CHAPTER 2

The nuclear revolution and the problem of credibility

This chapter elaborates a framework within which to place the more detailed and narrowly focused analyses of the subsequent chapters. That framework encompasses the broader themes that will connect the more specific issues examined in later chapters. This chapter develops the framework in three steps. The first step summarizes what will be called the classical logic of war: the stylized relation that existed between the use or threatened use of force and states' attempts to further their interests before the nuclear revolution. The nuclear revolution undercut the classical logic and made the problem of credibility the paramount theoretical concern. The second step then reviews the two apparently quite different ways in which nuclear deterrence theory has tried to solve this problem by explaining how the use or threatened use of force is related to states' political objectives after the nuclear revolution. The first is based on Schelling's "threats that leave something to chance" (1960, 1966). In this approach to the credibility problem, states take steps during a crisis that raise the risk that the crisis will go out of control and escalate to a general nuclear war. The second approach is not based on the risk of losing control. Instead, a state deliberately imposes severe but nevertheless limited sanctions on an adversary in order to make the threat of future punishment sufficiently credible that the adversary will come to terms. After outlining these two approaches, the third step is to show that despite their apparent differences, the two approaches are fundamentally alike. Each attempts to solve the credibility problem in essentially the same way. Each uses an array of limited options to bridge the gap between doing too much by launching a massive nuclear attack and doing too little by acquiescing to an adversary. The idea behind these limited options is that a state may be able to make the threat to use them more credible than the threat to launch a massive nuclear attack and in this way avoid having to submit to its adversary.

The classical logic of war

What is the nuclear revolution, and what are its consequences? How has it changed the logic of war? How, that is, has the nuclear revolution changed the relation between the use or threatened use of force and states' attempts

to secure their political ends? To address these questions, one must first have some notion of what the logic of war was before the nuclear revolution.[1]

Carl von Clausewitz, writing in the aftermath of the Napoleonic Wars, described the logic of war that existed before the nuclear revolution: "If the enemy is to be coerced you must put him in a situation that is even more unpleasant that the sacrifice you call on him to make" (Clausewitz 1976, p. 77). As it stands, this formulation of coercion is timeless. If an adversary is to make the political sacrifices demanded of it, then the cost of refusing to make them must appear to be still greater. What bounds this formulation and gives it meaning is the way that force or the threat of it can be used to make an adversary's situation unpleasant. Indeed, what will distinguish the relations between the use and threatened use of force and states' attempts to further their interests before and after the nuclear revolution are the different ways that force can be used to bring coercive pressure to bear.

Deterrence and the distinction between punitive and defensive capabilities are crucial to describing the different ways that coercive pressure may be exerted. Deterrence is a form of coercion. A state deters an adversary from doing something like attacking by convincing it that the cost of doing so would be greater than the potential gain.[2] A state's defensive capability is the state's physical ability to limit the costs an adversary can impose on it (Snyder 1961, p. 3). The greater a state's defensive capability, the less an adversary can hurt it. In addition to being able to limit the costs an adversary can impose on it, a state may be able to inflict costs on an adversary. These costs may include the invasion and occupation of some of its territory, the destruction of its military forces, the devastation of some of its industrial capability, or, more simply and gruesomely, the killing of some of its people. A state's punitive capability is its ability to inflict costs on an adversary. The greater a state's punitive capability, the more punishment it can impose.

The adjective "physical" in the definition of defensive capability is important. If a state deters an adversary from invading it by threatening to

[1] This discussion of the logic of war and the nature of the nuclear revolution draws heavily on the important contributions of Brodie (1959), Snyder (1961, pp. 3–51), and Schelling (1966, pp. 1–34).

[2] In some contexts it may be useful to distinguish between *deterring* an adversary from doing something and *compelling* an adversary to do something; see Schelling (1966, pp. 69–91) for a discussion of the difference between deterrence and compellence. This distinction is, however, conceptually elusive. The difference between deterring an adversary from attacking and compelling it not to attack is unclear. At a more general level, deterrence and compellence are alike: In each, a state is trying to coerce its adversary into acting in certain ways and not in others by shaping the adversary's estimates of the costs and benefits. No distinction will be made here between deterrence and compellence.

impose grave punishment, then the state will have succeeded in limiting its costs. But this does not imply that the state has a strong defense, for the state might have been physically unable to limit the cost of being invaded had its adversary actually decided to invade. Instead, the state, although physically unable to repel an invasion had it occurred, was able to coerce its adversary into not attacking by making the prospect of unacceptable punishment sufficiently likely that the adversary did not invade because the expected cost of doing so seemed greater than the expected gain.

Three aspects of the distinction between punitive and defensive capabilities should be emphasized. First, as just suggested, both capabilities may contribute to deterrence. Whether a state is trying to deter an adversary from invading or convince an adversary not to mount further resistance after it has been invaded, both capabilities are related to a state's ability to influence its adversary's actions. To make the cost of failing to comply greater than the cost of doing so, a state must have, or at least appear to have, the ability actually to impose sufficiently high costs on an adversary. This is the role of a state's punitive capabilities. But deterrence requires more than the ability to impose costs. An adversary must be sufficiently convinced that the state will use its punitive capabilities. This judgment would seem to be affected by the state's ability to limit the costs that an adversary can impose on it in retaliation. The greater a state's defensive capability, the less its adversary can hurt it, and the more likely it may be to use its punitive capabilities on its adversary. Accordingly, the ability to place an adversary in a situation the continuation of which will be more costly than the sacrifice it is being asked to make is related to both punitive and defensive capabilities.[3]

[3] Snyder (1961, pp. 14–16) and others, such as Schelling (1966) and Jervis (1984), point out that both capabilities may contribute to deterrence. However, trying to identify a separate form of deterrence with each type of capability, as Snyder does with his distinction between deterrence by denial and deterrence by punishment, is quite problematic.

In Snyder's formulation, a state deters an adversary from invading by denial by being physically able to "deny territorial gains to the enemy" (Snyder 1961, p. 14), or at least to make a successful invasion less likely. More generally, a state deters by denial by being physically able to deny an adversary its goal or, as Snyder puts it, by affecting "the probability of gaining his [the adversary's] objective" (1961, p. 15). Deterrence by punishment is different. Here, a state deters an adversary from invading not by being physically able to stop an invasion but by credibly threatening to impose enough punishment so that the costs of invading will seem greater than the potential gains. In this formulation, an army that would fight the invaders primarily contributes to deterrence by denial. A strategic nuclear force only capable of inflicting punishment by destroying an adversary's cities contributes mostly to deterrence by punishment.

The difficulty with this formulation is that it is more natural to think of a potential invader's objective not as simply to invade and occupy some territory but to do so at some acceptable cost. But then, as soon as a state's strategic nuclear arsenal can impose still higher costs, the state becomes physically able to deny its adversary its objective. Thus, what is perhaps the clearest example of a punitive capability (i.e., a state's strategic nuclear force capable only of destroying an adversary's cities) appears now to be deterring by denial, because this force is physically able to deny an adversary its broader political objective of occupying some of this state's territory at some acceptable cost. Identifying types of deterrence with types of capabilities is problematic. Snyder may be closer to the mark when he suggests that the difference between deterrence by denial and deterrence by punishment may have less to do with a formal distinction between the means of deterring and more to do with beliefs (which are perhaps based on historical experience) about the types of reactions that various actions, such as invasions, are likely to provoke (Snyder 1961, pp. 14–16).

The second point is that a state's punitive capability is related to its adversary's defensive capability. The greater a state's punitive capabilities, the higher the costs it can impose on an adversary, and thus the less physically able an adversary is to limit the costs that can be imposed on it. There is an inverse relation between a state's punitive capability and its adversary's defensive capability.

Finally, it is important to emphasize that although these two capabilities are conceptually separate, actual military forces may combine both of these capabilities. An army capable of repelling an invasion and thereby limiting the costs an adversary can impose may also be used to launch an invasion and inflict costs on an adversary by taking some of its territory. By helping to limit costs, the army contributes to the state's defensive capabilities. By being able to take what an adversary values, the army contributes to a state's punitive capabilities.

Much turns on whether or not punitive and defensive capabilities are generally conflated in the same forces. Indeed, this crucially affects the relation between the use and threatened use of force and states' attempts to secure their ends. The classical logic of war assumed that these capabilities were conflated. As will be seen, the separation of these capabilities and the development of states' punitive capabilities undercut the classical logic and marked the nuclear revolution (Snyder 1961, pp. 8–9; Schelling 1966, pp. 1–34; Jervis 1984, p. 26).

When these two capabilities were conflated, the same forces that limited the costs an adversary could impose also increased a state's ability to impose costs on its adversary, especially by taking its territory. Two consequences follow from the conflation of these two capabilities in the same forces. First, being militarily stronger could enhance deterrence by raising the expected cost an adversary would have to bear if it attacked. The state's greater punitive capability would mean that the adversary would have to pay a higher price if the state actually used its capability. Second, this state, because it would be less vulnerable to its adversary, might be more willing to use its capability. To the extent that raising the expected

cost of attacking for the adversary enhanced deterrence by reducing the chances of an attack, military strength was the key to security in the classical logic of war.[4]

The conflation of punitive and defensive capabilities also defined a reasonably clear relation between the use or threatened use of force and states' political ends, at least in the case in which a profound conflict of interest divided the states. If a supremely important political objective required an adversary to give up something of great value, then in this extreme a state might want to try to put its adversary in the worst of all possible positions. For Clausewitz, "the worst of all conditions in which a belligerent can find himself is to be utterly defenseless. Consequently, if you are to force the enemy, by making war on him, to do your bidding, you must either make him literally defenseless or at least put him in a position that makes this danger probable" (1976, p. 77).

Because punitive and defensive capabilities were conflated, rendering an adversary defenseless would also destroy its punitive capabilities. The destruction of both of these capabilities would give a state political control over its adversary. But that control would not be absolute. Even after becoming defenseless, a state still could refuse to do the victor's bidding. But in doing so, this state would have to suffer whatever punishment the victor decided to inflict and would be unable to retaliate against the victor in any meaningful way.[5] Before the nuclear revolution, "military victory was the price of admission," to use Schelling's apt description (Schelling 1966, p. 17). That is, the victor, after having already destroyed its adversary's military forces, might have to inflict still more punishment on the defeated state in order to convince it to do the victor's bidding. Its defeat did not assure that the adversary would do this automatically. But because the defenseless adversary could no longer pose a significant threat to hurt the victor in retaliation, the cost to the victor of carrying out its threats to hurt its adversary if the adversary refused to do the victor's bidding was as low as possible. In this way, the victor's being able to protect itself from

[4] Strengthening the state may make an adversary less likely to attack, but it may not. By becoming stronger, a state may raise the expected cost of attacking for an adversary. But because the strengthened state will have greater punitive capabilities, its adversary will be more vulnerable and, fearing still greater vulnerability, may find that the expected cost of not attacking is also rising. On balance, this may leave the adversary more likely to attack. This is the essence of the security dilemma, in which one state's effort to increase its own security by reducing its vulnerability reduces another state's security by increasing its vulnerability. For a discussion of this, see Jervis (1978).

[5] The assumed conflation of punitive and defensive capabilities is, of course, a simplification and a stylization. Even after losing its military forces, an adversary may still retain some punitive capabilities in the form of guerrilla resistance, for example. The assumption here is that any residual punitive capabilities are insignificant.

any retaliation made its threats to punish its adversary as believable as possible.

In the stylization of the classical logic of war, a great war fought over a profound conflict of interest could be thought of as a two-stage process. The first was a contest of relative military strength. During this stage each state still tried to coerce the other by making the cost of resistance seem greater than the cost of compliance. But these costs were primarily affected by course of the military struggle. Once the military struggle had been decided, then if the threat inherent in having been defeated had not already done so, the victor could use the "power to hurt" (Schelling 1966, p. 3) to raise the expected cost of further resistance to such a level that it would exceed the expected cost of coming to terms (Snyder 1961, p. 11; Schelling 1966, pp. 12–18; Jervis 1984, p. 27).

The assumption that the punitive and defensive capabilities were conflated in the same forces was crucial to the stylized relation between the use and threatened use of force and states' aims that existed before the nuclear revolution. This assumption meant that a state that had rendered its adversary defenseless would be able to defend itself. If, therefore, the victor chose to try to coerce a defeated yet defiant adversary into doing its bidding by punishing it, the victor could be confident of defending itself from any attempted retaliation. This is critical, for if the victor were unable to protect itself, it would not have control. "So long as I have not overthrown my opponent [i.e., rendered him defenseless] I am bound to fear that he may overthrow me. Thus I am not in control: he dictates to me as much as I dictate to him" (Clausewitz 1976, p. 77). The essence of the classical logic of war was that defense, at least for the victor, was possible (Brodie 1959, pp. 147–222; Schelling 1966, pp. 1–34). That defined the aim of warfare, at least in the extreme in which a profound conflict of interest divided two states. By rendering an adversary defenseless, a state, because it still could protect itself, would have put its adversary in the worst of all possible positions (Clausewitz 1976, p. 77).

The rise of strategic air power, the development of intercontinental ballistic missiles, and the advent of atomic and then thermonuclear weapons separated the ability to punish from the ability to limit the punishment one might have to suffer.[6] This growing separation cast doubt

[6] These two capabilities had in reality always been separate in varying degrees. The English forces used to carry out the *chevauchées* during the Hundred Years War, the British naval forces used to blockade Germany during World War I, the German submarines that attempted to blockade Great Britain, and the strategic air forces employed during World War II were more effective in punishing an adversary than in limiting the costs an adversary could impose. The classical logic of war and the assumption that punitive and defensive capabilities are conflated are at best useful stylizations and simplifications.

on the classical logic.[7] The nuclear revolution, that is, the advent of a technological condition of mutually assured destruction, completely undercut this logic. Rendering an adversary defenseless no longer meant that a state could also effectively limit the cost that its adversary could impose on it. Once two superpowers acquired secure second-strike capabilities, each state in effect had already rendered its adversary defenseless. Neither had the physical ability to limit the damage that the other could impose on it should the other decide to do so. But because the ability to punish was no longer conflated with the physical ability to limit punishment to oneself, rendering an adversary defenseless no longer brought political control. How did deterrence work when punitive and defensive capabilities were no longer reinforcing each other?

Coercion still required a state to be able to put its adversary in a situation the continuation of which would seem more costly than would complying with the state's demands. That, in turn, depended, first, on a state's being able to inflict a sufficient amount of punishment on an adversary and, second, on a state's being able to make the threat to use that capability sufficiently credible. In a condition of mutually assured destruction in which each state could destroy the other even after absorbing a first strike, the first condition certainly seemed to be satisfied. But what of the second? What of the credibility problem? Could a state convince its adversary that it would use its capability to punish when it was vulnerable to its adversary's retaliation? Could the threat to use these capabilities be credible? Once the rise of air power and then ballistic missiles had separated punitive and defensive capabilities, and once the nuclear revolution had made defense impossible, what was the relation between the use and threatened use of force and states' attempts to secure their interests?

The arrays of risk and punishment

The debate over the doctrine of massive retaliation brought the problem of credibility to the fore. In January 1954, Secretary of State Dulles, in a major speech on American foreign policy, reemphasized some of the themes of President Eisenhower's state-of-the-union address. Dulles's speech was ambiguous, but when seen in the context of the administration's efforts to limit military spending and its emphasis on nuclear weapons, the speech was interpreted by some to imply that the United States had adopted a strategy of massive retaliation: "in the event of another proxy or brushfire war in Korea, Indochina, Iran or anywhere else, the United States might retaliate instantly with atomic weapons against the U.S.S.R. or Red China"

[7] For discussions of the effects of the rise of air power and the advent of nuclear weapons, see Brodie (1959, pp. 3–222), Quester (1966), and Freedman (1989, pp. 3–44).

(Reston 1954). That is, the United States would rely on the threat of massive nuclear retaliation to protect the entire spectrum of American interests, ranging from the most peripheral to the most vital.[8]

Almost immediately the doctrine of massive retaliation was harshly criticized as being incredible and therefore ineffective.[9] This, however, was a debate about how best to protect less important American interests. As long as the United States was relatively invulnerable to a Soviet nuclear attack, massive retaliation seemed to be a credible means of protecting vital American interests like Western Europe (Brodie 1966, pp. 27–8). But once the United States became vulnerable to a devastating Soviet attack, the credibility of an American threat to launch a massive nuclear attack even in an attempt to protect vital interests became problematic.[10] How could a state credibly threaten to launch a massive nuclear attack when carrying out the threat would bring its own destruction? The policy debate focused on the problem of extended deterrence. How, that is, could the United States extend its ability to deter the Soviet Union from attacking the United States to deterring the Soviet Union from attacking Western Europe?[11] The more general issue was to understand how a state might credibly threaten to do what seemed to be inherently incredible.[12]

As technical and political circumstances and conditions changed after the demise of the doctrine of massive retaliation, many nuclear strategies and policies were devised. The 1960s saw Secretary of Defence McNamara propose the "no cities" doctrine, in which American strategic nuclear forces would not be aimed at Soviet cities but at Soviet military capabilities. The

[8] For a more detailed discussion of massive retaliation, see Gaddis (1981) and, especially, Wells (1981). Rosenberg (1983) offers an illuminating discussion of the Eisenhower administration's attitudes toward nuclear weapons and attempts to control them.

[9] Kaufmann (1956) has provided the classical criticism.

[10] Betts (1987, pp. 144–72) has traced the evolution of American assessments of American vulnerability to a Soviet nuclear attack.

[11] The problem of extending deterrence to cover vital American interests such as Western Europe has greatly influenced the evolution of American nuclear strategy. Freedman (1989) has provided a summary of this work and an extensive bibliography.

[12] There is an important implicit assumption here that should be made explicit. The credibility problem arises because it is assumed that the chances that a state will see any political objective as being worth the cost of bringing about its own destruction by launching a massive nuclear attack against an adversary are too small to deter this adversary. This assumption implies that the threat to order this attack and thereby bring a devastating counterattack in return is incredible, because sacrificing the political objective is less costly than launching a massive nuclear attack. This makes carrying out the threat inherent in the doctrine of massive retaliation irrational. If, however, some political objectives are worth certain destruction, then implementing the threat may be rational, and the doctrine of massive retaliation may be credible. In order to focus on the credibility problem, it will be assumed that no political objective is worth certain destruction.

United States would, in effect, try to hold Soviet cities hostage in order to give the Soviet Union the "strongest imaginable incentive to refrain from striking our own [i.e., American] cities" (McNamara 1962, p. 62). That strategy was quickly forsaken for the doctrine of assured destruction, which emphasized being able to destroy 20 to 25 percent of the Soviet population and 50 percent of its industrial capacity after absorbing a first strike (Enthoven and Smith 1971, p. 175). The North Atlantic Treaty Organization (NATO) debated and then in 1967 adopted the doctrine of flexible response, which, at least from the American perspective, was designed to enhance deterrence by raising the nuclear threshold by reducing NATO's dependence on the early use of nuclear weapons. The call for greater flexibility continued in the early 1970s with the Schlesinger doctrine, which tried to find implementable limited nuclear options that might be used to strike Soviet military targets or to demonstrate resolve. At the same time, President Nixon spoke of the strategy of "sufficiency," which meant "the maintenance of forces adequate to prevent us and our allies from being coerced" (Nixon 1971, p. 170). Later in the decade and into the early 1980s, nuclear policy centered on the countervailing strategy and the importance of having escalation dominance. In 1983, strategic defenses returned to center stage.[13]

The connections linking these policies and strategies to nuclear deterrence theory often were loose and rather tenuous. There were at least two reasons for this. First, nuclear policies and strategies have been the outcomes of bureaucratic and political processes that have reflected more than national security concerns.[14] A second and perhaps more important reason has to do with the weakness of deterrence theory itself. A powerful theory of nuclear deterrence would specify in more or less detail the likely consequences of various strategies and policies. If there were such a powerful theory, a state might then be expected to take it into account in formulating its nuclear strategies and policies. There would be a close connection between theory and policy. But when the theory is weak and often provides little insight into the detailed and pressing problems of policy, there is little reason to expect a state's strategies and policies to be anything more than vaguely related to nuclear deterrence theory's account of the relation between the use of force or the threat of it and states' attempts to further their ends.

Although there have been many nuclear policies and strategies, nuclear

[13] For discussions of these policies and strategies, see Enthoven and Smith (1971), Kahan (1975), Ball (1980), Freedman (1989), Schilling (1981), Slocombe (1981), Jervis (1984), and Sagan (1989b).

[14] Steinbruner (1974), Ball (1980), and Rosenberg (1983) have described examples of these processes.

deterrence theory has generally approached the credibility problem from one of two perspectives. Both approaches try to link the possible use of force to states' political objectives in what would seem to be the most difficult and demanding case. This is the stylization in which the condition of mutually assured destruction is interpreted in its strictest sense. In this stylization, there is no advantage to launching an unlimited nuclear attack first rather than second. In the event of a general exchange, it makes no difference if a state strikes first or is struck first. This is the strictest interpretation, because if there is no advantage to striking first, then as long as a state believes that there is the slightest chance that early warnings are erroneous and an adversary has not attacked, this state will not launch a general attack. It is always better to do something else. In this strictest interpretation, there is no situation in which it is rational for a state deliberately to launch an unlimited nuclear attack first.

The first approach to understanding the credibility problem is a direct conceptual descendant of the doctrine of massive retaliation, in that both appeal to the same sanction. In this approach, a state would still try to secure its interests by relying on the sanction of a massive nuclear attack. Schelling (1960, 1962b, 1966) provided the insight that explains how, at least in principle, this sanction might be linked to states' attempts to secure their ends after the nuclear revolution. Although in the strictest interpretation of mutually assured destruction the threat to launch a first strike deliberately would never be credible, deterrence could still be based on the fear of "things getting out of hand," on the fear that the crisis would go out of control and escalate to a general nuclear exchange (Schelling 1960, 1962b, 1966). It was unnecessary to rely on an incredible threat to launch a massive nuclear attack deliberately. Rather, a state could threaten to take steps that would increase the likelihood of uncontrolled escalation to an unlimited nuclear exchange. A state could make "threats that leave something to chance" (Schelling 1960). Credibility, then, was to be found in having a set of limited options, each of which, if exercised, would raise the risk of the crisis going out of control. Because exercising an option was not certain to trigger a general nuclear war, but only created the risk of it, the expected cost of exercising an option would be less than the expected cost of deliberately imposing the sanction of launching an unlimited attack. If, moreover, a state's stake in the crisis were high, the expected cost of escalating by exercising a limited option might be less than the expected cost of giving in to an adversary's demands. In that case, the threat to escalate would be credible. As Schelling put it, "a response that carries some risk of war can be plausible, even reasonable at a time when a final, ultimate decision to have a general war would be implausible or unreasonable" (1966, p. 98).

The set of limited options links the sanction of a massive nuclear attack to states' attempts to secure their interests. By being able to vary the risk of the crisis going out of control, the strategy that leaves something to chance offers a state a means of exerting coercive pressure on its adversary even in a condition of mutually assured destruction. Whether a state exercises a limited option in order to raise the risk or despite the greater risk, raising the risk of a general nuclear exchange increases the expected cost to its adversary of continuing the crisis. If this cost is greater than the cost of submitting, an adversary will quit the crisis. Of course, both states might take steps that would raise the risk. In the strategy that leaves something to chance, the crisis continues until one of the states finds the risk intolerable or until the crisis goes out of control and there is a general nuclear exchange.[15]

In this approach, limited options manipulate the risk of the crisis going out of control and escalating to an unlimited nuclear exchange. Accordingly, these options are not to be judged primarily in terms of their effects on the battlefield of a limited war. Relative military strength and superiority would seem to be unimportant. What matters would seem to be the ability and willingness to create risks. Limited options are to be judged by their effects on the risk of uncontrolled escalation to a general nuclear war (Schelling 1960, 1962b, 1966; Jervis 1979–80, 1984). In this way, the set of limited options, each of which carries a different risk of escalation, constitutes an array of risk.

The array of risk and a strategy based on threats that leave something to chance offered one means of coping with a situation in which mutually assured destruction was the technological state of affairs. Even if there were no advantage to striking first and no situation in which a state could rationally and deliberately launch what it knew to be an unlimited first strike, a state might still be able to use the sanction of a massive nuclear attack coercively to protect its interests by manipulating the risk that a crisis would go out of control and escalate to a general nuclear exchange. This is one of the ways that deterrence theory has addressed the problem of credibility.

There is also a second approach, in which deterrence is not based on an unlimited attack but on limited attacks or limited retaliation.[16] A state would no longer threaten the complete destruction of its adversary through a massive nuclear attack. Rather, a state would attempt to deter its adversary by threatening to extract a toll in pain and destruction that, although sufficiently large to outweigh any potential gains, would still be limited. Should this threat initially prove insufficiently credible to dissuade an adversary, then a state might try to make it more credible by actually carrying out a limited option and inflicting some punishment.

If limited options were to be used in this way, they had to satisfy two criteria. First, a state at least had to appear to be able to impose high enough costs on an adversary that it would rather back down than endure the punishment that could be inflicted. But, second, the options had to be sufficiently limited that even if they had been exercised, the adversary still would be left with something more to lose. That was the key to the credibility problem. If a state had been completely destroyed by an unlimited attack, so that it had nothing left to lose, it would have no incentive to limit its retaliation. If, however, a state had suffered a limited attack and was left with something more to lose, that state might be deterred from retaliating in order not to lose what was left. A state might, for example, threaten to destroy one of its adversary's cities in order to coerce that adversary into backing down during a severe confrontation in which vital national interests were at stake. If, during the course of that confrontation, that threat were carried out, then, despite the horrendous loss of a city, the adversary still would have much left that could be lost. Moreover, the fact that it had already lost one city might make the threat that it was about to lose another very credible. That, in turn, might convince it not to retaliate and to back down.

Clearly, whatever coercive pressure the exercise of a limited option creates in this approach arises only by increasing the credibility of the threat of future destruction. Coming to terms after a city has been destroyed does not rebuild the city or bring the dead back to life or alleviate the survivors' suffering. At most, it preserves what remains. "The hurting

[15] There are really two variants of this approach. In the first, states exercise a limited option in order to raise the risk of disaster. A crisis becomes a "competition in risk-taking" (Schelling 1966, p. 166) in which each state tries to demonstrate that its resolve, i.e., its willingness to run the risk that the crisis will go out of control, is greater than its adversary's resolve. In the second variant, the effect of exercising a limited option is to raise the risk, and it is this greater risk that actually exerts the coercive pressure, but the state exercises the option because it appears to further its ends in some other way. The greater risk is seen as an undesirable but unavoidable consequence of acting.

Historical evidence generally does not support the first variant. Leaders do not seem to take steps because they raise the risk of war (Snyder and Diesing 1977, p. 242; Trachtenberg 1985, p. 146). The second variant seems more viable: States do not act in order to increase the risk of war, but act in ways that do raise the risk, and this risk is the source of coercive pressure. This formulation, however, begs an important question: If states are not exercising limited options in order to raise the risk but because they appear to further their ends in some other way, what are these other ways, and what is the evidence for concluding that it is the greater risk and not these other ways that may coerce an adversary into submitting?

[16] For early studies of the strategy of limited retaliation, see Snyder (1961), Kaplan (1962), Kahn (1962), Knorr (1962), Schelling (1962a, 1965, 1966), and Halperin (1963).

does no good directly; it can only work indirectly. Coercion depends more on the threat of what is yet to come than on the damage already done" (Schelling 1966, p. 172). In sum, by exercising a limited option, a state attempts to demonstrate that its resolve is greater than that of its adversary, in the sense that it is more willing than is its adversary to inflict and endure future punishment in order to secure its ends.

In the strategy based on threats that leave something to chance, limited options raised the risk of the crisis going out of control and escalating to an unlimited nuclear exchange. The set of limited options thus constituted an array of risk. In the second approach to deterrence, limited options inflict limited amounts of punishment to make the threat of future punishment more credible. These options now form an array of punishment.[17]

Uncertainty and the struggle to control events play crucial roles in escalation and crisis bargaining. "The essence of the crisis is its unpredictability. The 'crisis' that involves no risk of things getting out of hand is no crisis.... It is the essence of a crisis that the participants are not fully in control of events" (Schelling 1966, p. 97). But "not [being] fully in control of events" has two interpretations, and the distinction between them is crucial to understanding the relations between the use or threatened use of force and states' attempts to secure their political objectives that underlie the strategies of leaving something to chance and of limited retaliation.

The first interpretation of the participants not being fully in control of events is that the participants do not have complete collective control. Even if the participants agree on a certain outcome and jointly act to effect this outcome, they cannot guarantee that this particular outcome will be realized. There is, to use Snyder and Diesing's phrase (1977, p. 210), some "autonomous risk" that some other outcome will eventuate.

Schelling (1966, pp. 99–105) offered a modified game of chess as an analogy for the strategy that leaves something to chance. This analogy also helps to clarify the first interpretation of events not being fully under control. To the three possible outcomes of the standard game of chess, win, lose, or draw, Schelling added a fourth, disaster, which is the analogue of a general nuclear exchange. If the game ends in disaster, each side will be worse off than if it had simply lost. The game may end in disaster in only one way: If a knight and queen of opposite colors cross the center line, then "the

[17] In an earlier essay (Powell 1985), these arrays were called the "spectrum of risk" and the "spectrum of violence." The word "spectrum" was a poor choice, for it connotes a continuum of limited options. That connotation was unintended and is inappropriate, for nuclear weapons may be very blunt, and there may be few limited options. For this reason, "array" is a better description of the set of limited options.

referee rolls a die. If an ace comes up the game is over and both sides are scored with a disaster, but if any other number comes up play goes on. If after the next move the queen and knight are still across the center line the [die is] rolled again, and so on" (Schelling 1966, p. 102). The addition of the referee and the die means that black and white are not in complete collective control of the game. Once a knight and queen of opposite colors cross the center line, the players can no longer guarantee that the game will not end in disaster. Although neither player would ever deliberately end the game in disaster, there is some chance of its ending that way. In moving a knight and queen across the center line, the players lose collective control of the outcome of the game. Their fate passes to the autonomous risk involved in the referee's throw of the die.

In the second interpretation of events not being fully under control, the participants are in complete collective control. If they agree on a particular outcome, the participants can effect any agreed outcome. Control is not something that can be lost. Events, however, are not fully under control, in the sense that no participant can control the actions and reactions of another.

The standard game of chess offers a good example of events not being fully under control in this second sense. If the players agree to a particular series of moves, then as long as this series is consistent with the rules of chess, the players, who collectively control all of the pieces, can effect this series of moves. If the players agree to a series of moves ending in white being checkmated, then the players can follow this series. But, of course, white does not want to be checkmated. White has no interest in following this series of moves and, not being under black's control, need not. This is the essence of the second interpretation of events not being fully under control.[18]

Failing to distinguish between these two interpretations can lead to apparently paradoxical conclusions about escalation in both nuclear and nonnuclear contexts. For example, concerning the crisis preceding World War I, the historian F. H. Hinsley wrote that if historians had gone as far as the evidence was trying to take them,

> they would have recognised that the dice had been set rolling for all the Powers before Russia mobilised – and not by any of the Powers but by a Balkan assassination. They would have seen that what makes some governments appear more responsible than others, or some governments more responsible at some stages and other governments more responsible at others, is not the fact that some governments were more instrumental than others in affecting the course of events. It is the fact that the positions

[18] The second interpretation of events not being fully under control is the permissive cause of war underlying Waltz's third image of international relations (Waltz 1959, p. 232).

> of the different governments varied with the course of events over which they had lost control. They would have recognised that, although it is theoretically possible to say that war would have been avoided if this or that government had acted otherwise, it was not possible for them to have acted otherwise. All the evidence goes to show that the beginning of the crisis which has been studied so largely with a view to discovering and distributing human responsibility, was one of those moments in history when events passed beyond men's control. [1963, p. 296]

According to Hinsley, people lost control of events after Sarajevo, and that resulted in war. But of the causes of World War II Hinsley said that "a war is always an alternative to some other course and is known to be so" (1963, p. 331). Juxtaposed, these comments seem paradoxical. How is it possible for one not to have control over war and peace and at the same time claim that war is always an alternative to some other course of action? If this other course is to have any meaning, it must be possible to follow it and thereby avoid war. But if one can avoid war by following another course, then one has control over war and peace.

The difficulty here is that the two interpretations of events not being fully under control have been conflated. The Balkan assassination that set the dice rolling was an example of the first interpretation. The shooting of the archduke was akin to the referee's throw of the die in Schelling's modified game of chess. It was an event beyond the collective control of the Great Powers. In that sense, the Great Powers lost control. But this did not lead directly to war. States acted and reacted to the actions of other states. The war, in Hinsley's account, was the result of the interaction of these reactions. The Great Powers did not lose collective control over whether or not there would be a world war. Had all of them agreed on a resolution of the crisis and acted jointly, they could have effected it. The war resulted from events not being fully under control in the second sense: No state could control the reactions of the other states.[19]

Distinguishing between these two interpretations resolves the apparent paradox. If a state can avoid war by submitting to its adversary, war is an alternative to some other course of action. But if a state does not believe that pursuing that course is in its best interest, given what is at stake in the confrontation, then because no state can control the actions of another, there will be war.

The distinction between the two interpretations of events not being fully under control is crucial to understanding the strategy of leaving something

[19] For nuclear deterrence theory, the July 1914 crisis is the archetypal crisis that goes out of control. See Trachtenberg (1989) for an historical reexamination of this thesis that casts doubt on this interpretation of the crisis.

to chance and the strategy of limited retaliation. If events are not fully under control only in the sense that no state can control the actions and reactions of other states, there is still risk and uncertainty. An adversary may escalate when it was expected to submit. But unless events are not fully under control in the first sense, the logic of the strategy that leaves something to chance will generally not be coherent. This dependence follows from the strategy's reliance on a sanction that would be so costly to impose that it would never be imposed deliberately. If the fear of suffering a sanction is to exert any coercive pressure, there must be some possibility of suffering it. If, therefore, no state would deliberately impose the sanction, there must be some other way for it to happen. Indeed, the something that is left to chance in the strategy of leaving something to chance is precisely that the sanction can arise in one of these other ways. Consequently, the states cannot be in complete collective control. There must be some autonomous risk underlying the strategy that leaves something to chance. The risk "must come from somewhere outside of the threatener's control" (Schelling 1960, p. 188). If the states always were in control in the first sense, there would be nothing to be left to chance. There would be no risk that could be manipulated in order to exert coercive pressure. There would no longer be an array of risk.

Two aspects of this dependence should be emphasized. First, the assumption that there is an autonomous risk is at once more and less demanding than it may initially seem. It may not appear to be very demanding at first because there are always events that are beyond collective control. There is always some autonomous risk of something. Some events are never fully under control in the first sense. But that is not sufficient for the strategy that leaves something to chance. A very specific event, the imposition of the sanction, must be imposed autonomously, and that is more demanding. At the height of the Cuban missile crisis, an American U-2 strayed into Soviet airspace. Soviet fighters were launched, and American interceptors, which because of the crisis and nuclear alert were armed with nuclear air-to-air missiles, were also scrambled. The interceptors did not make contact, and the U-2 found its way back to Alaska. Even so, President Kennedy was reported to have been concerned that Khrushchev might have thought that the U-2 was on a last-minute reconnaissance mission before an American nuclear attack (Sagan 1989a; 1989b, pp. 147–8). And the situation might have been much worse had one of the American fighters used a nuclear weapon.

The existence of an autonomous risk for events of this kind, serious and frightening as they may be, is not enough to ensure coherence for the logic of the strategy that leaves something to chance. The U-2 incident was akin to the Balkan assassination in Hinsley's account of the July 1914 crisis: It might have set the dice rolling. But just as the assassination did not lead

directly to war, but did so only indirectly through a series of actions and reactions, the U-2 incident could have led to the sanction's imposition only indirectly. It would have had to have been followed by a series of interacting decisions. But if, as is assumed in the strategy that leaves something to chance, the sanction would not have been imposed deliberately, the decision to launch a general nuclear war would not have been made, and the incident would not in the end have led to a general nuclear attack unless there had subsequently been a loss of collective control that had imposed the sanction directly. Describing the strategy of leaving something to chance as manipulating the risk that the crisis will go out of control obscures the fact that a very specific type of accident is required if there is to be a chance of losing control. The logic of this strategy generally depends on the autonomous risk of an event that will lead directly to the sanction without the participants having to make a series of decisions that ultimately will end in a deliberate decision to impose the sanction. The failure of, say, a computer chip would be required to launch a general nuclear attack directly. If there is no autonomous risk that the sanction will be imposed directly, there is nothing for the strategy that leaves something to chance to leave to chance.[20]

The fact that the sanction must be imposed directly makes the dependence on autonomous risks more demanding than it may initially appear. But another aspect of this dependence makes this requirement less demanding. Although it will be convenient to refer to the participants in a crisis as states, it is more reasonable to conceive of the participants as the leader of each state and the group of advisors who will be trying to deal with the crisis. This distinction is important because the source of the autonomous risk must lie beyond the participants' control; therefore, which sources are beyond the participants' collective control clearly depends on how one defines the participants. When, for example, the participants are taken to be a small group of advisors, risks that lie within the collective control of the state and its institutions and organizations, but beyond the control of the national command authorities, are still autonomous. For example, the risk that an order to carry out a limited option will result in a general nuclear attack because of organizational

[20] There is an exception to this strategy's general dependence on events not being fully under control in the first sense. That is, there is a way in which the sanction might be imposed without the states losing collective control. Suppose a state launches a limited attack deliberately, but because of poor attack assessment the adversary is absolutely convinced that it suffered an unlimited attack and then retaliates in kind. In retaliating, the adversary, believing itself to be launching an unlimited second strike, deliberately attacks and thereby intentionally launches what is actually a first strike. The problems of false alarms and their effects on escalation are examined in more detail in Chapters 5 and 6.

rigidities and routines is an autonomous risk.[21] Thus, this narrower description of the participants expands the scope of possible sources of autonomous risk, and that makes the logical dependence of the strategy that leaves something to chance on this risk empirically less demanding.

The second aspect of this strategy's dependence on autonomous risk that should be emphasized has to do with rationality. The credibility problem in the strategy that leaves something to chance arises because this strategy relies on a sanction that no rational actor would knowingly be the first to impose. As long as rationality is assumed, the logic of this strategy generally requires that it be physically possible for the states to lose collective control (e.g., because of technical failure). But if the rationality assumption is relaxed, there is another way in which the states may "lose" collective control: If there is some chance that under the stress of a crisis a state might act irrationally by imposing the sanction, then this possibility may exert coercive pressure during the confrontation. Recognizing this, a state might, at least in principle, pursue what has been called the strategy of the rationality of the irrational (Snyder 1961, pp. 24–7; Kahn 1965, pp. 57–8; Maxwell 1968) by trying to convince its adversary that it might act irrationally. In any case, allowing for irrationality does not fundamentally change the understanding of the credibility problem in the approach based on the strategy that leaves something to chance. If the stakes are high enough, then taking a step that leaves something to chance, when that something includes the possibility of an adversary acting irrationally, may still be rational, and so the threat to take the step may be credible. Indeed, models based on the assumption that a state believes that its adversary may act irrationally will be used in subsequent chapters to study the dynamics of strategies based on manipulation of risk and on limited retaliation.

The general dependence of the strategy of leaving something to chance on there being an autonomous risk of the sanction being imposed offers one means of assessing the empirical significance of this approach. Suppose that there is negligible risk from the narrow range of accidental or irrational acts that would impose the sanction directly. Then, although the array of risk might link the use or threatened use of force to states' political ends in principle, it would not seem to do so in practice. Although logically consistent, this approach would not seem adequate to account for the dynamics of escalation.[22]

[21] Sagan (1985, 1989a, 1989b) has described some of the accidents and problems of control that the United States has actually experienced during nuclear crises and alerts.

[22] Although not motivated by this issue, studies of the command and control systems of nuclear forces (Ball 1981, 1985–6; Bracken 1983; Blair 1985; Carter 1987), studies of previous accidents (Sagan 1985, 1989a, 1989b), and psychologically oriented studies of crises (Jervis 1977; Lebow 1981; Jervis *et al.* 1985) may shed some light on the size of this autonomous risk and, potentially, on the suitability of this description.

In sum, nuclear deterrence theory has approached the problem of credibility in two ways. It has generally linked force or the threat of it to states' political objectives after the nuclear revolution through arrays of risk and punishment. By raising the risk of unlimited destruction to an intolerably high level through the array of risk, or by posing too great a danger of limited but nevertheless terrible damage through the array of punishment, a state might be able to coerce its adversary into coming to terms and in that way be able to secure its interests. This, of course, is not to say that a state will exert coercive pressure in these ways. A state may not believe that doing so is in its best interest. The chance that an adversary will actually submit may seem too remote, and therefore the cost of pursuing these strategies may seem greater than the benefits. But whether or not a state actually uses or threatens to use force, these two approaches are solutions to the credibility problem facing nuclear deterrence theory, for they describe, at least in principle, the relation between the use or threatened use of force and states' attempts to achieve their ends within a stylized environment in which defense is impossible.

By limiting its focus to these two approaches, nuclear deterrence theory may seem much too narrow and entirely unrelated to many of the debates about counterforce strategies that have shaped American nuclear policy.[23] A more careful examination of the implicit assumptions that seem to underlie these strategies will show, however, that these approaches are more relevant than they may at first seem. Indeed, a better understanding of these approaches would seem to be a prerequisite to understanding these strategies.

In the counterforce strategy based on having escalation dominance,[24] for example, a state uses its counterforce capability and escalation dominance to force its adversary to bear the onus of escalation. Thus, it would seem that the closer a state can come to achieving escalation dominance at all levels, the more an adversary will have to bear the burden of escalation and the less likely it will be to escalate or to provoke a confrontation in the first place. Believing that to be the relation between force and states' political objectives after the nuclear revolution, a state may attempt to further its ends by trying to attain escalation dominance at as many levels as it can.

But if, because of relatively invulnerable strategic forces, a state has the ability to destroy its adversary, then even when faced with a military defeat

[23] For an overview of these debates, see Freedman (1989).

[24] A state has escalation dominance at a certain level of conflict because of its superior counterforce capabilities if that state's military capabilities are such that it can force its adversary to choose between accepting defeat at that level or escalating to another level of violence (Kahn 1965, p. 290).

at a given level of violence, that state does not have to accept defeat or escalate to a higher level of military conflict. The state may try to exert coercive pressure on its adversary through the arrays of risk and punishment. Exerting pressure in these ways does not, moreover, require significant counterforce capabilities. Accordingly, the assumption that escalation dominance will significantly enhance deterrence implicitly discounts the possibility that a state will turn to these other means of bringing coercive pressure to bear. That is, when facing defeat at a given level, a state will accept defeat and be deterred from turning to the arrays of risk and punishment. But is this implicit assumption well founded? What factors affect a state's decision whether or not to try to coerce an adversary in these ways? A better understanding of the two approaches to deterrence based on the arrays of risk and punishment and, especially, of the conditions in which a state will or will not turn to these coercive means will shed some light on these questions. In this way, a better appreciation of these two strategies will provide a deeper understanding of counterforce strategies.[25]

Limited options and the problem of credibility

Nuclear deterrence theory has linked force and states' political ends in two ways. The strategy that leaves something to chance works through an array of risk and ultimately appeals to the sanction of an unlimited nuclear attack or, more generally, to a sanction that no state would ever deliberately be the first to impose. The strategy of limited retaliation, however, never appeals to the possibility of an unlimited attack. The array of punishment is used to impose limited sanctions in order to make the threat of future destruction sufficiently credible that an adversary will be coerced into coming to terms. Although these two approaches seem quite different, they are at a general level fundamentally alike. Each attempts to solve the credibility problem in the same way. Each uses an array of limited options to bridge the gap between doing too much by launching an unlimited nuclear attack, as in the doctrine of massive retaliation, and doing too little by acquiescing. Because these options are limited, a state may be able to make the threat to use them more credible. The remainder of this chapter is devoted to bringing out the essential similarities underlying these approaches.

[25] Schelling (1965; 1966, pp. 190–204) has made a similar point. A counterforce contest as envisioned in McNamara's "no cities" doctrine, for example, would eventually confront the losing state with a choice between continuing to lose the counterforce struggle or turning to a strategy of limited retaliation. Accordingly, a better appreciation of this strategy and, especially, of the circumstances in which a state is likely to adopt it would also seem to be relevant to a deeper understanding of this counterforce doctrine.

A useful way to begin is to formalize the doctrine of massive retaliation and the criticisms made of it. The formalization furthers two ends. First, redescribing the doctrine of massive retaliation and its criticisms in different and more formal terms makes the fundamental similarities of the two approaches to the credibility problem easier to see. The second end is to build confidence in the formal tools that will be used in subsequent chapters to examine these two approaches in more detail. The doctrine of massive retaliation and its weaknesses are relatively straightforward. Formal analytic tools are not needed for adequate explication of the issues. If, however, the formal analysis corresponds well with a nonformal analysis of a given situation, like massive retaliation, in which one can be relatively confident of the nonformal analysis, then one may place more confidence in applying the formal analysis to more complicated situations in which, because of the greater complexity, a nonformal analysis would be much more problematic.

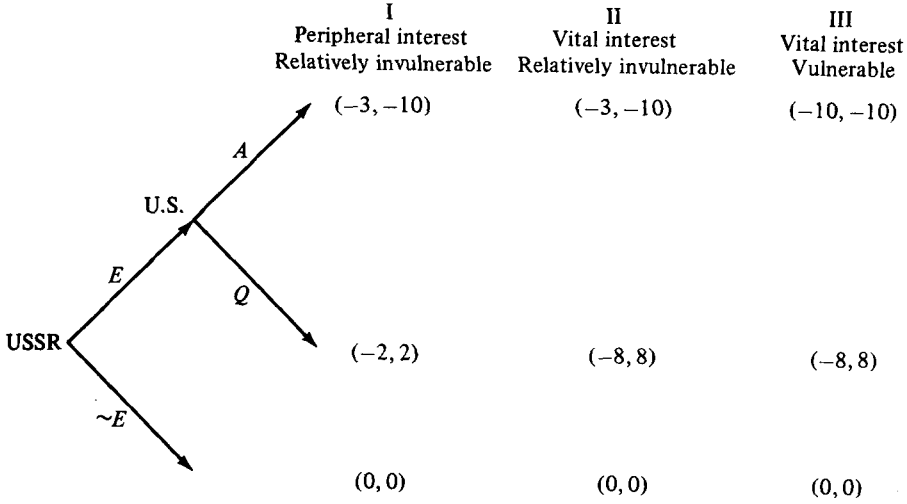
In its simplest form, the doctrine of massive retaliation relies on a threat to launch a massive nuclear attack in response to any challenge to any American interest, ranging from the most peripheral to the most vital. Recall further that the credibility and therefore the efficacy of this doctrine were initially criticized only when this doctrine was used as a means of protecting less important interests. As long as the United States was perceived to be relatively invulnerable to a Soviet nuclear attack, the threat to launch a massive nuclear attack in order to protect vital American interests like Western Europe seemed credible. But once the United States became vulnerable to a devastating Soviet retaliatory attack, the credibility of this threat, even if made only in the context of attempting to protect vital American interests, became problematic.

The game in Figure 2.1 illustrates the doctrine of massive retaliation. The game tree shows what the sequence of play is and what alternatives each state has when it must decide what to do. The Soviet Union begins the game by deciding whether or not to exploit a situation by challenging the status quo. These alternatives are denoted by E and $\sim E$, respectively. If the Soviet Union accepts the status quo by playing $\sim E$, the game ends. If the Soviet Union exploits an opportunity to challenge the status quo, then the United States must choose between two options.[26] It can launch a massive nuclear attack, A, or it can quit the confrontation, Q, by acquiescing to the Soviet challenge.

To complete the specification of the game, the payoffs must be defined. There are three different sets of payoffs, each corresponding to a different

[26] Although this simple version of the doctrine of massive retaliation may seem to be a caricature of what this doctrine actually was, it is the version of the doctrine that the critics seemed to have in mind. See, for example, Kaufmann (1956).

Figure 2.1. The credibility of massive retaliation.



situation in which the United States might try to rely on the doctrine of massive retaliation. Column I illustrates the situation in which the United States is relatively invulnerable and a peripheral American interest is involved. The United States is still assumed to be invulnerable in column II, but now a vital American interest is at issue. Finally, in column III, the United States is vulnerable to a Soviet nuclear counterattack, and a vital American interest is at stake.

To specify the payoffs in column I, where the United States is vulnerable and a peripheral interest is at risk, normalize the status quo payoffs to be (0, 0), where the first number is the American payoff and the second is the Soviet payoff. If the Soviet Union challenges the status quo in this situation and the United States replies with a massive nuclear attack, the Soviet Union will be completely destroyed. The United States will also suffer, but less so, because the United States is assumed to be relatively invulnerable. Let the payoffs to this outcome be, say, $(-3, -10)$.[27] If the Soviet Union challenges the status quo and the United States acquiesces, the Soviet Union improves its position compared with the status quo. The United States loses, but not much, for only a peripheral interest is assumed to be at stake. The payoffs corresponding to this outcome will be taken to be $(-2, 2)$.

Now consider the situation in which the United States remains relatively

[27] The specific numerical values of these payoffs are, of course, rather arbitrary. They are intended only to illustrate the differences between the three situations in a very simple way.

invulnerable, but a vital American interest is at risk. The status quo payoffs in column II are still (0, 0). The payoffs if there is a war are also the same: If the Soviet Union challenges the status quo and the United States attacks, they receive (−3, −10). The only payoffs that change from column I to II are those that obtain if the Soviet Union disputes the status quo and then the United States acquiesces. This change reflects the assumption that a vital American interest is at stake in column II. If the United States does not act in this situation and thereby lets the Soviet Union have its way, the United States will pay a high price. The payoffs depicting this condition are taken to be (−8, 8).

Finally, if a vital American interest is at stake and the United States is vulnerable to a devastating Soviet retaliatory attack, then the only difference between the payoffs in columns II and III is that the payoffs corresponding to a Soviet challenge followed by an American first strike now reflect the greater American vulnerability. As long as the United States was relatively invulnerable, its payoff to attacking and then having to endure Soviet retaliation was −3. Being more vulnerable, Soviet retaliation will impose higher costs and leave the United States with −10.

With the game thus defined, its equilibria may now be described. The doctrine of massive retaliation turns out to be a Nash equilibrium.[28] To see this, the states' strategies must be formally specified. The American strategy in the doctrine of massive retaliation is to attack the Soviet Union only if the Soviet Union challenges the status quo (i.e., the United States plays *A* in the game in Figure 2.1). The Soviet strategy is not to challenge the status quo. To show that this combination of strategies constitutes a Nash equilibrium, it need only be shown that no state has an incentive to deviate from its strategy given its adversary's strategy. Consider the situation in column I, in which the United States is relatively invulnerable, and only a peripheral American interest is at risk. Clearly, the Soviet Union has no incentive to alter its strategy of not disputing the status quo. Given the American strategy of responding to a challenge with an unlimited nuclear attack, if the Soviet Union deviates from its strategy by challenging the status quo its payoff will be −10, whereas following the strategy of accepting the status quo will assure the Soviet Union of 0. The United States also has no incentive to deviate from the doctrine of massive retaliation. Given that the Soviet Union is not challenging the status quo, the American payoff is always 0 regardless of what it would do if challenged. Because neither state has any incentive to deviate from its strategy given the strategy of the other state, the doctrine of massive

[28] See the Appendix following Chapter 8 for an introduction to the game-theoretic concepts, such as Nash equilibria, that are used in this and subsequent chapters.

retaliation in the game represented by the payoffs in column I is a Nash equilibrium. Similar arguments show that this doctrine is also a Nash equilibrium in the situations illustrated by columns II and III.

Although the doctrine of massive retaliation is a Nash equilibrium, there is a troubling feature about this equilibrium, and this goes to the heart of the criticisms of this doctrine. Suppose that the Soviet Union does, for some reason, challenge the status quo. The United States will then have to decide between launching a massive nuclear attack, which will bring a payoff of −3, or acquiescing, which, with only a peripheral interest involved, will yield −2. Assuming that states act in ways that they believe to be in their best interest, the United States will choose −2 when confronted with a choice between −2 and −3. The United States will not launch a massive nuclear attack, for the cost of carrying out its threat would be greater than the cost of not doing so. The Soviet Union, understanding this, will find the doctrine of massive retaliation incredible and will not be deterred by it. That is, in effect, the criticism William Kaufmann made in 1956: that even if the United States were relatively invulnerable, the doctrine of massive retaliation could not protect less important American interests.

This criticism can be stated more formally, and doing so helps to build confidence in the ability of formal methods to contribute to the analysis of more complicated situations. Although the doctrine of massive retaliation is a Nash equilibrium, it is not a sequential equilibrium.[29] In a sequential equilibrium, agents are required to act in their best interest everywhere in the game tree given their beliefs and the strategies of the other agents. Thus, when confronted with a choice between −2 and −3, the United States must choose the former. When the United States is relatively invulnerable, but only a peripheral interest at stake, there is a unique sequential equilibrium. In it, the Soviet Union challenges the status quo, and the United States acquiesces.[30] The doctrine of massive retaliation founders on the credibility problem in the sense that it is not a sequential equilibrium.

Viewing the credibility problem from the perspective of sequential equilibria also accounts for the other criticisms of the doctrine of massive

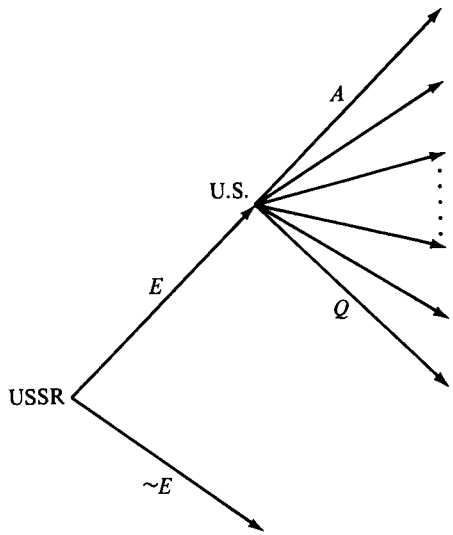
[29] It would suffice at this point to look only to the more appealing and less demanding notion of subgame perfection in order to eliminate this Nash equilibrium. Subgame perfection, however, will be insufficient in subsequent chapters, where incomplete-information games will be studied. For a discussion of the relation between subgame perfection and sequential equilibria, see Kreps and Wilson (1982b) or the Appendix following Chapter 8.

[30] Neither the Soviet Union nor the United States has any incentive to deviate from its strategy given the other's strategy. Challenging the status quo brings the Soviet Union 3, whereas forgoing a challenge yields 0. Similarly, the United States will lower its payoff from −2 to −3 if it deviates from acquiescing by attacking. This combination of strategies thus forms a Nash equilibrium. Moreover, no state has any incentive to deviate from its strategy anywhere in the game tree given its beliefs.

retaliation. That doctrine seemed to be credible as long as the United States was relatively invulnerable and a vital American interest was at risk. That situation corresponds to the payoffs in column II. With these payoffs, the game also has a unique sequential equilibrium. But this time, the United States will attack if challenged, for this brings -3, whereas surrendering a vital interest leaves the United States with -8. Given this American strategy, the Soviet Union's best response is not to dispute the status quo. When the cost of acquiescing is so high, the threat to retaliate massively rather than submit is credible, and the doctrine of massive retaliation is effective in protecting vital interests. But, of course, once the United States became vulnerable to a devastating Soviet retaliatory attack, this doctrine seemed incredible even with vital interests at stake. Again, insisting that equilibria be not only Nash but also sequential accounts for this. With the payoffs of column III, the Soviet Union challenges the status quo, and the United States acquiesces, preferring to suffer the large but limited loss rather than the still larger loss that a Soviet retaliation would cause.

All of this indicates that focusing on sequential equilibria in games modeling not only massive retaliation but also crises based on the arrays of risk and punishment will do much to provide an understanding of the credibility problem. Strategies that are part of a sequential equilibrium cannot rely on threats that are inherently incredible, because carrying them out would be more costly than not doing so. The close correspondence

Figure 2.2. The array of limited options.



between the formal and less formal critiques of the doctrine of massive retaliation suggests that these more formal tools will be useful in analyzing the more complicated approaches to the credibility problem based on the arrays of risk and punishment.

In addition to building confidence in the game-theoretic analysis of the credibility problem, the game in Figure 2.1 may be used to bring out the fundamental similarities underlying deterrence theory's two approaches to this problem. Intuitively, this problem arises because the two options in Figure 2.1 of launching a massive nuclear attack and of quitting are too far apart. One option does too much, the other too little. Credibility, then, would seem to require the creation of an array of limited options in which the distance between any two adjacent options in this array is less than the distance between the option of launching a massive attack and that of doing nothing. This array, in effect, bridges the gap between doing too much and too little. Figure 2.2 illustrates this array by filling the gap between these two extremes with a number of limited options. There are, however, two different ways to measure the distance between adjacent options, and each of these ways corresponds to one of the approaches deterrence theory has taken to the credibility problem.

The first way of measuring the distance is in terms of the probability that the crisis will end in an unlimited exchange. Measured in this way, the distance between launching an unlimited attack, which would end the crisis in a general nuclear exchange with probability 1, and quitting, which would ensure that there would be no exchange, is 1. No two options could be farther apart. The array of limited options fills this gap by making it possible to create intermediate levels of risk. Associated with each limited option is the level of risk that the exercise of this option will generate. Indeed, what distinguishes any option from any other is that they generate different levels of risk. In this way the set of limited options constitutes the array of risk that underlies the strategy that leaves something to chance.

Damage is the second way of measuring the distance between the extremes of launching a massive nuclear attack and submitting. The former inflicts complete destruction, and the latter inflicts none. Again, these options are very far apart, and, as before, limited options are used to bridge this gap. But in this case what defines and distinguishes one limited option from another is the amount of punishment it will impose if exercised. When measured in terms of damage, the set of limited options now forms an array of punishment.

So, when viewed from this more general perspective, the two seemingly disparate approaches to linking force or the threat of it to states' political ends appear to be fundamentally alike. Each addresses the credibility problem in essentially the same way, by creating an array of limited options

to bridge the gap between doing too much and too little. This is not to say that many options can be created such that there will be fine gradations in the levels of risk or damage separating these options. Nuclear weapons may be very blunt. It is only to say that what distinguishes these approaches is not the general role or relative importance of limited options but the particular way that this gap is measured.

The rise of strategic air power and the development of intercontinental ballistic missiles and atomic and then thermonuclear weapons separated the ability to defend from the ability to punish. These developments culminated in the nuclear revolution, in which mutually assured destruction became the technological state of affairs. The separation of these capabilities and the impossibility of defense undercut the classical logic of war.

After the nuclear revolution, a state's ability to impose costs that would exceed an adversary's gains was no longer at issue. But given that defense was impossible, could a state make the threat to use its punitive capabilities sufficiently credible? What was the relation between force or the threat of it and states' efforts to secure their ends after the nuclear revolution? That was the credibility problem.

Nuclear deterrence theory has generally approached this problem in two ways. Force or the threat of it is linked to states' political objectives through either an array of risk or an array of punishment. Although these two approaches initially appear quite different, they are, at a more general level, essentially alike. Each tries to solve the credibility problem in the same way.

But how far does this fundamental similarity extend? The following chapters examine this in two ways. The first is a more detailed study of the dynamics of crisis bargaining and escalation when force is related to political ends through the arrays of risk and punishment. The second is to elaborate the relation between these two approaches. Although both approaches are, at a high level of generality, attempts to solve the credibility problem, they are primarily concerned with different issues at somewhat lower levels of generalization. But these lower-level issues are connected. An analysis of the strategy that leaves something to chance leads naturally to questions about crisis stability and first-strike advantages, and these in turn raise questions about the role of limited sanctions and the strategy of limited retaliation. The following chapters trace these connections and in this way further clarify the relation between these two approaches to the credibility problem.