THE STRATEGY OF CONFLICT

THOMAS C. SCHELLING

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
LONDON, ENGLAND

Fifth printing, 1976

Printed in the United States of America
Library of Congress Catalog Card Number 60–11560
ISBN 0–674–84030–5

PREFACE

This is a series of closely interrelated essays in a field that is variously described as "theory of bargaining," "theory of conflict," or "theory of strategy." Strictly speaking, the subject falls within the *theory of games*, but within the part of game theory in which the least satisfactory progress has been made, the situations in which there is common interest as well as conflict between adversaries: negotiations, war and threats of war, criminal deterrence, tacit bargaining, extortion. The philosophy of the book is that in the strategy of conflict there are enlightening similarities between, say, maneuvering in limited war and jockeying in a traffic jam, between deterring the Russians and deterring one's own children, or between the modern balance of terror and the ancient institution of hostages.

The analysis is neither difficult nor so dependent on mathematics or analytical apparatus as to be inaccessible to any serious reader. A few chapters call for a rudimentary acquaintance with some concepts from game theory.

The first chapter (in a longer version) was originally presented in early 1959 to a conference on "International Relations in the Mid-twentieth Century," at Northwestern University; although the occasion and the audience were somewhat specialized, the paper represents the motivation and theme of the entire book. Chapters 2 and 3 were originally independent articles on "bargaining." It was evident, after they were written, that they belonged to the same field as the *theory of games*; an effort to fit them into the framework of game theory, stretching the framework if necessary, resulted in Chapters 4 through 6 and Appendices B and C. Chapters 7 through 10, and Appendix A, are extensions of the same method to particular problems in international strategy.

Appendices B and C will be of interest mainly to readers conversant with bargaining theory or game theory. Appendix A has been treated as an appendix only because its extended preoccupa-

tion with a particular policy problem is in some contrast to the style of Chapter 4, where it would otherwise belong.

The essays are a mixture of "pure" and "applied" research. To some extent the two can be separated, as in the companion pieces in Part IV. In my own thinking they have never been separate. Motivation for the purer theory came almost exclusively from preoccupation with (and fascination with) "applied" problems; and the clarification of theoretical ideas was absolutely dependent on an identification of live examples. For reasons inherent either in the subject or in the author, the interaction of the two levels of theory has been continuous and intense.

Three people have been most influential, probably more than they realize, in my continuing this work. They are Kenneth E. Boulding, Bernard F. Haley, and Charles J. Hitch. Numerous associates, particularly at The RAND Corporation, have lent me ideas and stimulated my own; I refer especially to Bernard Brodie, Daniel Ellsberg, Malcolm W. Hoag, Herman Kahn, William W. Kaufmann, and Albert J. Wohlstetter. William W. Taylor gave me valuable editorial help. And I owe a special word of appreciation to R. Duncan Luce and Howard Raiffa, whose *Games and Decisions* has been of immeasurable help; if I have often focused critical remarks on the book, it is only because the inevitable lot of a definitive survey is to serve as a definitive target.

During the year before this book went to press I was uniquely located to receive stimulation, provocation, advice, comment, disagreement, encouragement, and education. I spent the year with The RAND Corporation, in Santa Monica. As a collection of people, RAND is superb, and I have mentioned above only the few whose intellectual impact on me was powerful and persistent; many others, truly too numerous to list here, have as individuals affected the final shape of this book. But RAND is more than a collection of people; it is a social organism characterized by intellect, imagination, and good humor. RAND is not responsible for the shapes my ideas have taken — the "views herein expressed" — but I hope it will, as a corporation, take satisfaction from its responsibility for some of the ideas' taking any shape at all.

For readers who have come across some of the chapters before, the following may be of convenience. Chapter 2 appeared with the same title in *The American Economic Review*, Vol. XLVI No. 3, June 1956. Chapter 3 appeared with the same title in *The Journal of Conflict Resolution*, Vol. I No. 1, March 1957. Chapters 4, 5, and 6 are a somewhat rearranged version of "The Strategy of Conflict," *The Journal of Conflict Resolution*, Vol. II No. 3, September 1958, with parts eliminated that overlapped other chapters. Appendix B appeared, with the same title, in *The Review of Economics and Statistics*, Vol. XLI No. 3, August 1959. A longer version of Chapter 10, with the same title, is contained in Klaus Knorr (ed.), *NATO and American Security*, (Princeton: Princeton University Press, 1959). The several publishers have kindly allowed me to reprint these papers here, with modifications to make an integrated book.

THOMAS C. SCHELLING

Cambridge, Massachusetts

CONTENTS

PART I

ELEMENTS OF A THEORY OF STRATEGY

1

THE RETARDED SCIENCE OF INTERNATIONAL STRATEGY

Among diverse theories of conflict — corresponding to the diverse meanings of the word "conflict" — a main dividing line is between those that treat conflict as a pathological state and seek its causes and treatment, and those that take conflict for granted and study the behavior associated with it. Among the latter there is a further division between those that examine the participants in a conflict in all their complexity — with regard to both "rational" and "irrational" behavior, conscious and unconscious, and to motivations as well as to calculations — and those that focus on the more rational, conscious, artful kind of behavior. Crudely speaking, the latter treat conflict as a kind of contest, in which the participants are trying to "win." A study of conscious, intelligent, sophisticated conflict behavior — of successful behavior — is like a search for rules of "correct" behavior in a contest-winning sense.

We can call this field of study the *strategy* of conflict.[1] We can be interested in it for at least three reasons. We may be involved in a conflict ourselves; we all are, in fact, participants in international conflict, and we want to "win" in some proper sense. We may wish to understand how participants actually do conduct themselves in conflict situations; an understanding of "correct" play may give us a bench mark for the study of actual behavior.

[1] The term "strategy" is taken, here, from the *theory of games*, which distinguishes games of skill, games of chance, and games of strategy, the latter being those in which the best course of action for each player depends on what the other players do. The term is intended to focus on the interdependence of the adversaries' decisions and on their expectations about each other's behavior. This is not the military usage.

We may wish to control or influence the behavior of others in conflict, and we want, therefore, to know how the variables that are subject to our control can affect their behavior.

If we confine our study to the theory of strategy, we seriously restrict ourselves by the assumption of rational behavior — not just of intelligent behavior, but of behavior motivated by a conscious calculation of advantages, a calculation that in turn is based on an explicit and internally consistent value system. We thus limit the applicability of any results we reach. If our interest is the study of actual behavior, the results we reach under this constraint may prove to be either a good approximation of reality or a caricature. Any abstraction runs a risk of this sort, and we have to be prepared to use judgment with any results we reach.

The advantage of cultivating the area of "strategy" for theoretical development is not that, of all possible approaches, it is the one that evidently stays closest to the truth, but that the assumption of rational behavior is a productive one. It gives a grip on the subject that is peculiarly conducive to the development of theory. It permits us to identify our own analytical processes with those of the hypothetical participants in a conflict; and by demanding certain kinds of consistency in the behavior of our hypothetical participants, we can examine alternative courses of behavior according to whether or not they meet those standards of consistency. The premise of "rational behavior" is a potent one for the production of theory. Whether the resulting theory provides good or poor insight into actual behavior is, I repeat, a matter for subsequent judgment.

But, in taking conflict for granted, and working with an image of participants who try to "win," a theory of strategy does not deny that there are common as well as conflicting interests among the participants. In fact, the richness of the subject arises from the fact that, in international affairs, there is mutual dependence as well as opposition. Pure conflict, in which the interests of two antagonists are completely opposed, is a special case; it would arise in a war of complete extermination, otherwise not even in war. For this reason, "winning" in a conflict does not have a strictly competitive meaning; it is not winning relative to one's adversary. It means gaining relative to one's own value system;

and this may be done by bargaining, by mutual accommodation, and by the avoidance of mutually damaging behavior. If war to the finish has become inevitable, there is nothing left but pure conflict; but if there is any possibility of avoiding a mutually damaging war, of conducting warfare in a way that minimizes damage, or of coercing an adversary by threatening war rather than waging it, the possibility of mutual accommodation is as important and dramatic as the element of conflict. Concepts like deterrence, limited war, and disarmament, as well as negotiation, are concerned with the common interest and mutual dependence that can exist between participants in a conflict.

Thus, strategy — in the sense in which I am using it here — is not concerned with the efficient *application* of force but with the *exploitation of potential force*. It is concerned not just with enemies who dislike each other but with partners who distrust or disagree with each other. It is concerned not just with the division of gains and losses between two claimants but with the possibility that particular outcomes are worse (better) for *both* claimants than certain other outcomes. In the terminology of game theory, most interesting international conflicts are not "constant-sum games" but "variable-sum games": the sum of the gains of the participants involved is not fixed so that more for one inexorably means less for the other. There is a common interest in reaching outcomes that are mutually advantageous.

To study the strategy of conflict is to take the view that most conflict situations are essentially *bargaining* situations. They are situations in which the ability of one participant to gain his ends is dependent to an important degree on the choices or decisions that the other participant will make. The bargaining may be explicit, as when one offers a concession; or it may be by tacit maneuver, as when one occupies or evacuates strategic territory. It may, as in the ordinary haggling of the market-place, take the *status quo* as its zero point and seek arrangements that yield positive gains to both sides; or it may involve threats of damage, including mutual damage, as in a strike, boycott, or price war, or in extortion.

Viewing conflict behavior as a bargaining process is useful in keeping us from becoming exclusively preoccupied either with the

conflict or with the common interest. To characterize the maneuvers and actions of limited war as a bargaining process is to emphasize that, in addition to the divergence of interest over the variables in dispute, there is a powerful common interest in reaching an outcome that is not enormously destructive of values to both sides. A "successful" employees' strike is not one that destroys the employer financially, it may even be one that never takes place. Something similar can be true of war.

The idea of "deterrence" has had an evolution that is instructive for our purpose. It is a dozen years since deterrence was articulated as the keystone of our national strategy, and during those years the concept has been refined and improved. We have learned that a threat has to be credible to be efficacious, and that its credibility may depend on the costs and risks associated with fulfillment for the party making the threat. We have developed the idea of making a threat credible by getting ourselves committed to its fulfillment, through the stretching of a "trip wire" across the enemy's path of advance, or by making fulfillment a matter of national honor and prestige — as in the case, say, of the Formosa Resolution. We have recognized that a readiness to fight limited war in particular areas may detract from the threat of massive retaliation, by preserving the choice of a lesser evil if the contingency arises. We have considered the possibility that a retaliatory threat may be more credible if the means of carrying it out and the responsibility for retaliation are placed in the hands of those whose resolution is strongest, as in recent suggestions for "nuclear sharing." We have observed that the rationality of the adversary is pertinent to the efficacy of a threat, and that madmen, like small children, can often not be controlled by threats. We have recognized that the efficacy of the threat may depend on what alternatives are available to the potential enemy, who, if he is not to react like a trapped lion, must be left some tolerable recourse. We have come to realize that a threat of all-out retaliation gives the enemy every incentive, in the event he should choose not to heed the threat, to initiate his transgression with an all-out strike at us; it eliminates lesser courses of action and forces him to choose between extremes. We have learned that the

threat of massive destruction may deter an enemy only if there is a corresponding implicit promise of nondestruction in the event he complies, so that we must consider whether too great a capacity to strike him by surprise may induce him to strike first to avoid being disarmed by a first strike from us. And recently, in connection with the so-called "measures to safeguard against surprise attack," we have begun to consider the possibility of improving mutual deterrence through arms control.

What is impressive is not how complicated the idea of deterrence has become, and how carefully it has been refined and developed, but how slow the process has been, how vague the concepts still are, and how inelegant the current theory of deterrence is. This is not said to depreciate the efforts of people who have struggled with the deterrence concept over the last dozen years. On strategic matters of which deterrence is an example, those who have tried to devise policies to meet urgent problems have had little or no help from an already existing body of theory, but have had to create their own as they went along. There is no scientific literature on deterrence that begins to compare with, say, the literature on inflation, Asiatic flu, elementary-school reading, or smog.

Furthermore, those who have grappled with ideas like deterrence, being motivated largely by immediate problems, have not primarily been concerned with the cumulative process of developing a theoretical structure. This seems to be true not only of policy-makers and journalists but of the more scholarly as well. Whether it reflects the scholars' interests or that of the editors, the literature on deterrence and related concepts has been mainly preoccupied with solving immediate problems rather than with a methodology for dealing with problems.[2] We do not even have a

[2] There are some excellent examples to the contrary, like C. W. Sherwin, "Securing Peace Through Military Technology," *Bulletin of the Atomic Scientists*, 12:159–164 (May 1956). And Sherwin's reference there to a paper by Warren Amster reminds us that when theory is stimulated by military problems, as so much of it currently is, it may not receive open publication. There are undoubtedly, also, serious editorial obstacles; journals in international affairs appeal to a dominantly nontheoretical audience, and articles with high theoretical content must often be purged of it and focused on immediate problems. The recent devotion of an entire issue of *Conflict Resolution* to Anatol Rapoport's magnificent essay on "Lewis F. Richardson's Mathematical

decent terminology; occasional terms like "active" and "passive" deterrence do not begin to fill the need.

How do we account for this lack of theoretical development? I think one significant fact is that the military services, in contrast to almost any other sizable and respectable profession, have no identifiable academic counterpart. Those who make policy in the fields of economics, medicine, public health, soil conservation, education, or criminal law, can readily identify their scholarly counterpart in the academic world. (In economics the number of trained people who are doing research and writing books compares well with the number engaged in economic policy or administration.) But where is the academic counterpart of the military profession?

It is not — on any great scale — in the service academies; these are undergraduate schools, devoted mainly to teaching rather than to research. Not — or not yet on any great scale — in the war colleges and other nontechnical advanced educational institutions within the military services; these have not yet developed the permanent faculty, the research orientation, and the value system required for sustained and systematic theoretical development.

Within the universities, military strategy in this country has been the preoccupation of a small number of historians and political scientists, supported on a scale that suggests that deterring the Russians from a conquest of Europe is about as important as enforcing the antitrust laws. This is said not to disparage the accomplishments, but to emphasize that within the universities there has usually been no directly identifiable department or line of inquiry that can be associated with the military professions and the role of force in foreign relations. (ROTC programs have recently become a limited exception to this point, at least to the extent that they induce the organization of pertinent courses in history and political science.) The defense-studies programs and institutes now found on a number of campuses, and the attention given to international security problems by the foundations, are a novel and significant development. New quasi-governmental

Theory of War" (vol. I, No. 3, September 1957) is a heartening sign in the other direction.

research institutions like The RAND Corporation and the Institute for Defense Analysis are importantly helping to fill the need but, for our purpose, can be cited as evidence of the need.

One may ask whether the military services themselves might not be able to produce a growing body of theory to illuminate ideas like deterrence or limited war. After all, theory does not have to be developed solely by specialists isolated in universities. If the military services are intellectually prepared to make effective *use* of military force, it might seem that they are equipped to theorize about it. But here a useful distinction can be made between the *application* of force and the *threat* of force. Deterrence is concerned with the exploitation of potential force. It is concerned with persuading a potential enemy that he should in his own interest avoid certain courses of activity. There is an important difference between the intellectual skills required for carrying out a military mission and for using *potential* military capability to pursue a nation's objectives. A theory of deterrence would be, in effect, a theory of the skillful *nonuse* of military forces, and for this purpose deterrence requires something broader than military skills. The military professions may have these broader skills, but they do not automatically have them as a result of meeting their primary responsibilities, and those primary responsibilities place full-time demands on their time.[3]

A new kind of inquiry that gave promise, fifteen years ago, of leading to such a theory of strategy is *game theory*. Game theory is concerned with situations — games of "strategy," in contrast to games of skill or games of chance — in which the best course of action for each participant depends on what he expects the

[3] The lack of a vigorous intellectual tradition in the field of military strategy is forcefully discussed by Bernard Brodie in the first chapters of his *Strategy in the Missile Age* (Princeton, 1959). Pertinent also is Colonel Joseph I. Greene's foreword to the Modern Library edition of Clausewitz, *On War* (New York, 1943): "During most of the years between the great wars, the two highest schools of our Army were limited to a single course of some ten months' duration for all officers selected to attend them. . . . There could be no time at either place for study of the long development of military thought and theory. . . . If ever more extensive periods of higher training become possible in our Army — periods of two or three years' duration — the greatest of the military thinkers would surely deserve a course of study in themselves" (pp. xi–xii).

other participants to do. A deterrent threat meets this definition nicely; it works only because of what the other player expects us to do in response to his choice of moves, and we can afford to make the threat only because we expect it to have an influence on his choice. But in international strategy the promise of game theory is so far unfulfilled. Game theory has been extremely helpful in the formulation of problems and the clarification of concepts, but its greatest successes have been in other fields. It has, on the whole, been pitched at a level of abstraction where it has made little contact with the elements of a problem like deterrence.[4]

The idea of deterrence figures so prominently in some areas of conflict other than international affairs that one might have supposed the existence of a well-cultivated theory already available to be exploited for international applications. Deterrence has been an important concept in criminal law for a long time. Legislators, jurists, lawyers, and legal scholars might be supposed to have subjected the concept to rigorous and systematic scrutiny for many generations. To be sure, deterrence is not the sole consideration involved in criminal law, nor even necessarily the most important; still, it has figured prominently enough for one to suppose the existence of a theory that would take into account the kinds and sizes of penalties available to be imposed on a convicted criminal, the potential criminal's value system, the profitability of crime, the law-enforcement system's ability to apprehend criminals and to get them convicted, the criminal's awareness of the law and of the probability of apprehension and conviction, the extent to which different types of crime are motivated by rational calculation, the resoluteness of society to be neither niggardly nor soft-hearted in the expensive and disagreeable application of the penalty and how well this reso-

[4] Jessie Bernard, writing on "The Theory of Games as a Modern Sociology of Conflict," gives a somewhat similar appraisal but adds that "we may expect that the mathematics required to make a fruitful application of the theory of games to sociological phenomena will emerge in the not-too-distant future" (*The American Journal of Sociology*, 59:418, March 1954). My own view is that the present deficiencies are not in the mathematics, and that the theory of strategy has suffered from too great a willingness of social scientists to treat the subject as though it were, or should be, solely a branch of mathematics.

luteness (or lack of it) is known to the criminal, the likelihood of mistakes in the system, the possibilities for third parties to exploit the system for personal gain, the role of communication between organized society and the criminal, the organization of criminals to defeat the system, and so on.

It is not only criminals, however, but our own children that have to be deterred. Some aspects of deterrence stand out vividly in child discipline: the importance of rationality and self-discipline on the part of the person to be deterred, of his ability to comprehend the threat if he hears it and to hear it through the din and noise, of the threatener's determination to fulfill the threat if need be — and, more important, of the threatened party's conviction that the threat will be carried out. Clearer perhaps in child discipline than in criminal deterrence is the important possibility that the threatened punishment will hurt the threatener as much as it will the one threatened, perhaps more. There is an analogy between a parent's threat to a child and the threat that a wealthy paternalistic nation makes to the weak and disorganized government of a poor nation in, say, extending foreign aid and demanding "sound" economic policies or cooperative military policies in return.

And the analogy reminds us that, even in international affairs, deterrence is as relevant to relations between friends as between potential enemies. (The threat to withdraw to a "peripheral strategy" if France failed to ratify the European Defense Community Treaty was subject to many of the same disabilities as a threat of retaliation.) The deterrence concept requires that there be both conflict and common interest between the parties involved; it is as inapplicable to a situation of pure and complete antagonism of interest as it is to the case of pure and complete common interest. Between these extremes, deterring an ally and deterring an enemy differ only by degrees, and in fact we may have to develop a more coherent theory before we can even say in a meaningful way whether we have more in common with Russia or with Greece, relative to the conflicts between us.[5]

[5] It may be important to emphasize that, in referring to a "common interest," I do not mean that they must have what is usually referred to as a similarity in their value systems. They may just be in the same boat together:

The deterrence idea also crops up casually in everyday affairs. Automobile drivers have an evident common interest in avoiding collision and a conflict of interest over who shall go first and who shall slam on his brakes and let the other through. Collision being about as mutual as anything can be, and often the only thing that one can threaten, the maneuvers by which one conveys a threat of mutual damage to another driver aggressing on one's right of way are an instructive example of the kind of threat that is conveyed not by words but by actions, and of the threat in which the pledge to fulfill is made not by verbal announcement but by losing the power to do otherwise.

Finally, there is the important area of the underworld. Gang war and international war have a lot in common. Nations and outlaws both lack enforceable legal systems to help them govern their affairs. Both engage in the ultimate in violence. Both have an interest in avoiding violence, but the threat of violence is continually on call. It is interesting that racketeers, as well as gangs of delinquents, engage in limited war, disarmament and disengagement, surprise attack, retaliation and threat of retaliation; they worry about "appeasement" and loss of face; and they make alliances and agreements with the same disability that nations are subject to — the inability to appeal to higher authority in the interest of contract enforcement.

There are consequently a number of other areas available for study that may yield insight into the one that concerns us, the international area. Often a principle that in our own field of interest is hidden in a mass of detail, or has too complicated a structure, or that we cannot see because of a predisposition, is easier to perceive in another field where it enjoys simplicity and vividness or where we are not blinded by our predispositions. It may be easier to articulate the peculiar difficulty of constraining

---

they may even be there only because one of them perceived it a strategic advantage to get in that position — to couple their interests in not tipping the boat. If being overturned together in the same boat is a potential outcome, given the array of alternatives available to both parties, they have a "common interest" in the sense intended in the text. "Potential common interest" might seem more descriptive. Deterrence, for example, is concerned with coupling one's own course of action with the other's course of action in a way that exploits that potential common interest.

a Mossadeq by the use of threats when one is fresh from a vain attempt at using threats to keep a small child from hurting a dog or a small dog from hurting a child.

None of these other areas of conflict seems to have been mastered by a well-developed theory that can, with modification, be used in the analysis of international affairs. Sociologists, including those who study criminal behavior in underworld conflict, have not traditionally been much concerned with what we would call the *strategy* of conflict. Nor does the literature on law and criminology reveal an appreciable body of explicit theory on the subject. I cannot confidently assert that there are no handbooks, textbooks, or original works on the pure theory of blackmail circulating in the underworld; but certainly no expurgated version, showing how to use extortion and how to resist it, has shown up as "New Ways in Child Guidance," in spite of the demand for it.[6]

What would "theory" in this field of strategy consist of? What questions would it try to answer? What ideas would it try to unify, clarify, or communicate more effectively? To begin with, it should define the essentials of the situation and of the behavior in question. Deterrence — to continue with deterrence as a typical strategic concept — is concerned with influencing the choices that another party will make, and doing it by influencing his expectations of how we will behave. It involves confronting him with evidence for believing that our behavior will be determined by his behavior.

But what configuration of value systems for the two participants — of the "payoffs," in the language of game theory — makes a deterrent threat credible? How do we measure the mixture of conflict and common interest required to generate a "deterrence" situation? What communication is required, and what means of authenticating the evidence communicated? What kind of "rationality" is required of the party to be deterred — a knowledge of his own value system, an ability to perceive alternatives

[6] Progress is being made. Daniel Ellsberg included a lecture on "The Theory and Practice of Blackmail," and one on "The Political Uses of Madness," in his series on "The Art of Coercion," sponsored by the Lowell Institute, Boston, March 1959.

and to calculate with probabilities, an ability to demonstrate (or an inability to conceal) his own rationality?

What is the need for trust, or enforcement of promises? Specifically, in addition to threatening damage, need one also guarantee to withhold the damage if compliance is forthcoming; or does this depend on the configuration of "payoffs" involved? What "legal system," communication system, or information structure is needed to make the necessary promises enforceable?

Can one threaten that he will "probably" fulfill a threat; or must he threaten that he certainly will? What is the meaning of a threat that one will "probably" fulfill when it is clear that, if he retained any choice, he'd have no incentive to fulfill it after the act? More generally, what are the devices by which one gets committed to fulfillment that he would otherwise be known to shrink from, considering that if a commitment makes the threat credible enough to be effective it need not be carried out. What is the difference, if any, between a threat that deters action and one that compels action, or a threat designed to safeguard a second party from his own mistakes? Are there any logical differences among deterrent, disciplinary, and extortionate threats?

How is the situation affected by a third participant, who has his own mixture of conflict and common interest with those already present, who has access to or control of the communication system, whose behavior is rational or irrational in one sense or another, who enjoys trust or some means of contract enforcement with one or another of the two principals? How are these questions affected by the existence of a legal system that permits and prohibits certain actions, that is available to inflict penalty on nonfulfillment of contract, or that can demand authentic information from the participants. To what extent can we rationalize concepts like "reputation," "face," or "trust," in terms of a real or hypothetical legal system, in terms of modification of the participants' value systems, or in terms of relationships of the players concerned to additional participants, real or hypothetical?

This brief sample of questions may suggest that there is scope for the creation of "theory." There is something here that looks like a mixture of game theory, organization theory, communication theory, theory of evidence, theory of choice, and theory of collective decision. It is faithful to our definition of "strategy": it takes conflict for granted, but also assumes common interest between the adversaries; it assumes a "rational" value-maximizing mode of behavior; and it focuses on the fact that each participant's "best" choice of action depends on what he expects the other to do, and that "strategic behavior" is concerned with influencing another's choice by working on his expectation of how one's own behavior is related to his.

There are two points worth stressing. One is that, though "strategy of conflict" sounds cold-blooded, the theory is not concerned with the efficient *application* of violence or anything of the sort; it is not essentially a theory of aggression or of resistance or of war. *Threats* of war, yes, or threats of anything else; but it is the employment of threats, or of threats and promises, or more generally of the conditioning of one's own behavior on the behavior of others, that the theory is about.

Second, such a theory is nondiscriminatory as between the conflict and the common interest, as between its applicability to potential enemies and its applicability to potential friends. The theory degenerates at one extreme if there is no scope for mutual accommodation, no common interest at all even in avoiding mutual disaster; it degenerates at the other extreme if there is no conflict at all and no problem in identifying and reaching common goals. But in the area between those two extremes the theory is noncommittal about the mixture of conflict and common interest; we can equally well call it the theory of precarious partnership or the theory of incomplete antagonism.[7] (In Chapter 9 it is pointed out that some central aspects of the problem of surprise attack in international affairs are structurally identical with the problem of mutually suspicious partners.)

Both of these points — the neutrality of the theory with respect to the degree of conflict involved, and the definition of "strategy" as concerned with constraining an adversary through

[7] In using the word "threat" I have not intended any necessarily aggressive or hostile connotations. In an explicit negotiation between friends or in tacit cooperation between them, the threat of disagreement or of reduced cooperation, expressed or implied, is a sanction by which they support their demands, just as in a commercial transaction an offer is enforced by threat of "no sale."

his expectation of the consequences of his actions — suggest that we might call our subject the *theory of interdependent decision.*

Threats and responses to threats, reprisals and counter-reprisals, limited war, arms races, brinkmanship, surprise attack, trusting and cheating can be viewed as either hot-headed or cool-headed activities. In suggesting that they can usefully be viewed, in the development of theory, as cool-headed activities, it is not asserted that they are in fact entirely cool-headed. Rather it is asserted that the assumption of rational behavior is a productive one in the generation of systematic theory. If behavior were actually cool-headed, valid and relevant theory would probably be easier to create than it actually is. If we view our results as a bench mark for further approximation to reality, not as a fully adequate theory, we should manage to protect ourselves from the worst results of a biased theory.

Furthermore, theory that is based on the assumption that the participants coolly and "rationally" calculate their advantages according to a consistent value system forces us to think more thoroughly about the meaning of "irrationality." Decision-makers are not simply distributed along a one-dimensional scale that stretches from complete rationality at one end to complete irrationality at the other. Rationality is a collection of attributes, and departures from complete rationality may be in many different directions. Irrationality can imply a disorderly and inconsistent value system, faulty calculation, an inability to receive messages or to communicate efficiently; it can imply random or haphazard influences in the reaching of decisions or the transmission of them, or in the receipt or conveyance of information; and it sometimes merely reflects the collective nature of a decision among individuals who do not have identical value systems and whose organizational arrangements and communication systems do not cause them to act like a single entity.

As a matter of fact, many of the critical elements that go into a model of rational behavior can be identified with particular types of rationality or irrationality. The value system, the communication system, the information system, the collective decision process, or a parameter representing the probability of error

or loss of control, can be viewed as an effort to formalize the study of "irrationality." Hitler, the French Parliament, the commander of a bomber, the radar operators at Pearl Harbor, Khrushchev, and the American electorate may all suffer from some kinds of "irrationality," but by no means the same kinds. Some of them can be accounted for within a theory of rational behavior. (Even the neurotic, with inconsistent values and no method of reconciling them, motivated to suppress rather than to reconcile his conflicting goals, may for some purposes be viewed as a *pair* of "rational" entities with distinct value systems, reaching collective decisions through a voting process that has some haphazard or random element, asymmetrical communications, and so forth.)

The apparent restrictiveness of an assumption of "rational" behavior — of a calculating, value-maximizing strategy of decision — is mitigated by two additional observations. One, which I can only allege at second hand, is that even among the emotionally unbalanced, among the certified "irrationals," there is often observed an intuitive appreciation of the principles of strategy, or at least of particular applications of them. I am told that inmates of mental hospitals often seem to cultivate, deliberately or instinctively, value systems that make them less susceptible to disciplinary threats and more capable of exercising coercion themselves. A careless or even self-destructive attitude toward injury — "I'll cut a vein in my arm if you don't let me . . ." — can be a genuine strategic advantage; so can a cultivated inability to hear or to comprehend, or a reputation for frequent lapses of self-control that make punitive threats ineffectual as deterrents. (Again I am reminded of my children.) As a matter of fact, one of the advantages of an explicit theory of "rational" strategic decision in situations of mixed conflict and common interest is that, by showing the strategic basis of certain paradoxical tactics, it can display how sound and rational some of the tactics are that are practiced by the untutored and the infirm. It may not be an exaggeration to say that our sophistication sometimes suppresses sound intuitions, and one of the effects of an explicit theory may be to restore some intuitive notions that were only superficially "irrational."

The second observation is related to the first. It is that an explicit theory of "rational" decision, and of the strategic consequences of such decisions, makes perfectly clear that it is not a universal advantage in situations of conflict to be inalienably and manifestly rational in decision and motivation. Many of the attributes of rationality, as in several illustrations mentioned earlier, are strategic disabilities in certain conflict situations. It may be perfectly rational to wish oneself not altogether rational, or — if that language is philosophically objectionable — to wish for the power to suspend certain rational capabilities in particular situations. And one *can* suspend or destroy his own "rationality," at least to a limited extent; one can do this because the attributes that go to make up rationality are not inalienable, deeply personal, integral attributes of the human soul, but include such things as one's hearing aid, the reliability of the mails, the legal system, and the rationality of one's agents and partners. In principle, one might evade extortion equally well by drugging his brain, conspicuously isolating himself geographically, getting his assets legally impounded, or breaking the hand that he uses in signing checks. In a theory of strategy, several of these defenses can be represented as impairments of rationality if we wish to represent them so. A theory that makes rationality an explicit postulate is able not only to modify the postulate and examine its meaning but to take some of the mystery out of it. As a matter of fact, the paradoxical role of "rationality" in these conflict situations is evidence of the likely help that a systematic theory could provide.

And the results reached by a theoretical analysis of strategic behavior *are* often somewhat paradoxical; they often do contradict common sense or accepted rules. It is not true, as illustrated in the example of extortion, that in the face of a threat it is invariably an advantage to be rational, particularly if the fact of being rational or irrational cannot be concealed. It is not invariably an advantage, in the face of a threat, to have a communication system in good order, to have complete information, or to be in full command of one's own actions or of one's own assets. Mossadeq and my small children have already been referred to; but the same tactic is illustrated by the burning of bridges behind

oneself to persuade an adversary that one cannot be induced to retreat. An old English law that made it a serious crime to *pay* tribute to coastal pirates does not necessarily appear either cruel or anomalous in the light of a theory of strategy. It is interesting that political democracy itself relies on a particular communication system in which the transmittal of authentic evidence is precluded: the mandatory secret ballot is a scheme to deny the voter any means of proving which way he voted. Being stripped of his power to prove how he voted, he is stripped of his power to be intimidated. Powerless to prove whether or not he complied with a threat, he knows — and so do those who would threaten him — that any punishment would be unrelated to the way he actually voted.

The well-known principle that one should pick good negotiators to represent him and then give them complete flexibility and authority — a principle commonly voiced by negotiators themselves — is by no means as self-evident as its proponents suggest; the power of a negotiator often rests on a manifest inability to make concessions and to meet demands.[8] Similarly, while prudence suggests leaving open a way of escape when one threatens an adversary with mutually painful reprisal, any visible means of escape may make the threat less credible. The very notion that it may be a strategic advantage to relinquish certain options deliberately, or even to give up all control over one's future actions and make his responses automatic, seems to be a hard one to swallow.

Many of these examples involve some denial of the value of skill, resourcefulness, rationality, knowledge, control, or freedom of choice. They are all, in principle, valid in certain circumstances; but seeing through their strangeness and comprehending the logic behind them is often a good deal easier if one has formalized the problem, studied it in the abstract, and identified analogies in other contexts where the strangeness is less of an obstacle to comprehension.

Another principle contrary to the usual first impression con-

[8] The administration of foreign aid presents numerous examples. See, for example, T. C. Schelling, "American Foreign Assistance," *World Politics* (July 1955), pp. 614–15.

cerns the relative virtues of clean and dirty bombs. Bernard Brodie has pointed out that when one considers the special requirements of deterrence, in contrast to the requirements of a war that one expects to fight, one may see some utility in the super-dirty bomb.[9] As remarked in Chapter 10, this conclusion is not so strange if we recognize the "balance of terror" as simply a massive modern version of an ancient institution, the exchange of hostages.

Here perhaps we perceive a disadvantage peculiar to civilized modern students of international affairs, by contrast with, say, Machiavelli or the ancient Chinese. We tend to identify peace, stability, and the quiescence of conflict with notions like trust, good faith, and mutual respect. To the extent that this point of view actually encourages trust and respect it is good. But where trust and good faith do not exist and cannot be made to by our acting as though they did, we may wish to solicit advice from the underworld, or from ancient despotisms, on how to make agreements work when trust and good faith are lacking and there is no legal recourse for breach of contract. The ancients exchanged hostages, drank wine from the same glass to demonstrate the absence of poison, met in public places to inhibit the massacre of one by the other, and even deliberately exchanged spies to facilitate transmittal of authentic information. It seems likely that a well-developed theory of strategy could throw light on the efficacy of some of those old devices, suggest the circumstances to which they apply, and discover modern equivalents that, though offensive to our taste, may be desperately needed in the regulation of conflict.

[9] Compare p. 239 below.

2

AN ESSAY ON BARGAINING

This chapter presents a tactical approach to the analysis of bargaining. The subject includes both explicit bargaining and the tacit kind in which adversaries watch and interpret each other's behavior, each aware that his own actions are being interpreted and anticipated, each acting with a view to the expectations that he creates. In economics the subject covers wage negotiations, tariff negotiations, competition where competitors are few, settlements out of court, and the real estate agent and his customer. Outside economics it ranges from the threat of massive retaliation to taking the right of way from a taxi.

Our concern will *not* be with the part of bargaining that consists of exploring for mutually profitable adjustments, and that might be called the "efficiency" aspect of bargaining. For example, can an insurance firm save money, and make a client happier, by offering a cash settlement rather than repairing the client's car; can an employer save money by granting a voluntary wage increase to employees who agree to take a substantial part of their wages in merchandise? Instead, we shall be concerned with what might be called the "distributional" aspect of bargaining: the situations in which a better bargain for one means less for the other. When the business is finally sold to the one interested buyer, what price does it go for? When two dynamite trucks meet on a road wide enough for one, who backs up?

These are situations that ultimately involve an element of pure bargaining — bargaining in which each party is guided mainly by his expectations of what the other will accept. But with each guided by expectations and knowing that the other is too, expectations become compounded. A bargain is struck when somebody makes a final, sufficient concession. Why does he concede?

Because he thinks the other will not. "I must concede because he won't. He won't because he thinks I will. He thinks I will because he thinks I think he thinks so. . . ." There is some range of alternative outcomes in which any point is better for both sides than no agreement at all. To insist on any such point is pure bargaining, since one always *would* take less rather than reach no agreement at all, and since one always *can* recede if retreat proves necessary to agreement. Yet if both parties are aware of the limits to this range, *any* outcome is a point from which at least one party would have been willing to retreat and the other knows it! There is no resting place.

There is, however, an outcome; and if we cannot find it in the logic of the situation we may find it in the tactics employed. The purpose of this chapter is to call attention to an important class of tactics, of a kind that is peculiarly appropriate to the logic of indeterminate situations. The essence of these tactics is some voluntary but irreversible sacrifice of freedom of choice. They rest on the paradox that the power to constrain an adversary may depend on the power to bind oneself; that, in bargaining, weakness is often strength, freedom may be freedom to capitulate, and to burn bridges behind one may suffice to undo an opponent.

BARGAINING POWER: THE POWER TO BIND ONESELF

"Bargaining power," "bargaining strength," "bargaining skill" suggest that the advantage goes to the powerful, the strong, or the skillful. It does, of course, if those qualities are defined to mean only that negotiations are won by those who win. But, if the terms imply that it is an advantage to be more intelligent or more skilled in debate, or to have more financial resources, more physical strength, more military potency, or more ability to withstand losses, then the term does a disservice. These qualities are by no means universal advantages in bargaining situations; they often have a contrary value.

The sophisticated negotiator may find it difficult to seem as obstinate as a truly obstinate man. If a man knocks at a door and says that he will stab himself on the porch unless given $10, he is more likely to get the $10 if his eyes are bloodshot. The

threat of mutual destruction cannot be used to deter an adversary who is too unintelligent to comprehend it or too weak to enforce his will on those he represents. The government that cannot control its balance of payments, or collect taxes, or muster the political unity to defend itself, may enjoy assistance that would be denied it if it could control its own resources. And, to cite an example familiar from economic theory, "price leadership" in oligopoly may be an unprofitable distinction evaded by the small firms and assumed perforce by the large one.

Bargaining power has also been described as the power to fool and bluff, "the ability to set the best price for yourself and fool the other man into thinking this was your maximum offer."[1] Fooling and bluffing are certainly involved; but there are two kinds of fooling. One is deceiving about the facts; a buyer may lie about his income or misrepresent the size of his family. The other is purely tactical. Suppose each knows everything about the other, and each knows what the other knows. What is there to fool about? The buyer may say that, though he'd really pay up to twenty and the seller knows it, he is firmly resolved as a tactical matter not to budge above sixteen. If the seller capitulates, was he fooled? Or was he convinced of the truth? Or did the buyer really not know what he would do next if the tactic failed? If the buyer really "feels" himself firmly resolved, and bases his resolve on the conviction that the seller will capitulate, and the seller does, the buyer may say afterwards that he was "not fooling." Whatever has occurred, it is not adequately conveyed by the notions of bluffing and fooling.

How does one person make another believe something? The answer depends importantly on the factual question, "Is it true?" It is easier to prove the truth of something that is true than of something false. To prove the truth about our health we can call on a reputable doctor; to prove the truth about our costs or income we may let the person look at books that have been audited by a reputable firm or the Bureau of Internal Revenue. But to persuade him of something false we may have no such convincing evidence.

[1] J. N. Morgan, "Bilateral Monopoly and the Competitive Output," *Quarterly Journal of Economics*, 63:376n6 (August 1949).

When one wishes to persuade someone that he would not pay more than $16,000 for a house that is really worth $20,000 to him, what can he do to take advantage of the usually superior credibility of the truth over a false assertion? Answer: make it true. How can a buyer make it true? If he likes the house because it is near his business, he might move his business, persuading the seller that the house is really now worth only $16,000 to him. This would be unprofitable; he is no better off than if he had paid the higher price.

But suppose the buyer could make an irrevocable and enforceable bet with some third party, duly recorded and certified, according to which he would pay for the house no more than $16,000, or forfeit $5,000. The seller has lost; the buyer need simply present the truth. Unless the seller is enraged and withholds the house in sheer spite, the situation has been rigged against him; the "objective" situation — the buyer's true incentive — has been voluntarily, conspicuously, and irreversibly changed. The seller can take it or leave it. This example demonstrates that if the buyer can accept an irrevocable *commitment*, in a way that is unambiguously visible to the seller, he can squeeze the range of indeterminacy down to the point most favorable to him. It also suggests, by its artificiality, that the tactic is one that may or may not be available; whether the buyer can find an effective device for committing himself may depend on who he is, who the seller is, where they live, and a number of legal and institutional arrangements (including, in our artificial example, whether bets are legally enforceable).

If both men live in a culture where "cross my heart" is universally accepted as potent, all the buyer has to do is allege that he will pay no more than $16,000, using this invocation of penalty, and he wins — or at least he wins if the seller does not beat him to it by shouting "$19,000, cross my heart." If the buyer is an agent authorized by a board of directors to buy at $16,000 but not a cent more, and the directors cannot constitutionally meet again for several months and the buyer cannot exceed his authority, and if all this can be made known to the seller, then the buyer "wins" — if, again, the seller has not tied himself up with a commitment to $19,000. Or, if the buyer can assert that he will pay

no more than $16,000 so firmly that he would suffer intolerable loss of personal prestige or bargaining reputation by paying more, and if the fact of his paying more would necessarily be known, and if the seller appreciates all this, then a loud declaration by itself may provide the commitment. The device, of course, is a needless surrender of flexibility unless it can be made fully evident and understandable to the seller.

Incidentally, some of the more contractual kinds of commitments are not as effective as they at first seem. In the example of the self-inflicted penalty through the bet, it remains possible for the seller to seek out the third party and offer a modest sum in consideration of the latter's releasing the buyer from the bet, threatening to sell the house for $16,000 if the release is not forthcoming. The effect of the bet — as of most such contractual commitments — is to shift the locus and personnel of the negotiation, in the hope that the third party will be less available for negotiation or less subject to an incentive to concede. To put it differently, a *contractual* commitment is usually the assumption of a contingent "transfer cost," not a "real cost"; and if all interested parties can be brought into the negotiation the range of indeterminacy remains as it was. But if the third party were available only at substantial transportation cost, to that extent a truly irrevocable commitment would have been assumed. (If bets were made with a number of people, the "real costs" of bringing them into the negotiation might be made prohibitive.) [2]

[2] Perhaps the "ideal" solution to the bilateral monopoly problem is as follows. One member of the pair shifts his marginal cost curve so that joint profits are now zero at the output at which joint profits originally would have been maximized. He does this through an irrevocable sale-leaseback arrangement; he sells a royalty contract to some third party for a lump sum, the royalties so related to his output that joint costs exceed joint revenue at all other outputs. He cannot now afford to produce at any price or output except that price and output at which the entire original joint profits accrue to him; the other member of the bilateral monopoly sees the contract, appreciates the situation, and accepts his true minimum profits. The "winner" really gains the entire original profit via the lump sum for which he sold royalty rights; this profit does not affect his incentives because it is independent of what he produces. The third party pays the lump sum (minus a small discount for inducement) because he knows that the second party will have to capitulate and that therefore he will in fact get his contingent royalty. The hitch is that the royalty-rights buyer must not be available to the "losing member"; otherwise the latter can force him to renounce his royalty claim by threatening not to reach a bargain, thus restoring