russian strategic posture may turn out not to be a relevant one.

While a deliberate initiation of central nuclear war is mutually deterred, it is less clear what else is deterred. Some argue that the strategic balance acts as an umbrella which "makes the world safe for conventional war." Resort to central strategic exchanges being "rationally" excluded, neither side believes that the other would escalate to the level, nor therefore to a just lower level, that would trigger that final escalation, nor to . . . etc. Thus, limited conventional engagements can be relatively safely engaged in.

But this is by no means certain. The issue is the extent of responsible behavior, and how this extent is subsumed into "rational" behavior. If a deterrent posture fails, and the enemy initiates a war,

is it rational for the attacked side to carry out its prewar threat of massive *countercity* retaliation? Not especially; indeed, especially not if the enemy has retained further countervalue weapons which he threatens to use in response to retaliation. If the attacked side could be counted on dependably to back down "rationally" on its deterring threats after that deterrence fails, then deterrence scarcely exists at all. But deterrence does exist in such a prewar posture because the potential attacker cannot be sure that, when subjected to the crisis of a surprise attack, the victim will be capable of guaranteeing responsible "rational" behavior. Psychological studies indicate the breakdown of efficient individual and group decision-making under extreme stress. Under attack, the victim may, despite his "best" interests, retaliate uncontrollably. So it is rational for the potential attacker to anticipate the possibility of "irrational" behavior on the part of his enemy. And indeed, since it makes deterrence credible, it is "rational" for the potential victim sometimes to behave, or seem to behave, or be expected to behave, "irrationally."

The same line of reasoning implies that deterrence of surprise attack may also establish some deterrence of conventional war because, while "final" escalation is apparently deterred, neither combatant can be sure how much of a conventional war can be safely engaged in without setting off an escalatory dynamic that will get out of hand. The strategic balance does suggest that conventional limited wars will not be substantially escalated with rapidity but only, if at all, haltingly. Thus, conventional wars are not entirely safe, but not clearly unsafe. The lower the initial scale of hostilities envisaged or the smaller are the stakes involved—i.e., the more limited the war—the greater the likelihood that the strategic balance will prevent it from getting out of control.

We may generalize. The balance of nuclear deterrence probably deters major provocations of all sorts by both sides, but neither has any effect on, and may even stimulate, minor provocations—the latter as long as the elimination of the threat of surprise attack does not itself substantially decrease international tensions. To the extent that the Cold War is largely reactive and based on mistrust of the other's intentions, mutual nuclear deterrence can promote a detente which will decrease *all* provocations. It provides a period during which underlying nonmilitary problems between the antagonists have a better chance to be resolved constructively.

B. FUTURE TRENDS OF CURRENT PROGRAMS

If the present trends of procurement, deployment and strategy are continued, there is some belief that the postures of the U.S. and Soviet Union will become less disparate with respect to power. The U.S. may be willing to see some erosion in its strategic superiority, since a first-strike capacity will henceforth appear unfeasible; but it will not allow any substantial erosion to occur. The Soviet Union, on the other hand, has probably given up on a purely quantitative challenge. It hopes to offset U.S. superiority through qualitative changes. This introduces the uncertainities of technological change, with their attendant potentialities for strategic destabilizing surprises. Responded to by R-and-D activity by the U.S., the prospects for a continuing, or even accelerating trend of qualitative arms race make prediction difficult. All in all, one might predict an attrition of some of the U.S. options for controlled counterforce responses. Both nations may find themselves, *on the average,* possessing finite deterrence but only finite deterrence—a balance fraught with instabilities.

C. THE STRATEGIC FREEZE PROPOSAL

I assume this proposal calls for a freezing of the stockpiles of U.S. and Soviet Union nuclear weapons—warheads and their delivery systems—at their present levels. This will perpetuate the present strategic balance which is characterized by greater U.S. flexibility, but with a difference. While numbers will remain constant, there may exist a possibility of improving present weapons. *New* weapons systems will be, however, forbidden testing or production. Because of the asymmetry between the postures of the U.S. and the Soviet Union in terms of strategy and weapon characteristics, it is sometimes argued that the improvement which may be permissible under the freeze (e.g., in CEP, reliability, SLBM range, penetration aids, salvo capability) would differentially favor the U.S. Thus, the freeze would not preclude the U.S. from continuing a damage-limiting strategy through controlled response. It would, if anything, enhance it. Since, if allowed to do so, the Soviet Union would presumably move in the direction of greater hardening and dispersion of its missiles, the improved U.S. advantage is not likely to approach a real counterforce capability.

The impact of these relative postures on various types of deterrence is likely to be substantially as indicated for current programs.

D. THE U.S. GENERAL AND COMPLETE DISARMAMENT PROPOSAL, STAGE 1

This proposal calls for a proportionate 30 per cent decrease in the level of all strategic weapons, including ABM systems. This would call for a larger absolute decrease in long-range missiles and aircraft for the U.S. than for the Soviet Union (in the same proportions as present U.S. superiority), and a larger absolute decrease in intermediate range missiles for the Soviet Union. The result is some diminution of Russia's overwhelming intermediate range missile threat against Europe.

Of the two, the first is probably more significant. The Soviet missile advantage against Europe is so great that the differential disarmament here is not likely to alter relative postures appreciably. She will retain Europe as a hostage against U.S. long-range superiority. On the other hand, the American long-range superiority, while great in absolute numbers, gives the U.S. the opportunity for only a partial damage-limiting strategy. The significance of even this has been questioned above. The differential disarmament here can serve to shave further the flexibility of option possessed by the U.S. While the U.S. probably still could hold to a controlled response policy, the combination of differential disarmament and Soviet actions to decrease the vulnerability of its strategic forces, if allowed, might decrease its attractiveness relative to countervalue strategy close to the vanishing point. Disarmament appreciably beyond 30 per cent would quite likely convert the U.S. to a deterrence strategy. But where the point of indifference occurs is of course very difficult to gauge and requires highly detailed information.

There is an offsetting factor to this prediction, however. Insofar as the asymmetry of postures means that permissible qualitative improvements favor the U.S., a damage-limiting controlled-response strategy will be more securely attractive.

E. THE GILPATRICK LONG VIEW: 1970

This is possibly the most difficult AC-and-D policy to place in perspective, because it is the least well spelled out. It represents a voluntary, tacitly agreed-upon, breathing spell in the arms race, taking the

opportunity to withdraw weapons systems which are soft or nearly dominated in the armament menu—i.e., weapons not yet obsolete, but which contribute unstable elements to the over-all confrontation, or whose effectiveness is clearly secondary even in a diversified weapons mix. In addition, much of the same self-restraint in procuring and deploying significantly new strategic offense and defense systems as under the freeze would be called upon. Over and above this, however, there would be a budget cut of 25 per cent, reflecting partly the phasing out of soft missiles, manned bombers and interceptors, and partly the decreased needs of new procurement as current programs are progressively fulfilled.

At first sight, the impact of this policy would seem to be somewhere between that of General and Complete Disarmament (GCD), Stage 1, and the Freeze. But there are complicating features: the tacitness of the agreement, the nonproportional force level cuts, the control over aggregate expenditures instead of individual items.

The tacitness of the agreement introduces an additional strategic dimension to policy and also an environmental parameter. If one side acts unilaterally upon a supposed detente, in the expectation that the other side will follow, is it to the advantage of the other side in fact to follow? Would such an initiating move by the U.S. be seized upon by the Soviet Union as an opportunity (since there is no explicit inspection machinery to be set up) to achieve clandestinely the long-sought catching-up in strategic strength without openly challenging the superior productive capacity of the U.S.? Or would the opportunity for genuine detente be deemed more fruitful? The answer to this question is important because it throws light on whether or not the environment under which limited disarmament would be accomplished showed a significant diminution of tension and mistrust. In an environment of significantly eased tensions, emphasis by the U.S. upon a damage limiting strategy based on counterforce strikes might seem more active (leading toward the possibility of a pre-emptive first strike)—and hence provocative—to the Russians than a purely deterrent strategy. Therefore, it is not impossible that to preserve the spirit of detente U.S. policy might shift partly—and at least overtly—toward deterrent emphasis. Any communicated change of emphasis would represent a real policy change, since an important payoff of damage-limiting strategy, the establishing of ground rules for restraint in actual nuclear war, re-

quires specific structuring of the enemy's prewar expectations as to how one intends to fight such a war. And such expectations are created through the prewar communication.

The nonproportional force cuts, together with an absence of restraint on research and development, make possible significant transformations in over-all posture to adapt to the lower force levels. Such transformed postures could possess greater strategic power than the initial states, and could shift the strategic balance nontrivially. Insofar as the nonproportional cuts represent the phasing out of less effective and more destabilizing (because soft) weapons systems, the over-all balance will be more stable despite changes in differential power (the Soviet Union probably can gain more from this aspect, both because it has fewer offensive weapons subject to net decrease and because its weapons can use more additional hardening than those of the U.S.). Insofar as additional restructuring of force composition occurs, fed by R and D within the implicit "letter" of the Gilpatrick policy, stability could be seriously diminished. Choice of U.S. strategy could well hinge on such changes.

Since disarmament targets are expressed in terms of over-all budgetary restrictions for reasons other than the differential elimination of certain weapons systems already discussed, this adds to the range of changes which can be achieved in relative strategic power between the two countries under the Gilpatrick approach. As such, it complicates predictions about availability and desirability of different strategies.

These three closely related complications may turn out to be unimportant. If so, the unmistakable bent of the Gilpatrick proposal is to enhance the stability of mutual deterrence by decreasing the invulnerability of retaliatory systems and bring the two opposing postures closer together, thereby leading strategic emphasis away from counterforce toward deterrence alone. The explicit changes in force composition called for support this. But the lack of explicit restrictions about some other composition changes, and the encouragement of improved command and control, reliability and flexibility, may pull the other way.

F. SYMMETRY AND ASYMMETRY IN AC-AND-D AGREEMENTS

One must distinguish between equality and symmetry in arms control agreements. Symmetry need not involve equality of effects, whether it takes the form of an equal number of each type of weapon to be eliminated by both sides, or an equal percentage. The very fact that it is ex-

pressible either in absolute or percentage terms suggests that it differs from equality of effect. Actually, the ambiguity is greater than this.

Equality of disarmament in absolute or relative terms may lead to inequality of effect because of the existence of substitutive and complementary relationships within the compositions of military forces. These relations operate among the various target-goals, among the various offensive systems, among defensive systems, and between offensive and defensive systems. What is an optimal relative composition of systems at one force level may be suboptimal at a different level. This effect of force level on optimal relative proportions is likely to be greater the greater is the diversification of systems within the force. Thus, where the two arms agreement participants begin with highly dissimilar, highly diversified force structures, equal absolute or relative changes may worsen the efficiency of internal force composition for one much more than for the other. In order to achieve equality of effect, some asymmetry would have to be provided for in the agreement—either explicitly, or by stipulating equal disarmament in terms of some aggregate index of force level (say, the value of the stock and/or military budget) and allowing each participant to vary its composition as it desires.

Since equality of effect has often been spoken of in terms of equal numbers of weapons, our argument can be rephrased to say that there is more than one definition of equality of effect. And this assertion can be extended further, because equality of effect is in fact a highly ambiguous concept. It may have any of the following meanings at least: (1) an equal absolute reduction for the participants in the number of specific weapons; (2) an equal absolute across-the-board reduction in all weapons; (3) an equal absolute reduction in an index of force size; (4) an equal absolute reduction in the *ceteris paribus* national security significance of each weapons reduction; (5) to (8) equal proportional changes in each of the above; (9) reductions that result in equal improvement in over-all resource utilization (considering military and nonmilitary uses of resources: the fulfillment of national security and other goals); (10) reductions that lead to equal levels of national security; (11) reductions that lead to the greatest total improvement in the national security of both participants.

Any of these might be held to characterize equality, although one could make a good case that the first eight (excluding 4 and 8) represent practical approximations to the underlying "utility" dimensions referred

to in the last three. My purpose is not, however, to anatomize the concept of equality but to indicate that only the first eight bear any relation to the concept of symmetry; and these refer to different conceptions of symmetry. Only the fully across-the-board reductions (2 and 6) represent both intra and interforce symmetry. This means that the pursuit of most of these criteria would entail agreements with certain asymmetric elements. I have already argued above that pursuit of the third or seventh criteria would entail compositional asymmetries for the participants, while aggregate force-size reductions were symmetrical. But pursuit of the more ultimate criteria (9 to 11) would entail asymmetries even in the over-all size of force reductions, the more so the greater is the disparity of initial over-all postures between the participants.

Thus, asymmetrical arms control proposals are not necessarily inconsistent with equality of effect. Thus, it is not inconceivable that such proposals might obtain acceptance by both (or more) participants. Consider, for example, the spirit of the Soviet GCD proposals. By calling for the elimination of foreign bases and Polaris submarines in an early stage, and by scaling reductions in arms to achieve an essentially equal final position, these proposals call for a greater sacrifice of strength from the U.S. than from the Soviet Union. Yet it is not so much the inequality in force reductions that makes the proposals unsatisfactory as the inadequate provisions for inspection and stability of postures during the course of disarmament. As long as an essential nuclear standoff is the alternative to disarmament, so that any power differential between the contestants is largely irrelevant, an ultimate situation of weapons—and possibly national security—parity is not obviously unsatisfactory to either side. Even the erstwhile stronger side loses little in relative security, yet gains a significantly decreased resource burden in the defense budget.

We may generalize this. The mutual advantage necessary to a successful arms control agreement requires an approximation to equality of effect which is typically expressed in terms of symmetrical force reduction. But asymmetrical reductions may also be mutually acceptable if: (1) whatever power differentials the initial situation possesses do not confer decisive advantage to either side, (2) the terminal situation stipulates an underlying parity of security and/or force structure, (3) substantial force reduction is offered. The last is both the bait that can offset the antipathy to asymmetrical force reductions, and the reason

why asymmetry is needed (since at very low force levels, small absolute differences would be significant and could be decisive).

Where such proposals are mutually acceptable, the scaling down of differences would seem to suggest that both sides are satisfied with a simple deterrence strategy. If adequately safeguarded for stability during the reduction process, only a second-strike countervalue posture would be cultivated.

A further possibility should be noted. Agreements providing for large reductions in force structures might distinguish between offensive and defensive weapons (in the spirit of some Soviet proposals), stipulating stringent limits on the former but much more lenient limits on the latter. Such agreements could radically change the basis of military posture from deterrence to defense. The loss to both sides of an "assured destruction" capability—the loss of human and economic hostages—is made up by the corresponding ability of each to limit damage to itself from a first attack by the other. I will discuss below some of the deeper strategic issues involved in, and the stability of, such postures. For the present, all I need indicate is that none of the strategies I have discussed so far, all based on deterrence, would be fully relevant to this situation.

III. The Deployment of Defensive Weapons Systems and the Structure of Deterrence

A. WITHOUT ARMS CONTROL

1. *Ceteris Paribus Situations.* In discussing the effect of defensive postures on the balance of deterrence without arms control we must distinguish among three types of situations. One is where one side has deployed nontrivial packages of active and passive defense systems and the other has not yet had time to respond: *ceteris paribus* situations. A second is where the noninitiator has had the opportunity to react to the initial deployment, but has chosen to do so without resort to a defensive package: asymmetric *mutatis mutandis* situations. The third is where the response involves defensive deployments as well: symmetric *mutatis mutandis* situations. I begin with the first.

Any first deployment of a defensive package imparts an immediate military advantage to the deployer, regardless of the composition of the package. But the events leading to deployment, and the precise nature

of the advantage depends on the characteristics of the defense package. The first refers to the lead time required for procurement and deployment, and the proportion of this time during which the initiator can achieve, or wishes to achieve, secrecy about his action. Given equal initial capabilities on the part of both parties, the higher the values for the two elements, the larger is the potential military advantage to be obtained by deploying first; but also, therefore, the greater is the incentive for the noninitiating party to try to offset this with additions to short lead-time offensive capacity or an actual preemptive strike. Indeed, a very high value combination of the two generates instability: neither party can afford to be second in such deployment. A party engages in deployment as soon as R-and-D progress promises successful deployment. There is no mutual deterrence in deployment. The absence of secrecy reverses this: the longer is the lead-time for procurement and deployment, and the more visible are its early stages, the greater is the likelihood that each side can wait to see whether the other has begun before deciding to respond in kind. Shorter lead-times combined with lack of secrecy require quicker responses, since the noninitiator may reason that systems which are expected to confer only brief advantage are likely to by planned for use in that brief period.

Secrecy during the deployment process is thus of real importance in determining the extent and duration of the advantage conferred by deployment. But secrecy has a different function to play after deployment. Its worth then depends on whether the defense deployment is intended more to enhance deterrence or to improve effectiveness in waging war. In the former context the deploying party *wants* the other to know that his weapons will have less impact than he could heretofore expect—and wants him to know this *before* he decides to use them. So secrecy here would be dysfunctional. In the latter context secrecy can succeed in tricking the enemy into *using* unexpectedly ineffective weapons. It can indeed trick him into *starting a war* which he might not have begun if he knew his offensive weapons were going to be blunted.

The characteristics of the defensive package deployed helps determine the nature and extent of military advantage conferred, partly because they determine lead-time and secrecy, and also for reasons to be noted below. Realistically, given something like present technology, most elements of a defensive package have long lead-times and are

grossly visible. This is most nearly so with respect to ABM for cities, and fallout- and blast-sheltering, especially the latter, for urban populations. ABM to protect offensive weapons—airfields and missile silos, for example—has similar lead-times but is susceptible of greater secrecy. ASW is probably amenable to the greatest secrecy. Moreover, the lead-time involved depends on the nature of the technological breakthroughs that make ASW a truly effective instrumentality. Bomber defense systems already exist in the posture of both the U.S. and Soviet Union at a level of effectiveness deemed adequate to meet the prospective bomber threat. So no important new deployment need be considered.

Thus, in the large, one should expect that first deployment of a defensive package will not give rise to long-lasting *ceteris paribus* advantages.

The specifics of deployment will affect what kind of advantage is possible. ASW, for example, can serve as an adjunct to ICBM in a counterforce first strike, whether preemptive or otherwise. ABM for hardpoint defense is a way of decreasing the vulnerability of retaliatory weapons and therefore enhances deterrence of a first strike. Population-sheltering can improve the stakes for making a counterforce first strike, since the total damage-limitation impact of the counterforce strike and of sheltering against a weakened retaliatory blow could reduce the expected damage from a first-strike policy to a tolerable level. Actually, sheltering alone (unless it is of the current politically unrealistic blast variety) is unlikely to have this much damage-limiting effectiveness against even moderate countercity attacks unless combined with ABM capacity. Its chief *solo* effectiveness is likely to come as fallout protection against counterforce attacks. The combination of fallout-sheltering of populations with city ABM is a natural one. Each enhances the effectiveness of the other. Indeed, city ABM without fallout-sheltering would produce heavy casualties because of the radiation by-products of its "successful" use, and the possibility of enemy bypass attacks to waft radioactivity into the unprotected city.

If combined, active and passive city defenses could enhance the attractiveness of a first strike. But it might also detract from the need for making a first strike. So long as the nation's retaliatory weapons are reasonably secure against a first strike, the ability to limit population damage as well in an enemy strike gives the nation an ability to wait out a crisis without having to resort to a preemptive strike.

The deterrence effects of such first deployment depend in good measure on the credibility of the threats that may be made under its umbrella. With respect to central nuclear war, the key threat is: "If you attack me, I'll attack you and impose intolerable casualties." It is *not*: "If you attack me, I'll attack you and impose greater casualties on you than you impose on me." The distinction is important. If side A attacks side B in a counterforce blow, side B must now decide what it should do. If it should retaliate against side A—instead of perhaps suing for peace—it would have to expect an additional, and this time countervalue, blow. Now the difference in expected casualties from such a retaliation before and after it has deployed a defense package may be large enough to convert intolerable into "tolerable" casualties. Country B would have the incentive with defense deployment but not without such deployment, to retaliate even after a counterforce first strike by country A. This retaliatory attack could always be a countervalue blow unless some alternative target complex promised more effective impact. So B could threaten in advance of any first counterforce strike by A that it would surely retaliate and impose damage upon A in the event of such a strike, because it would have rational grounds for retaliating even after such a strike.

If A's first strike is countervalue, however, B's subsequent choice is different. The deployed defensive package having already been employed to decrease damage in the first strike, the consequences of a retaliatory strike now would be less affected—if at all—by the presence of the defense package. Consequently, the presence or absence of defense no longer has as much chance of making the difference between a rationally justified or unjustified retaliation. Retaliation by B now has to be threatened in advance on grounds of nonrational loss of control under crisis—the basic case of deterrence without defense. And the deterrent effect of the defense package rests on A's belief that the countervalue damage that it can impose in its *first strike* will not be large enough to be decisive because of the defense package.

The defense package decreases the number of hostages to any strike. But its differential effect would seem to be greater where the initial aggressor's first strike would leave the other's defensive capability still essentially intact. Thus, it would appear to enhance deterrence of a central nuclear war primarily where the potential aggressor found initiation of such a war rational on counterforce but not on countervalue terms.

Given the nuclear postures of the United States and Soviet Union at present and in the foreseeable future, a *deliberate* first strike against population centers seems unreasonable indeed. Only an initiating counterforce strike—which at present does not look reasonable either—would seem at all likely to commend itself "rationally" to either party. So defense probably enhances deterrence of central nuclear war.

Is any other form of deterrence enhanced? In this asymmetric situation of one-sided defense additional deterrence *is* possible. Say that side A is not seen likely to initiate a central nuclear strike but simply a nonnuclear, but serious, provocation. Side B, possessing the defense package, may indeed be able to threaten nuclear retaliation on grounds that its enemy's lack of comparable defenses and its own ability to blunt an enemy subsequent retaliation would *rationally* warrant such a nuclear first strike as a winning strategy. Such a threat is not incredible. Consequently, a significant range of non-nuclear provocations—lesser armed conflicts and threats, and political aggressions—may be deterred in the *ceteris paribus* situation.

Let us examine how this notion applies to the actual situation. Short of a military provocation technically relevant to U.S. defense, the area where a military provocation by the Soviet Union would be most likely to evoke a U.S. threat of central nuclear retaliation is Europe. The situation of Europe is importantly asymmetrical for the U.S. and Soviet Union. The significance of Europe as ally or prize, as military asset or threat, differs for the two. Its differential location vis-à-vis the two has substantial strategic implications. The Soviet Union has internal land links with Europe; the U.S. requires external, far more distant, sea links. For the Soviet Union, home-based intermediate and even short-range weapons have a potential offensive function against Europe; they do not for the U.S.: only foreign-based intermediate weapons have a comparable function. Finally, the relative military postures of the two vis-à-vis Europe, nuclear and conventional, differ markedly. The Soviet Union can bring a far more overwhelming attack against Western Europe than the U.S. and Western Europe can offset or can bring against the Soviet Union.

One of the problems of the developing strategic nuclear standoff between the two countries is that if the Soviet Union chose to invade Western Europe on a large, conventional scale, NATO would not be able to defend itself with conventional forces alone and, given the con-

tingent use of Russian intermediate and shorter range nuclear weapons, not with tactical nuclear weapons either. Under these circumstances, only a U.S. threat to offer strategic nuclear retaliation might succeed. But because of strategic nuclear stalemate (parity), such a threat would not be especially credible to the Soviet, since even a counterforce first strike could not guarantee only "tolerable" U.S. destruction under a retaliatory countervalue nuclear strike by the Soviets. The increasing lack of credibility of this threat to the Soviet Union as they have increased their assured destruction capability toward "parity" has spread to the NATO allies. These feel less assured that the U.S. would in fact retaliate centrally. A noncentral tactical nuclear response is far less satisfactory to them, since it makes of Europe a nuclear battlefield. And the overwhelming Russian nuclear capability in this limited arena would cause so much destruction as to make a mockery of Europe being "defended". Thus, in both Soviet and NATO eyes the U.S. strategic deterrent of limited aggression in Europe is seriously degraded. One response to this is the French desire for a national nuclear capability whose retaliatory use has higher credibility, especially when seen as a trigger to provoke U.S.-Soviet Union nuclear exchanges and therefore, in some respects, to enhance U.S. deterrent power again.

In this context, if the U.S. had first deployment of a significant defense package, it might substantially reestablish an effective deterrent against Soviet provocation in Europe. It would again make credible that the U.S. could rationally retaliate even if not directly attacked. Thus, it could perform a most important function. But we must remember that the military advantage that confers this virtue is a short-lasting transitional one. After Soviet response has been allowed for, and subsequent dynamic interplay, the story is quite different. And since defense of Western Europe is a long-term goal of American policy, it *is* long-run deterrence that is wanted. As we shall see below, this may nullify the profitability of first deployment of a defense package, except insofar as U.S. long-run interests suggest that a U.S. first strike, made temporarily tolerable by the umbrella of asymmetrical defense, would be desirable. So, despite the fact that first deployment would augment deterrence, its only practical advantage under the circumstances might be to make a first-strike strategy rational. Moreover, if the Soviet Union duplicated this reasoning, they would see in a U.S. first deployment of defense, not a defensive, deterrent purpose, but an aggressive first-strike one.

They might be driven to desperate aggressive countermeasures, perhaps even a preemptive nuclear strike. Thus, if the advantages of first deployment are in reality very short lasting, this could impart important instability into the relative postures. This will be discussed below.

Deterrence against provocations elsewhere is not likely to be much enhanced by first deployment, even in the short-run. While expected damage from a retaliatory interplay might be decreased nontrivially, and to an "acceptable" level with respect to the goal of defending Europe, this damage is still enormous by any human standard, and is likely to be unacceptably high for any U.S. interest less important than Europe (and probably Latin America). Consequently, the defense deployment would not markedly increase the credibility of threats relating to lesser provocations. Indeed, it is not likely even to lead to the making of such threats.

One last point. I have been treating the credibility of threats as essentially unitary: whatever an opponent would believe (disbelieve) our allies would believe (disbelieve) and vice versa. This is probably a good approximation, but at times significant discrepancies may creep in. The general closeness of the approximation is due not only to the fact that both enemies and friends are reading the same objective evidence, but that each group reads the actions of the other as evidence on how well *they* believe certain threats. Thus, the Soviet Union may well read France's efforts to establish an independent nuclear deterrent as disbelief in the threats on which U.S. deterrence depends, and its own belief in the U.S. threats may be thereby affected (since an ally can be assumed to be privy to information not available to an enemy).

Discrepancies arise because the information available to allies and enemies *is* different. Moreover, the vantage is different, and this introduces a strategic variable. Allies have an incentive to *act as though* they believe a threat made by one of them to bolster its credibility for the common enemy. On the other hand, an ally may act in a way that appears to be relevant to belief of a threat, but in reality is part of an attempt to improve its bargaining position within the alliance. A motivation of this sort has sometimes been attributed to the French.

2. *Asymmetric Mutatis Mutandis Situations.* I can be quite brief, since most of the basic elements have been treated. The act of deploying a defensive package will be responded to insofar as the other party expects it to have military effectiveness *or* believes that the deployer expects it to have effectiveness. Given anything like present technology the

exchange rate greatly favors offense over defense. A reasonable response may well involve adapting and/or augmenting offensive capacity, with no attempt to institute a direct defensive capability. Offensive adaptation includes, for example, retrofitting penetration aids against ABM and submarine performance changes against ASW. Augmentation includes simply increasing the number of offensive weapons. Both measures are designed to offset directly (by saturation or performance or operational tactics) the defensive weaponry on its own terms and thus replace the operational effectiveness of the offense to a level comparable to the pre-defense deployment situation. Since the responding country would wish to achieve this replacement at a given level of confidence in the face of an increased uncertainty about actual effectiveness, such would require an adaptation augmentation objectively greater than what would be necessary to impose the predeployment level of damage.[1]

A response of this sort has every likelihood of being able in fact to offset the military advantage imparted by the defense deployment. Indeed, as suggested, in actual use it would probably more than offset such advantage. Since under anything like present technology the possibility of such offset seems generally appreciated, a known response in offensive capability would wipe out the defense deployer's felt advantage. It would cancel out any enhanced deterrence. On the other hand, for reasons to be discussed below and having to do with uncertainty levels, even a slightly more than exact offset through enhanced offense is not likely to establish increased deterrence for the responding country. The defensive package is likely to increase the threshold of disparity in military capabilities, which is necessary if the opposing side is to have *significant* advantage.

If country A deployed a defensive package, and country B responded with enhanced offensive capability, the interaction might continue in the form of an arms race of the same character—that is, a further defense

[1] A qualification is necessary. The enemy's defense deployment raises the cost of inflicting any given degree of damage on him. Resources available for national security are not free to the nation: they have as opportunity cost the nonmilitary resource uses foregone. Since national security and other goals are to some extent substitutive, a substantial increase in the cost of the former relative to the latter could well induce some diminution in the level of security sought. This means that the level of assured destruction sought (at given degree of confidence)—i.e., the extent of direct offset to defense deployment—could fall below that of the predeployment situation. In the present example, the cost change is too small to justify serious attention to this qualification.

buildup by A, triggering a further offensive buildup by B, and so on. But this is unlikely. The initial defense deployment would probably have been associated with R-and-D successes. The temporary advantage rendered might or might not have seemed worth the expense and resulting postures. But a new significant technological achievement is not likely to be so conveniently available just in time for A to respond with a new round. And in the absence of such an achievement, the fact that a simple extension of measures involved in the first deployment would have had lesser returns even before B's offsetting response, together with the demonstration that B could respond advantageously to offset A's defense (because of the current technological imbalance between offense and defense), would very likely deter A from initiating a new round as response. The dynamic interaction unit therefore seems to be one round at a time. Each round is a *new* episode, initiated in its own special terms.

3. *Symmetric Mutatis Mutandis Situations.* Here it is envisaged that country B would respond to A's deployment of a defense package by itself deploying a comparable defense package. In view of the advantage of offense over defense this kind of response has to be motivated. It is likely to come about where both countries have comparable defense technologies available, but where B has been hitherto deterred from deploying its own defense package by the desire to keep A from such deployment. Deterrence having failed with respect to A, it now fails for B as well.

By responding with a defense package, despite the advantages of offense over defense, B is not committing an irrelevancy. It will be recalled that one of the options for A which was enhanced by the defense deployment was a first strike. Insofar as this is so, defense deployment by B has a direct defensive function: it can save lives in the event of attack. But there is a deterrent function as well—exactly the same as for A. If already subject to attack (counterforce especially) by A, the possession of some population defense, for example, could make the difference for B between a rational incentive to retaliate and to surrender. Thus, its presence will tend to deter A from attack.

But the presence of a defense capability on both sides complicates deterrence for both. By decreasing the number of hostages subject to the enemy's attack, it decreases the amount of assured destruction which that enemy can count on in a retaliatory attack. Thus, it de-

creases the ability of each to deter the other's attack. Each has reason to believe that if it attacks the other, the other will probably retaliate; but if the first attack is counterforce and one's own defenses are not inconsiderable, the combination can reduce over-all damage to oneself to within tolerable limits relative to the possibility of victory. An initiatory counterforce blow bypasses the enemy's *defenses,* while one's own defenses function to absorb the evoked retaliation. The enemy's defenses ultimately enter consideration in that the initiator, having degraded the enemy's offensive capacity by his initial attack and the absorbed retaliation, can obtain victory only by finally threatening a countervalue blow. The impact of this threat depends on the efficacy of defense against the remaining offensive capability of the aggressor. If the willingness to terminate the war depends more on one's *relative defenselessness* vis-à-vis the enemy at some point in the hostilities, rather than on the relative casualties so far inflicted—a not unlikely circumstance—then the mutual possession of a defensive capability can symmetrically weaken deterrence. The defensive packages on both sides do not simply cancel one another out; they can operate almost independently to unhinge a balance of mutual deterrence (except for the effect of uncertainty, as will be discussed below).

Indeed, it is partly for this reason that the mutual postures we have just considered are incomplete. It is very unlikely that B will respond to A's defense capability simply by deploying its own defense. It will in addition attempt to offset the enemy's defense, as we saw in the last section. It will enhance its offensive capability. And the enemy will similarly respond to the defensive capability induced. Thus, if A deploys a defense package, B will probably do likewise and augment offense as well. A, in turn, will offset B's new defense by augmenting its offensive capability. Thus, a defensive and offensive arms race is the likely outcome.

The consequences of this balanced interplay differ from the preceding case. Offensive capability having increased something like proportionally for both sides, and in like form,[2] an initial counterforce blow by A cannot, given approximately the relevant exchange rates obtaining today,

[2] Like form is specified—and assumed—because if A augmented offense in terms of numbers and accuracy, while B solely in terms of penetration against ABM, a counterforce blow by A would be more effective after such proportional augmentation than before.

reduce expected retaliatory damage to itself much—if at all—below the level obtaining before any defense deployment, despite the presence of that defensive capability.

Thus, with additional offense offsetting defense, no degradation of deterrence occurs for either side—unless, of course, the arms buildup is disproportionate for the two (a different case). The result is similar to the status quo ante in relative strengths, and almost the same in terms of deterrence. A standoff has been purchased at a higher level of armaments—and at great expense—so that in case of war by accident a higher total level of destruction is possible. Moreover, the dynamics of such arms procurement can lead to instability, as will be noted below. The only useful thing which has been purchased is an increased uncertainty as to outcomes of actual operations. This enhances the stability of deterrence somewhat, by increasing the margin of superiority which either side must believe it has in order to be willing to initiate hostilities. But much the same effect is obtainable by increasing the absolute level of offensive capability alone. The higher is this level, the greater the proportion of superiority which one side must possess in order to be able to initiate a counterforce strike and reduce retaliatory damage to a preassigned level. Even the stability bought through uncertainty should not be exaggerated. As long as there are no constraints on offensive capabilities, the induced augmentation is likely to be great enough to prevent any very significant increase in operational uncertainty. It is primarily where offensive constraints exist that this may be an important factor. I shall return to this below.

In sum, where there are no agreed-upon constraints in force structure, deployment of defense packages can confer short-term advantages on the initiator—translated in part into primary and secondary deterrence—but these are largely or wholly offset when the other side responds. The most likely outcome is an expensive cancelling-out of advantage, worsened by an induced dynamic arms acquisition process which can destabilize whatever balance of deterrence existed before defense deployment began.

B. WITH ARMS CONTROL

The analysis differs if arms control measures are in force. If they are, defensive commitments come at the expense of offensive options, or at least cannot be freely offset by offensive augmentation. The *ceteris*

paribus case is the same. Only the *mutatis mutandis* case must be reconsidered.

The situation depends on what kind of arms control obtains. The two main types for this issue are: (1) limitations on offensive weapons systems but not on defensive systems (an approximation to a Soviet preference), (2) over-all force limitations which allow for substitutions between defense and offense. (Limitations specifying particular offensive *and* defensive force levels would not permit us to consider defensive force levels as a decision variable for the two participants.)

Under the first, it is not unlikely that a parallel defensive arms buildup will ensue since, offensive levels being fixed by agreement, every increment of defense for either side decreases expected damage from a war (accidental or not). The incremental reduction of expected damage for each is not undermined by the other's similar defensive additions. With respect to this factor, then, a defense buildup represents an atomistic, nonstrategic choice limited only by the nonmilitary opportunity cost of resources.

But a strategic factor does enter in as well. I noted earlier that defensive deployment may have either a defensive *or* an offensive strategic impact, or a combination of both. By decreasing its own expected damage, defensive deployment decreases the other side's basic ingredient for deterrence. Thus, the parallel defense buildup can be played for aggressive intent and become an arms race—as offensive in character as if offensive and not defensive weapons were being acquired. In such an eventuality the arms agreement itself would be endangered. Thus, both sides would have to realize that the pace and extent of their defensive buildups were constrained by their strategic valuation of the offensive weapons arms control agreement. It is likely, therefore, that a viable, purely offensive weapons-systems limitation would impose indirect limitations on defensive systems as well.[3]

The situation resembles the symmetric *mutatis mutandis* case without arms control except that: (1) offensive levels are prevented from rising

[3] This is not to say that such an initial agreement might not be amended to allow for an ultimately complete supersession of defensive posture for deterrence. But such supersession is likely to come about by explicit agreement rather than atomistically or even tacitly, because the pathway involves serious strategic instabilities. If progress toward the goal is made possible by an easing of international tensions, then this is more likely to be brought about by a decrease in allowable offensive force levels than by an increase in defensive levels. The former is far cheaper and probably less dangerous.

to offset the defensive deployment (2) the mutual defensive buildup can proceed much further here before being discouraged by arms control viability and nonmilitary opportunity costs. Together, these mean that the assured destruction arm of deterrence can be substantially degraded. Taken by itself, this could be seriously destabilizing, and war might be initiated before even the telltale warning of a breakdown in the arms control agreement into upward spiralling offensive buildups. But another factor assuages this possibility of instability. Deterrence does not simply depend on what an attacker will lose if he attacks his enemy, but also on what he will gain by making such an attack. It is a question of the *difference* in outcomes between attacking and not attacking. Despite his standing to lose little by an attack, the attractiveness of such a course will be weakened as the potential attacker stands to *gain* less and less. Parallel defensive buildups decrease the assured gains just as they decrease the expected losses. The net effect is not clear.

There are really two effects on the gains from attack: a decrease in the expected level of destruction to one's enemy, and an increase in the variance of possible outcomes. Both make a first strike strategy—and a first strike itself—less attractive. The effect of expected level is obvious. The effect of variance (uncertainty) is somewhat less obvious, but probably compelling none the less. For one thing, the simple inability to be sure of what will happen when the stakes are fantastically high and the knowledge that what is done cannot be undone and that all humanity would pay the cost of a miscalculation would seem to exert strong deterrence against central nuclear attack. For another, there are measures a potential attacker can take to offset some of the effects of this uncertainty—reconnaissance capabilities, flexibility of targeting, salvo capabilities, etc. But all of these are expensive and, under some arms control agreements of the type I am considering, are either limited in quantity or are obtainable only by sacrificing the absolute level of offensive capability. Really effective countering of the enhanced uncertainty—an offensive buildup—is prevented by the arms control agreement.

One may guess at the over-all consequences on deterrence. Both the expected gains and retaliatory losses from an initial attack are decreased. But attached to *both* is an increased variance of outcome. Not only may one's attack prove to be singularly ineffective but also one's defenses against retaliation. The increased uncertainty of both work against a potential attack: he will wish both a high probability that *at least* a

specified amount of damage be done to his enemy, and a high probability (possibly higher) that *no more than* a specified amount of damage be done to him in return. Both require overcompensating for variance —the larger the variance the greater the degree of overcompensation. Again, the critical constraints imposed by the arms control agreement prevent effective insurance. One could easily predict that deterrence against central nuclear war will be enhanced under these conditions: only extreme inadvertent exacerbation of political tensions would seem likely to make an attack worth considering.

Much the same would seem to characterize the second type of arms control—over-all force limitations with substitutability between offensive and defensive weapons systems. The main difference is that the opportunity cost of defensive capability is offensive capability. Given something like present technology, and force levels not very dissimilar from present U.S. and Soviet levels, not much defensive capability will be deliberately acquired at the cost of offensive capability. It is possible that more will be purchased, however, then in the no arms control situation, since the over-all force level constraint will prevent the opponent from easily augmenting his offensive to offset any defensive deployment. Thus, when some R-and-D achievement in defensive systems promises reasonable performance against the current offensive posture, a nontrivial acquisition and deployment of a defensive package may well occur. There is less reason to expect symmetrical (parallel) defensive deployment by both parties under the present case than under previous cases. The over-all constraint puts a heavy premium on each side's *relative* valuation of the different elements in its posture. The over-all situation of the U.S. and Soviet Union (resources, relative costs, preferences) differs in important respects. It is likely that they differ with respect to the usefulness of defense capability. Certainly the past weapons acquisition experience of the two nations supports this presumption. Under over-all force level constraints, the Soviets are likely to buy considerably more defense than the U.S.

The consequences of asymmetric defense deployment will be a pastiche of effects that we have already considered. The constraints against easy offset policies will prevent the strategic interplay from cancelling out. Some decrease in assured destructiveness will be "suffered" by both sides, and some increase in operational uncertainty. As in the case just

discussed, these may result in a net enhancement of mutual deterrence against central nuclear war.

For both types of arms control, we may now ask whether deterrence against nonnuclear provocation will also be enhanced, as under the *ceteris paribus* no arms control situation. It is true that possession of some defensive capability makes either a first or second strike more conceivable, and therefore makes it all the more important to prevent an uncontrolled escalation to the point of central nuclear attack. But such possession can also be looked on as conferring a kind of insurance against error. A nation with considerable defense might believe that it can dabble in non-nuclear provocations—limited wars with some escalation—taking the risk that if it has miscalculated the costs will not be horrendous. A conservative, essentially status quo power might be more influenced by the former consideration—that nuclear war is not unthinkable to a nation willing to take risks (i.e., to gamble within the broader sea of operational uncertainty)—and be deterred from infranuclear provocations. An adventurous, but not necessarily implacably aggressive, nation might be more influenced by its damage-limiting insurance and be led to probe provocatively. Thus, despite what may well be an enhancement of mutual deterrence against central nuclear war, one cannot gauge the effect on deterrence against infranuclear provocation. The difficulty is not due to the weakness of the over-all effect, but to the opposing direction of influences of its components.

C. DISTINCTIVE INGREDIENTS OF DEFENSIVE POSTURES RELEVANT TO DETERRENCE

In my discussion so far I have emphasized the impact of defensive postures in reducing the enemy's assured destruction on either a first or second strike, and for either counterforce or countervalue targets. I have also emphasized that active and passive defenses are likely to increase operational uncertainty for an attacker. It pays to examine this a bit further, and also to discuss briefly some other aspects of defensive postures which bear on deterrence.

1. *Uncertainties for the Attacker.* In advance of an attack the potential attacker cannot be sure how successfully his attack will go. Compounding the unreliability of his own weapons—which he must contend with regardless of the defender's capabilities—are the ability

of active defenses to intercept his weapons, and the ability of passive defenses to prevent damage from his weapons. At best, he can estimate only a probability distribution of outcomes—a distribution characterized by lack of truly relevant experience in its formulation.

Some of these uncertainties can be dispelled after the attack. To do so will typically require nontrivial investments in specialized resources—reconnaissance and other forms of intelligence. But some of the uncertainties are extremely difficult to dispel even after the attack. Destruction of hidden enemy weapons may remain largely unknown to the attacker. Even an approximate estimate of human casualties may be out of the question if the population is sheltered in blast (especially) or even fallout shelters. Yet knowledge about this destruction after the attack may well be crucial in determining how the war should be pursued. The ability of the attacker's continuing capabilities to perform depends on such information, regardless of the extent to which the attacker has perfected continuity of command and control, flexibility of targeting, salvo capabilities, and penetration aids, as (expensive) adaptations to a defensive environment.

The existence of these uncertainties substantially complicates the problem of conducting a successful war. These uncertainties bear on both initial attacker and defender once the war has begun. But they bear especially on the potential attacker in deciding whether to start the war, since that decision rests on the burden of anticipated success. One would expect the existence of these uncertainties to engender substantial caution—deterrence—in any would-be attacker.

2. *Crisis Management.* The role of defense systems in crisis management has been treated by Schelling and others. I will briefly refer to some of the issues in these treatments. The crucial point is that to be used during crisis situations passive population defense systems must be activated while active defense systems (and passive weapons defense) may for the most part remain passive. Fixed or fixed orbiting ABM and antibomber systems achieve and maintain a state of alert without any overt change. Manned interceptors, ASW, and airborne alert for strategic bombers do require an overt change in deployment, but these are not likely to be the main lines of defense. On the other hand, large-scale sheltering of populations requires very overt alerting and mobilizing actions. Thus, it is passive population defense that performs the function

of the signaling and communicating with one's allies and enemies that used to be performed by military mobilization as a whole.

This communication provides a strategic medium for conveying threats and assurances which help determine the course of the crisis. As such, it is of real importance for deterrence. The circumstances under which deterrence is most necessary to prevent war are not isolated periods of relative tranquility but periods of crisis. Deterrence consists in providing that even in the midst of crisis the net payoff for making a no-fight decision will substantially exceed that for a decision to fight. The dynamics of mobilization have been historically known to turn relative payoffs suddenly in favor of fighting, even though longer-view considerations presumably favored peace. (World War I may be a case in point.) So too this public medium for conveying fears, intentions, and threats, will affect the outcome of crisis just by its very existence, and despite the intentions of the nations which deploy such defensive systems. Its use or nonuse is inextricably caught up in the dynamics of crisis development.

Unfortunately, population sheltering is a crude instrument for such a function. It is awkward in that so many people have to be mobilized into unnatural and inconvenient postures. This can be accomplished only to the accompaniment of widespread and deep psychological alarm —terror, fear, anxiety. But these emotional states cannot be turned on and off at will. They have a life and an inertia of their own. Thus, the defensive weapon of sheltering cannot be used with subtlety as a strategic tool. Just because of this, leads and lags become important, along with other dynamic properties possessed by general military mobilization in a former era. True, it can be employed to try to allay crisis—by scrupulous avoidance of use—but even here its nonuse by one side gives the other side an advantage in using it. Furthermore, in its absence, other tools of strategic interplay would be employed—it is not indispensable. But, on balance, its existence probably adds a not inconsiderable element of instability into crisis management.

3. *The Defender's Options.* One final issue is the impact of defense capabilities on the options open to a potential defender. In the absence of such capabilities, if an enemy makes menacing gestures in the course of a crisis, the current military advantage in being the first to attack could well stampede a potential defender into becoming an actual at-

tacker. In other words, preemptive or even preventive attack looks more attractive than being helplessly smashed. But this knife-edge trigger orientation can lead to many—and fatal—false alarms. Both sides benefit if either can wait to see whether what looks like an impending or already-begun attack is really that.

The possession of significant defensive capabilities provides just such an option to wait. False alarms, accidental enemy firings, or third-party "catalytic" firings, can be waited out with little damage and distinguished from serious enemy hostilities. This flexibility of option introduces an extremely important element of stability into a balance of deterrence. If one believes—as quite a number do—that neither the U.S. nor the Soviet Union has aggressive enough intents to attack the other just to take advantage of a chink in the armor of mutual deterrence (except possibly when in the throes of an agonizing crisis), then the whole previous analysis of defense and deterrence becomes largely irrelevant. The most truly relevant issue becomes the problem of accidental or mischievous, catalytic war. And it is measures like "the hot line" and the defense-imparted ability to wait out false alarms that most truly deter central nuclear war. Whatever the relative importance of different elements in deterrence, the present one is considerable in absolute terms.

IV. Defensive Systems and the Stability of Mutual Deterrence

A. INTRODUCTION

I have already considered many of the elements that are included in an analysis of the stability of deterrence. It will be useful to draw them together at this time, along with some additional material.

Deterrence refers to a situation where the military capability of one antagonist and its avowed or implied conditional employment, result in a situation in which the other antagonist finds a policy of no-fight preferable to one of fight. The stability of deterrence refers to the range of circumstances over which the no-fight policy will continue to seem preferable to the fight policy, the larger the range the greater the stability. In other words, are not-unreasonably changing events likely to change an antagonist's no-fight choice?

A systematic answer would examine various types of changes in

circumstance—technological change, sheer quantitative expansion of force size, force-level reductions and structural changes in the course of disarmament and other arms control measures, changes in political goals, etc. My purpose here is to concentrate on the effects of defensive postures.

B. DEFENSE DEPLOYMENT AND FIRST STRIKE INCENTIVES

I have already examined this in general terms. The first antagonist deploying a reasonably efficient defensive package can experience a radical change in the incentive to make a first strike in the period before his enemy has had time to respond. The increase in his damage-limiting potential can degrade his enemy's deterrence to the vanishing point. Whether or not a specific nation at a specific time and place, will find it actually worthwhile to make a first strike depends, of course, on the specifics of the situation—the efficacy of his defense package, the precise nature of his and his enemy's military force structures, and the operational strategies and tactics of the two (and of course the tension level.)

This potential instability disappears once the enemy has had time to adapt his forces for a direct offset (through offensive enhancement) to the defensive deployment. Exact data are not so necessary to justify this prediction so long as the antagonist is relatively free to change his forces, and information about approximate weapons performance is available to both sides. Inadequate, misleading information can make any objective situation unstable, by leading to miscalculation.[4]

When the enemy's response to an initial defense deployment is a similar deployment without offensive augmentation—an unlikely response when there is no arms control—there may be a complicated change in first-strike incentives: both parties may find such a strike attractive. How likely this is depends heavily on the specifics, as before; but more, it depends also on the particular goal evaluations of the two sides. What does each nation hope to obtain through war? What is its weighing of the various costs of war?

Where arms control agreements exert constraints on force structures, defense deployments can enhance first strike incentives if they provoke a defensive arms race, or if they give rise to significantly asymmetric

[4] Except, as noted above, that when information is known to be inadequate more caution rather than less is likely to be observed.

postures. I have argued that built-in strategic constraints on defense deployments under these circumstances are likely to produce configurations where mutual deterrence is enhanced rather than degraded. Specific empirical evidence (including simulation outcomes) would serve to demarcate the two types of outcomes.

In sum, our qualitative conclusions support the possibility of instability in the very short run, but enhancement of deterrence in a longer-run perspective in cases presumed to be "normal". Verification, and the identification of circumstances which conform to these "typical" results, require empirical observation.

C. DEFENSE FOR OFFENSIVE WEAPONS VS. DEFENSE FOR URBAN AREAS

Urban areas and strategic offensive weapons can be subjected to both passive and active defenses—for example, by hardening (sheltering) and ABM, respectively. Indeed, by locating the two near enough together, the same ABM systems can protect the two. However, the efficacy of defense for the two, the feasibility (resource and "human" costs), the indispensability, the strategic consequences, and the ultimate significance, differ markedly.

Briefly, offensive weapons are easier to defend both actively and passively. They can be dug in more feasibly and cheaply; and can be kept dug in—and alerted—indefinitely (subject, of course, to normal depreciation). Populations cannot: sheltering is unnatural and inconvenient for them; it represents only a temporary, emergency possibility. Weapons, moreover can be easily and decisively dispersed; populations cannot be. The way of life, whose protection is after all one of the main ultimate reasons—if not *the* main reason—for the whole state of international tension which gives relevance to military forces—requires heavy clustering. Moreover, weapons can be expanded in number to whatever "reasonable" levels are desired; and they are expendable. Populations are whatever numbers *they* themselves determine (i.e., prewar); and they are not expendable. Lastly, the combination of expandibility, dispersability, and expendability for weapons means that adequate sheltering and/or active defense are not an indispensable means of defending an offensive force structure. There is the alternative of proliferation or dispersal, or mobility or hiding, or combinations of these. In terms of cost alone, then, defense of weapons is by far the better buy.

The strategic implications of the two differ markedly. Defense of strategic attack weapons adds to the invulnerability of a retaliatory force. This has almost entirely a deterrent, not an aggressive, impact. The defense is most relevant where the enemy has made a counterforce first strike. It guarantees a larger potential retaliation, and therefore enhances deterrence against such a first strike. Thus, defense of strategic weapons, while relatively inexpensive, is also relatively nonprovocative.[5]

Defense of urban populations can also, as we have seen, be part of a purely defensive, nonaggressive posture. But by withdrawing a portion of the antagonist's hostages which form the bulwark of his deterrence, it *could* also have the aggressive significance of making a first strike attractive. Therefore, no matter whether *intended* defensively or aggressively, if international tension is high enough, the enemy will have to interpret it as an aggressive act for safety's sake, and act accordingly. This gives it an aggressive significance, whatever its intent. Thus, defense of urban populations can seriously destabilize mutual deterrence. It need not under all circumstances, of course, but it may.

A final point on this score is that while the various forms of defense for retaliatory weapons can be hidden, urban defense is not capable of much secrecy. So an attempt to avoid the potential instability caused by urban defense by keeping it covert is almost certain to fail.

Thus, on grounds of both effectiveness (cost) and strategic impact, defense of retaliatory weapons would seem much preferable to urban defense. This paradoxical conclusion has another possibly even more paradoxical implication. To avoid provocation and enhance deterrence, one should locate weapons away from urban areas, lest the defense of these weapons (especially active defense—e.g., ABM) be thought to have a provocative, population defense function as well. Space weapons are the logical extreme for this kind of reasoning.[6]

The logic of a deployment of extreme separation between weapons and ultimate value targets is that these latter have a better chance to

[5] Its only aggressive function is as a protection of weapons held back by an attacker who wants to be able to follow up his initial attack with a second one.

[6] This is not the only reason for separating weapons location from populations. Another reason is that if either side finds a first strike attractive it will probably be a counterforce strike. Separation will tend to minimize the collateral value damage to any such counterforce first strike. Weapons placed in space are obviously ideal for this purpose. A counterforce war can be fought with no value damage whatever.

survive under a situation of balanced mutual deterrence without value defense than under a value defense situation which destabilized mutual deterrence. Indeed, in the absence of arms control, not only is strategic instability possible, with its attendant risk of war, but also the defense deployment is likely to be directly and conveniently offset by offensive enhancement by the other side. The only situation in which this argument fails is where an arms control agreement makes possible a lasting protective function for urban defense. As indicated in an earlier section, population defense under such circumstances could actually enhance deterrence. But the failure of the argument for pure deterrence does not mean that only urban and not weapons defense would be warranted here. Indeed, most of the defense procurement would still go to insure the invulnerability of attack weapons. Besides, the alternative argument given above for separation of weapons from cities would still strongly argue for such separation.

D. DURABILITY OF DEFENSE

How long does the effectiveness of a defensive system last? This is relevant to stability because unless the defense can efficiently endure as long as the threat against it, time will degrade it and change the relative power of the participants accordingly. The question of durability can be asked both of a short-run operational period, during which neither offensive threat nor defensive shield can be replaced or augmented, and of a long-run period during which quantitative and qualitative changes of note can be made in relative force structures. Both are relevant to some aspect of the stability of deterrence, but the first is somewhat more immediate. We may compare the durability in both these senses for both urban and retaliatory weapon defenses.

On the operational period level, possibly the key issue is the movement into and out of, and the continued maintenance of, an alert status. As indicated above, a significance is imparted because the population-sheltering component of urban defense requires explicit and cumbersome mobilization procedures in order to attain and maintain alert status, while hard-point active and passive defense, as well as the active defense component of urban defense, do not. Moreover, sheltering is an unnatural and costly situation for the population. Except for the capital cost of the installation (which exists for population sheltering as well), measures for insuring the survivability of second-strike forces are not

"unnatural" or costly (but there *is* an operating cost involved in hiding submarine-launched ballistic missiles [SLBM]).

There are two implications. First, despite the *existence* of passive city defense, a surprise attack might succeed in catching it in a nonalert status, so that, in effect, it cannot be used to decrease population damage: it is bypassed. The active defense component would be operative, but its productivity would be hampered by the bypassing of the complementary passive component. The prospective decrease in defense effectiveness might be sufficient to make a first-strike policy attractive to a potential attacker.

Second, once a population has been mobilized into shelters—as, for example, during a serious crisis—it is costly and difficult to keep them there. There are two types of cost involved. One is the sheer psychological strain of living underground in cramped, crowded, unhealthy conditions, lacking in most of the amenities of normal life. Possibly most important is the high and unrelieved sense of anxiety, fear, terror involved in waiting for the worst to happen where, because of sheltering, the worst seems imminent. Continued for any length of time, this could inflict extremely serious psychic trauma. The other type of cost is economic. Sheltering precludes normal urban economic activity. The loss of income is directly related to the amount of time economic activities are suspended. Thus, maintenance of a state of alert in passive population defense incurs costs which rise inexorably and probably at an increasing rate (because of increasing economic start-up costs and deepening psychic difficulties).

These increasingly burdensome costs of continued alert, in a situation where an expected attack has failed to materialize, could lead to unsheltering before the end of the crisis is reached. A potential attacker, incurring far less cost in being on offensive alert (as with much of active defense), could pick just this demobilization to initiate his attack. Indeed, he has a strategic interest in provoking mobilization and bringing on "shelter fatigue"—or, even more so, in evoking the second if the first has already come about—in order to take advantage of the resulting expected unsheltering. The risk, of course, is that the enemy's loss of patience and defensive readiness may shift the initiative to him: the erstwhile potential victim may become the actual attacker. In any case, the attrition brought about by sheer duration imparts a potentially unstable element to crisis management, an element which is lacking in the measures necessary to guarantee the invulnerability of second-strike forces.

On the long-run duration level the story is different. Here, we are not concerned with the maintenance of a continued alert, but with the relative effectiveness of a whole military posture vis-à-vis the enemy's posture. The nemesis of a given set of measures for insuring second-strike invulnerability is that the enemy may overtake them by means either of sheer numerical acquisition or of qualitative changes in his weaponry: technological change. First-strike weapons may be improved enough with respect to reliability, accuracy, yield, or production of new effects so as seriously to degrade a previously effective deterrent posture by the enemy. A counterforce first strike could succeed in destroying all but an ineffectual remnant of a once-adequate retaliatory capability.

Urban defense, both active and passive, is subject to just this same obsolescence through the enemy's technological advance. But it is neither more nor less subject than offensive systems. A qualitative arms race is especially charged with unstable episodes, because *any* set of initial postures for two parties, whatever the degree of balance between them, or whatever their internal composition, can be transformed into an unbalanced situation in which one or both parties have a military [7] incentive to attack the other. This is not a flat prediction of actual instability, but recognition of a probability distribution for which the probability of zero potential [8] instability over an extended period with unconstrained R and D is very, very low.

Thus, in terms of short-run operational durability, the prospects for a stable balance of deterrence are affected by whether urban defense is or is not deployed in addition to weapons defense. In terms of long-run durability, these prospects are a function of technological change, and do not depend in advance on the type of defense chosen.

V. Deterrence vs. Defense

A. THE TRANSITION FROM DETERRENCE TO DEFENSE STRATEGY

Given anything like present technology, the only situation where a defensive, rather than a deterrent, strategy could be viable is where a substantial defensive capacity buildup proceeded under the umbrella of mutually agreed limitation on offensive power. Such a mutual posture

[7] A "military" incentive to attack is not necessarily a *net* incentive, since the over-all level of tension, the conflicts separating the parties, may change in an offsetting direction during the same period.

[8] "Potential" in the sense of "military" incentives to initiate an attack.

would have the attractiveness of making a first strike unprofitable and any unintended or third party firing of little effect. The deterrence felt might be weaker than at present, but, in hostilities, damage levels and risks of escalation would be far lower.

Transition from a deterrent to a defensive strategy is complicated by their basic incompatibility: effective defense by one side degrades the possibility of the other's deterring attack through offensive retaliation. One cannot have an adequate deterrent while the other has an adequate defense. Thus, movement toward defense strategy can only succeed if it is parallel for the two.

A parallel defensive arms race is largely similar to an offensive race in that both sides competitively seek to reduce retaliatory damage to an acceptable level. Here too differential achievement can lead to strategic instability. But there are differences. First, unlike offensive races, even nondifferential achievement can lead to instability, since both sides may have simultaneous attractive first-strike options. In offensive races, offensive weapon is pitted against offensive weapon, so only one side at most can predominate over the other. Here, defensive weapon is pitted against offensive weapon, so it is possible for defense to predominate simultaneously for both sides without either having an over-all differential advantage. This source of instability may not be serious, since inflicted damage falls along with sustained damage. Willingness to attack probably depends on a relationship between gains and costs, not on costs alone.

A second difference is that the defensive race is viable only in the context of agreements limiting offensive weapons. Either or both sides may strategically choose to break the agreement as a response to a grossly disadvantageous development of the defensive race. This possibility of lumpy escalation serves to set limits to differential advantage within the race and introduce a distinctive form of instability.

B. TYPES OF DAMAGE AND DAMAGE LIMITATION AND THE COMPATIBILITY OF DETERRENCE AND DEFENSE

Up to now I have tacitly assumed that defense referred either to defense of weapons or defense of populations. Defense of urban areas represented an attempt to defend populations efficiently, by taking advantage of their clustering. While installation of hard-point defense had relevance mostly to the enemy's calculations about first strike desirability, installation of the second degraded the enemy's deterrent posi-

tion. By removing one's own population from the status of human hostages, the defense-deployer could paradoxically jeopardize that very population. It was for this reason that I considered deterrence and defense orientations incompatible in spirit.

This assumption in effect categorizes targets into only two classes. But there is a third type of target, and its existence may importantly revise my judgment. Populations are not the only kind of target deemed to represent the nation's ultimately valued elements. Economic resources are also prized, complementarily with human beings. Indeed the situation of humans without economic resources is hardly deemed more worthwhile then the situation of no humans at all. Moreover, the issue of many international conflicts concerns augmentation, protection, or use of a nation's economic base.

To some extent, consideration of economic resources does not change the analysis. An important portion of these resources is clustered along with populations in urban areas. Defense of these areas is partly a defense of both. But even here there are differences. While active defense may protect both, fall-out sheltering will protect populations but not production facilities located in cities. The disparity is even more striking for resources located outside urban areas. Industries located in small towns and rural areas cannot economically be defended. Much more important is agriculture. Agriculture can be almost completely hostage to enemy action with high yield, extremely dirty nuclear weapons, and no economically feasible defense against such targeting exists under presently foreseeable technology. Only highly efficient antilift stage weapons could be relevant, and these do not seem to be in the offing. Indeed, even if they were, they might not be the strategic weapons chosen for acquisition and deployment.

What this means is that even the massive defensive superiority envisioned in the previous section does not completely withdraw all useful value targets from under an enemy's retaliatory sights. Massive destruction of the arability of land may not have the immediate dramatic impact of huge human casualties, but the enormity of the loss would become apparent as existing food and material supplies became depleted. And the hopelessness of a human future under such circumstances would approximate—might even exceed—that of heavy human losses. The 30–50 per cent human decimation of the Black Death in Europe was after all supportable since it left resources intact (indeed, agricultural productivity even rose as a result). But catastrophic destruction of such

resources destroys the springboard from which the human spirit can spring back.

Highly prized values would therefore remain as targets even though a heavy defense buildup for populations took place. Offensive weaponry, if adapted to the new type of target, and without necessitating substantial augmentation, could still maintain deterrent power. This means that large-scale defense deployment would not have succeeded in changing basic strategic orientation away from deterrence to defense. It would simply have changed the relative attractiveness of different types of targets. The prospective payoffs that help create the conflicts which wars are called upon to settle, and which help to define victory and defeat in such wars, would remain hostage to offensive power despite heavy defensive commitment; thus, one is again led fundamentally to question the rationale behind a resource allocation to large-scale defense such as I have been discussing.

Finally, these considerations strike a chord that has been struck again and again throughout my study. In order to answer complicated questions, I am forced to go back to raise some very elementary questions. What are the issues that are likely to lead to wars? What are the aims that will be followed in carrying out such wars? Are they abstract aims like destroying the enemy, or doing him more damage than he does you, or minimizing your own damage? What kinds of damage are important? What are the trade-offs between different types of damage? Or are the aims of war instead more closely associated to the concrete aims that engendered the crisis and war? Are they some combination of the abstract and concrete? One may even have to ask: When the carrying out of hot or even cold war becomes terribly onerous, how does one side, or both, get out? What is the value of losing a given struggle in order to be able to extricate oneself from it, in order to resume a larger contest in a more favorable arena? How does a participant communicate such considerations?

These questions do not refer solely to the problem of defense. They are relevant here because they are relevant to broader problems of overall strategy; and if any single conclusion should follow from my entire discussion, it is that the techniques, the worthwhileness, the consequences of various defensive measures—the very significance of defense—relate intimately to strategic considerations of the broadest scope. I am not speaking of one egg in a basketful—I am speaking of the hen herself.