THE STRATEGY OF CONFLICT

THOMAS C. SCHELLING

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
LONDON, ENGLAND

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

The most interesting parts of our topic concern whether and how commitments can be taken; but it is worth while to consider briefly a model in which practical problems are absent — a world in which absolute commitments are freely available. Consider a culture in which "cross my heart" is universally recognized as absolutely binding. Any offer accompanied by this invocation is a final offer, and is so recognized. If each party knows the other's true reservation price, the object is to be first with a firm offer. Complete responsibility for the outcome then rests with the other, who can take it or leave it as he chooses (and who chooses to take it). Bargaining is all over; the commitment (that is, the first offer) wins.

Interpose some communication difficulty. They must bargain by letter; the invocation becomes effective when signed but cannot be known to the other until its arrival. Now when one party writes such a letter the other may already have signed his own, or may yet do so before the letter of the first arrives. There is then no sale; both are bound to incompatible positions. Each must now recognize this possibility of stalemate and take into account the likelihood that the other already has, or will have, signed his own commitment.

An asymmetry in communication may well favor the one who is (and is known to be) unavailable for the receipt of messages, for he is the one who cannot be deterred from his own commitment by receipt of the other's. (On the other hand, if the one who cannot communicate can feign ignorance of his own inability, the other too may be deterred from his own commitment by fear of the first's unwitting commitment.) If the commitments depend not just on words but on special forms or ceremonies, ignorance of the other party's commitment ceremonies may be an advantage if the ignorance is fully appreciated, since it makes the other aware that only his own restraint can avert stalemate.

Suppose only part of the population belongs to the cult in which "cross my heart" is (or is believed to be) absolutely bind-

---

the original marginal cost situation. But we may imagine the development of institutions that specialize in royalty purchases, whose ultimate success depends on a reputation for never renegotiating, and whose incentives can thus not be appealed to in any single negotiation.

ing. If everyone knows (and is known to know) everyone else's affiliation, those belonging to this particular cult have the advantage. They can commit themselves, the others cannot. If the buyer says "$16,000, cross my heart" his offer is final; if the seller says "$19,000" he is (and is known to be) only "bargaining."

If each does not know the other's true reservation price there is an initial stage in which each tries to discover the other's and misrepresent his own, as in ordinary bargaining. But the process of discovery and revelation becomes quickly merged with the process of creating and discovering commitments; the commitments permanently change, for all practical purposes, the "true" reservation prices. If one party has, and the other has not, the belief in a binding ceremony, the latter pursues the "ordinary" bargaining technique of *asserting* his reservation price, while the former proceeds to *make* his.

The foregoing discussion has tried to suggest both the plausibility and the logic of self-commitment. Some examples may suggest the relevance of the tactic, although an observer can seldom distinguish with confidence the consciously logical, the intuitive, or the inadvertent use of a visible tactic. First, it has not been uncommon for union officials to stir up excitement and determination on the part of the membership during or prior to a wage negotiation. If the union is going to insist on $2 and expects the management to counter with $1.60, an effort is made to persuade the membership not only that the management could pay $2 but even perhaps that the negotiators themselves are incompetent if they fail to obtain close to $2. The purpose — or, rather, a plausible purpose suggested by our analysis — is to make clear to the management that the negotiators could not accept less than $2 *even if they wished to* because they no longer control the members or because they would lose their own positions if they tried. In other words, the negotiators reduce the scope of their own authority and confront the management with the threat of a strike that the union itself cannot avert, even though it was the union's own action that eliminated its power to prevent the strike.

Something similar occurs when the United States Government negotiates with other governments on, say, the uses to which foreign assistance will be put, or tariff reduction. If the executive branch is free to negotiate the best arrangement it can, it may be unable to make any position stick and may end by conceding controversial points because its partners know, or believe obstinately, that the United States would rather concede than terminate the negotiations. But, if the executive branch negotiates under legislative authority, with its position constrained by law, and it is evident that Congress will not be reconvened to change the law within the necessary time period, then the executive branch has a firm position that is visible to its negotiating partners.

When national representatives go to international negotiations knowing that there is a wide range of potential agreement within which the outcome will depend on bargaining, they seem often to create a bargaining position by public statements, statements calculated to arouse a public opinion that permits no concessions to be made. If a binding public opinion can be cultivated and made evident to the other side, the initial position can thereby be made visibly "final."

These examples have certain characteristics in common. First, they clearly depend not only on incurring a commitment but on communicating it persuasively to the other party. Second, it is by no means easy to establish the commitment, nor is it entirely clear to either of the parties concerned just how strong the commitment is. Third, similar activity may be available to the parties on both sides. Fourth, the possibility of commitment, though perhaps available to both sides, is by no means equally available; the ability of a democratic government to get itself tied by public opinion may be different from the ability of a totalitarian government to incur such a commitment. Fifth, they all run the risk of establishing an immovable position that goes beyond the ability of the other to concede, and thereby provoke the likelihood of stalemate or breakdown.

INSTITUTIONAL AND STRUCTURAL CHARACTERISTICS OF THE NEGOTIATION

Some institutional and structural characteristics of bargaining situations may make the commitment tactic easy or difficult to use, or make it more available to one party than the other, or affect the likelihood of simultaneous commitment or stalemate.

Use of a Bargaining Agent. The use of a bargaining agent affects the power of commitment in at least two ways. First, the agent may be given instructions that are difficult or impossible to change, such instructions (and their inflexibility) being visible to the opposite party. The principle applies in distinguishing the legislative from the executive branch, or the management from the board of directors, as well as to a messenger-carried offer when the bargaining process has a time limit and the principal has interposed sufficient distance between himself and his messenger to make further communication evidently impossible before the time runs out.

Second, an "agent" may be brought in as a principal in his own right, with an incentive structure of his own that differs from his principal's. This device is involved in automobile insurance; the private citizen, in settling out of court, cannot threaten suit as effectively as the insurance company since the latter is more conspicuously obliged to carry out such threats to maintain its own reputation for subsequent accidents.[3]

Secrecy vs. Publicity. A potent means of commitment, and sometimes the only means, is the pledge of one's reputation. If national representatives can arrange to be charged with appeasement for every small concession, they place concession visibly beyond their own reach. If a union with other plants to deal with can arrange to make any retreat dramatically visible, it places its bargaining reputation in jeopardy and thereby becomes visibly incapable of serious compromise. (The same convenient jeopardy is the basis for the universally exploited defense, "If I did it for you I'd have to do it for everyone else.") But to commit in this

[3] The formal solution to the right-of-way problem in automobile traffic may be that the winner is the one who first becomes fully and visibly insured against all contingencies; since he then has no incentive to avoid accident, the other must yield and knows it. (The latter cannot counter in kind; no company will insure him now that the first is insured.) More seriously, the pooling of strike funds among unions reduces the visible incentive on each individual union to avoid a strike. As in the bilateral monopoly solution suggested earlier, there is a transfer of interest to a third party with a resulting visible shift in one's own incentive structure.

fashion publicity is required. Both the initial offer and the final outcome would have to be known; and if secrecy surrounds either point, or if the outcome is inherently not observable, the device is unavailable. If one party has a "public" and the other has not, the latter may try to neutralize his disadvantage by excluding the relevant public; or if both parties fear the potentialities for stalemate in the simultaneous use of this tactic, they may try to enforce an agreement on secrecy.

Intersecting Negotiations. If a union is simultaneously engaged, or will shortly be engaged, in many negotiations while the management has no other plants and deals with no other unions, the management cannot convincingly stake its bargaining reputation while the union can. The advantage goes to the party that can persuasively point to an array of other negotiations in which its own position would be prejudiced if it made a concession in this one. (The "reputation value" of the bargain may be less related to the outcome than to the firmness with which some initial bargaining position is adhered to.) Defense against this tactic may involve, among other things, both misinterpretation of the other party's position and an effort to make the eventual outcome incommensurable with the initial positions. If the subjects under negotiation can be enlarged in the process of negotiation, or the wage figure replaced by fringe benefits that cannot be reduced to a wage equivalent, an "out" is provided to the party that has committed itself; and the availability of this "out" weakens the commitment itself, to the disadvantage of the committed party.

Continuous Negotiations. A special case of interrelated negotiations occurs when the same two parties are to negotiate other topics, simultaneously or in the future. The logic of this case is more subtle; to persuade the other that one cannot afford to recede, one says in effect, "If I conceded to you here, you would revise your estimate of me in our other negotiations; to protect my reputation with you I must stand firm." The second party is simultaneously the "third party" to whom one's bargaining reputation can be pledged. This situation occurs in the threat of local resistance to local aggression. The party threatening achieves its commitment, and hence the credibility of its threat, not by referring to what it would gain from carrying out the threat in this particular instance but by pointing to the long-run value of a fulfilled threat in enhancing the credibility of future threats.

The Restrictive Agenda. When there are two objects to negotiate, the decision to negotiate them simultaneously or in separate forums or at separate times is by no means neutral to the outcome, particularly when there is a latent extortionate threat that can be exploited only if it can be attached to some more ordinary, legitimate, bargaining situation. The protection against extortion depends on refusal, unavailability, or inability, to negotiate. But if the object of the extortionate threat can be brought onto the agenda with the other topic, the latent threat becomes effective.

Tariff bargaining is an example. If reciprocal tariffs on cheese and automobiles are to be negotiated, one party may alter the outcome by threatening a purely punitive change in some other tariff. But if the bargaining representatives of the threatened party are confined to the cheese-automobile agenda, and have no instructions that permit them even to take cognizance of other commodities, or if there are ground rules that forbid mention of other tariffs while cheese and automobiles remain unsettled, this extortionate weapon must await another opportunity. If the threat that would be brought to the conference table is one that cannot stand publicity, publicity itself may prevent its effective communication.

The Possibility of Compensation. As Fellner has pointed out, agreement may be dependent on some means of redistributing costs or gains.[4] If duopolists, for example, divide markets in a way that maximizes their combined profits, some initial accrual of profits is thereby determined; any other division of the profits requires that one firm be able to compensate the other. If the fact of compensation would be evidence of illegal collusion, or if the motive for compensation would be misunderstood by the stockholders, or if the two do not sufficiently trust each other, some less optimum level of *joint* profits may be required in order that the

[4] W. Fellner, *Competition Among the Few* (New York, 1949), pp. 34–35, 191–97, 231–32, 234.

initial accrual of profits to the two firms be in closer accordance with an agreed division of gains between them.

When agreement must be reached on something that is inherently a one-man act, any division of the cost depends on compensation. The "agenda" assumes particular importance in these cases, since a principal means of compensation is a concession on some other object. If two simultaneous negotiations can be brought into a contingent relationship with each other, a means of compensation is available. If they are kept separate, each remains an indivisible object.

It may be to the advantage of one party to keep a bargain isolated, and to the other to join it to some second bargain. If there are two projects, each with a cost of three, and each with a value of two to A and a value of four to B, and each is inherently a "one-man" project in its execution, and if compensation is institutionally impossible, B will be forced to pay the entire cost of each as long as the two projects are kept separate. He cannot usefully threaten nonperformance, since A has no incentive to carry out either project by himself. But if B can link the projects together, offering to carry out one while A carries out the other, and can effectively threaten to abandon both unless A carries out one of them, A is left an option with a gain of four and a cost of three, which he takes, and B cuts his cost in half.

An important limitation of economic problems, as prototypes of bargaining situations, is that they tend disproportionately to involve divisible objects and compensable activities. If a drainage ditch in the back of one house will protect both houses; and if it costs \$1,000 and is worth \$800 to each home-owner; neither would undertake it separately, but we nevertheless usually assume that they will get together and see that this project worth \$1,600 to the two of them gets carried out. But if it costs 10 hours a week to be scoutmaster, and each considers it worth 8 hours of his time to have a scout troop but one man must do the whole job, it is far from certain that the neighbors will reach a deal according to which one puts 10 hours on the job and the other pays him cash or does 5 hours' gardening for him. When two cars meet on a narrow road, the ensuing deadlock is aggravated by the absence of a custom of bidding to pay for the right of way. Parliamentary deadlocks occur when logrolling is impracticable. Measures that require unanimous agreement can often be initiated only if several are bundled together.[5]

The Mechanics of Negotiation. A number of other characteristics deserve mention, although we shall not work out their implications. Is there a penalty on the conveyance of false information? Is there a penalty on called bluffs, that is, can one put forth an offer and withdraw it after it has been accepted? Is there a penalty on hiring an agent who pretends to be an interested party and makes insincere offers, simply to test the position of the other party? Can all interested parties be recognized? Is there a time limit on the bargaining? Does the bargaining take the particular structure of an auction, a Dutch auction, a sealed bid system, or some other formal arrangement? Is there *a status quo,* so that unavailability for negotiation can win the *status quo* for the party that prefers it? Is renegotiation possible in case of stalemate? What are the costs of stalemate? Can compliance with the agreement be observed? What, in general, are the means of communication, and are any of them susceptible of being put out of order by one party or the other? If there are several items to negotiate, are they negotiated in one comprehensive negotiation, separately in a particular order so that each piece is finished before the next is taken up, or simultaneously through different agents or under different rules.

The importance of many of these structural questions becomes evident when one reflects on parliamentary technique. Rules that permit a president to veto an appropriation bill only in its entirety, or that require each amendment to be voted before the original act is voted on, or a priority system accorded to different kinds of motions, substantially alter the incentives that are brought to bear on each action. One who might be pressured into choosing second best is relieved of his vulnerability if he can vote earlier to eliminate that possibility, thereby leaving only

[5] Inclusion of a provision on the Saar in the "Paris Agreements" that ended the occupation of Western Germany may have reflected either this principle or the one in the preceding paragraph.

first and third choices about which his preference is known to be so strong that no threat will be made.

Principles and Precedents. To be convincing, commitments usually have to be qualitative rather than quantitative, and to rest on some rationale. It may be difficult to conceive of a really firm commitment to $2.07½; why not $2.02¼? The numerical scale is too continuous to provide good resting places, except at nice round numbers like $2.00. But a commitment to the *principle* of "profit sharing," "cost-of-living increases," or any other basis for a numerical calculation that comes out at $2.07½, may provide a foothold for a commitment. Furthermore, one may create something of a commitment by putting the principles and precedents themselves in jeopardy. If in the past one has successfully maintained the principle of, say, nonrecognition of governments imposed by force, and elects to nail his demands to that principle in the present negotiation, he not only adduces precedent behind his claim but risks the principle itself. Having pledged it, he may persuade his adversary that he would accept stalemate rather than capitulate and discredit the principle.

Casuistry. If one reaches the point where concession is advisable, he has to recognize two effects: it puts him closer to his opponent's position, and it affects his opponent's estimate of his firmness. Concession not only may be construed as capitulation, it may mark a prior commitment as a fraud, and make the adversary skeptical of any new pretense at commitment. One, therefore, needs an "excuse" for accommodating his opponent, preferably a rationalized reinterpretation of the original commitment, one that is persuasive to the adversary himself.

More interesting is the use of casuistry to release an opponent from a commitment. If one can demonstrate to an opponent that the latter is not committed, or that he has miscalculated his commitment, one may in fact undo or revise the opponent's commitment. Or if one can confuse the opponent's commitment, so that his constituents or principals or audience cannot exactly identify compliance with the commitment — show that "productivity" is ambiguous, or that "proportionate contributions" has several meanings — one may undo it or lower its value. In these cases it is to the opponent's disadvantage that this commitment be successfully refuted by argument. But when the opponent has resolved to make a moderate concession one may help him by proving that he *can* make a moderate concession consistent with his former position, and that if he does there are no grounds for believing it to reflect on his original principles. One must seek, in other words, a rationalization by which to deny oneself too great a reward from the opponent's concession, otherwise the concession will not be made.[6]

THE THREAT

When one threatens to fight if attacked or to cut his price if his competitor does, the threat is no more than a communication of one's own incentives, designed to impress on the other the automatic consequences of his act. And, incidentally, if it succeeds in deterring, it benefits both parties.

But more than communication is involved when one threatens an act that he would have no incentive to perform but that is designed to deter through its promise of mutual harm. To threaten massive retaliation against small encroachments is of this nature, as is the threat to bump a car that does not yield the right of way or to call a costly strike if the wage is not raised a few cents. The distinctive feature of this threat is that the threatener has no in-

[6] In many textbook problems, such as bilateral monopoly between firms, the ends of the bargaining range are points of zero profits for one or the other party; and to settle for one's minimum position is no better than no settlement at all. But, apart from certain buying and selling situations, there are commonly limits on the range of acceptable outcomes, and the least favorable outcome that one is free to accept may be substantially superior to stalemate. In these cases one's overriding purpose may be to forestall any misguided commitment by the other party. If the truth is more demonstrable than a false position, a conservative initial position is indicated, as it is if any withdrawal from an initial "advanced" position would discredit any subsequent attempt to convey the truth. Actually, though a person does not commonly invite penalties on his own behavior, the existence of an enforceable penalty on falsehood would be of assistance; if one can demonstrate, for example, his cost or income position by showing his income tax return, the penalties on fraud may enhance the value of this evidence.

Even the "pure" bilateral monopoly case becomes somewhat of this nature if the bargaining is conducted by agents or employees whose rewards are more dependent on *whether* agreement is reached than on how favorable the terms of the agreement are.

centive to carry it out either before the event or after. He does have an incentive to bind himself to fulfill the threat, if he thinks the threat may be successful, because the threat and not its fulfillment gains the end; and fulfillment is not required if the threat succeeds. The more certain the contingent fulfillment is, the less likely is actual fulfillment. But the threat's efficacy depends on the credulity of the other party, and the threat is ineffectual unless the threatener can rearrange or display his own incentives so as to demonstrate that he would, *ex post*, have an incentive to carry it out.[7]

We are back again at the commitment. How can one commit himself in advance to an act that he would in fact prefer not to carry out in the event, in order that his commitment may deter the other party? One can of course bluff, to persuade the other falsely that the costs or damages to the threatener would be minor or negative. More interesting, the one making the threat may pretend that he himself erroneously believes his own costs to be small, and therefore would mistakenly go ahead and fulfill the threat. Or perhaps he can pretend a revenge motivation so strong as to overcome the prospect of self-damage; but this option is probably most readily available to the truly revengeful. Otherwise he must find a way to commit himself.

One may try to stake his reputation on fulfillment, in a manner that impresses the threatened person. One may even stake his reputation *with the threatened person himself*, on grounds that it would be worth the costs and pains to give a lesson to the latter if he fails to heed the threat. Or one may try to arrange a legal commitment, perhaps through contracting with a third party.[8]

[7] Incidentally, the deterrent threat has some interesting quantitative characteristics, reflecting the general asymmetry between rewards and punishments. It is not necessary, for example, that the threat promise more damage to the party threatened than to the party carrying it out. The threat to smash an old car with a new one may succeed if believed, or to sue expensively for small damages, or to start a price war. Also, as far as the power to deter is concerned, there is no such thing as "too large" a threat; if it is large enough to succeed, it is not carried out anyway. A threat is only "too large" if its very size interferes with its credibility. Atomic destruction for small misdemeanors, like expensive incarceration for overtime parking, would be superfluous but not exorbitant unless the threatened person considered it too awful to be real and ignored it.

[8] Mutual defense treaties among strong and weak nations might best be

Or if one can turn the whole business over to an agent whose salary (or business reputation) depends on carrying out the threat but who is unalterably relieved of any responsibility for the further costs, one may shift the incentive.

The commitment problem is nicely illustrated by the legal doctrine of the "last clear chance" which recognizes that, in the events that led up to an accident, there was some point at which the accident became inevitable as a result of prior actions, and that the abilities of the two parties to prevent it may not have expired at the same time. In bargaining, the commitment is a device to leave the last clear chance to decide the outcome with the other party, in a manner that he fully appreciates; it is to relinquish further initiative, having rigged the incentives so that the other party must choose in one's favor. If one driver speeds up so that he cannot stop, and the other realizes it, the latter has to yield. A legislative rider at the end of a session leaves the President the last clear chance to pass the bill. This doctrine helps to understand some of those cases in which bargaining "strength" inheres in what is weakness by other standards. When a person — or a country — has lost the power to help himself, or the power to avert mutual damage, the other interested party has no choice but to assume the cost or responsibility. "Coercive deficiency" is the term Arthur Smithies uses to describe the tactic of deliberately exhausting one's annual budgetary allowance so early in the year that the need for more funds is irresistibly urgent.[9]

A related tactic is maneuvering into a *status quo* from which one can be dislodged only by an overt act, an act that precipitates mutual damage because the maneuvering party has relinquished the power to retreat. If one carries explosives visibly on his person, in a manner that makes destruction obviously inevitable for himself and for any assailant, he may deter assault much more than if he retained any control over the explosives. If one com-

---

viewed in this light, that is, not as undertaken to reassure the small nations nor in exchange for a *quid pro quo*, but rather as a device for surrendering an embarrassing freedom of choice.

[9] A. Smithies, *The Budgetary Process in the United States* (New York, 1955), pp. 40, 56. One solution is the short tether of an apportionment process. See also T. C. Schelling, "American Foreign Assistance," *World Politics*, 7:609–625 (July 1955), regarding the same principle in foreign aid allocations.

mits a token force of troops that would be unable to escape, the commitment to full resistance is increased. Walter Lippmann has used the analogy of the plate glass window that helps to protect a jewelry store: anyone can break it easily enough, but not without creating an uproar.

Similar techniques may be available to the one threatened. His best defense, of course, is to carry out the act before the threat is made; in that case there is neither incentive nor commitment for retaliation. If he cannot hasten the act itself, he may commit himself to it; if the person to be threatened is already committed, the one who would threaten cannot deter with his threat, he can only make certain the mutually disastrous consequences that he threatens.[10] If the person to be threatened can arrange before the threat is made to share the risk with others (as suggested by the insurance solution to the right-of-way problem mentioned earlier) he may become so visibly unsusceptible to the threat as to dissuade the threatener. Or if by any other means he can either change or misrepresent his own incentives, to make it appear that he would gain in spite of threat fulfillment (or perhaps only that he thinks he would), the threatener may have to give up the threat as costly and fruitless; or if one can misrepresent himself as either unable to comprehend a threat, or too obstinate to heed it, he may deter the threat itself. Best of all may be *genuine* ignorance, obstinacy, or simple disbelief, since it may be more convincing to the prospective threatener; but of course if it fails to persuade him and he commits himself to the threat, both sides lose. Finally, both the threat and the commitment have to be communicated; if the threatened person can be unavailable for messages, or can destroy the communication channels, even though

[10] The system of supplying the police with traffic tickets that are numbered and incapable of erasures makes it possible for the officer, by writing in the license number of the car before speaking to the driver, to preclude the latter's threat. Some trucks carry signs that say, "Alarm and lock system not subject to the driver's control." The time lock on bank vaults serves much the same purpose, as does the mandatory secret ballot in elections. So does starting an invasion with a small advance force that, though too small and premature to win the objective, attaches too much "face" to the enterprise to permit withdrawal: the larger force can then be readied without fear of inviting a purely deterrent threat. At many universities the faculty is protected by a rule that denies instructors the power to change a course grade once it has been recorded.

he does so in an obvious effort to avert threat, he may deter the threat itself.[11] But the time to show disbelief or obstinacy is before the threat is made, that is, before the commitment is taken, not just before the threat is fulfilled; it does no good to be incredulous, or out of town, when the messenger arrives with the committed threat.

In threat situations, as in ordinary bargaining, commitments are not altogether clear; each party cannot exactly estimate the costs and values to the other side of the two related actions involved in the threat; the process of commitment may be a progressive one, the commitments acquiring their firmness by a sequence of actions. Communication is often neither entirely impossible nor entirely reliable; while certain evidence of one's commitment can be communicated directly, other evidence must travel by newspaper or hearsay, or be demonstrated by actions. In these cases the unhappy possibility of both acts occurring, as a result of simultaneous commitment, is increased. Furthermore, the recognition of this possibility of simultaneous commitment becomes itself a deterrent to the taking of commitments.[12]

In case a threat is made and fails to deter, there is a second stage prior to fulfillment in which *both* parties have an interest in undoing the commitment. The purpose of the threat is gone, its deterrence value is zero, and only the commitment exists to motivate fulfillment. This feature has, of course, an analogy with stalemate in ordinary bargaining, stalemate resulting from both parties' getting committed to incompatible positions, or one party's

[11] The racketeer cannot sell protection if he cannot find his customer at home; nor can the kidnapper expect any ransom if he cannot communicate with friends or relatives. Thus, as a perhaps impractical suggestion, a law that required the immediate confinement of all interested friends and relatives when a kidnapping occurred might make the prospects for ransom unprofitably dim. The rotation of watchmen and policemen, or their assignment in random pairs, not only limits their exploitation of bribes but protects them from threats.

[12] It is a remarkable institutional fact that there is no simple, universal way for persons or nations to assume commitments of the kind we have been discussing. There are numerous ways they can try, but most of them are quite ambiguous, unsure, or only occasionally available. In the "cross-my-heart" society adverted to earlier, bargaining theory would reduce itself to game strategy and the mechanics of communication; but in most of the contemporary world the topic is mainly an empirical and institutional one of who can commit, how, and with what assurance of appreciation by the other side.

mistakenly committing himself to a position that the other truly would not accept. If there appears a possibility of undoing the commitment, *both* parties have an interest in doing so. How to undo it is a matter on which their interests diverge, since different ways of undoing it lead to different outcomes. Furthermore, "undoing" does not mean neglecting a commitment regardless of reputation; "undoing," if the commitment of reputation was real, means disconnecting the threat from one's reputation, perhaps one's own reputation with the threatened person himself. It is therefore a subtle and tenuous situation in which, though both have an interest in undoing the commitment, they may be quite unable to collaborate in undoing it.

Special care may be needed in defining the threat, both the act that is threatened against and the counter act that is threatened. The difficulty arises from the fact, just noted, that once the former has been done the incentive to perform the later has disappeared. The credibility of the threat before the act depends on how visible to the threatened party is the inability of the threatening party to rationalize his way out of his commitment once it has failed its purpose. Any loopholes the threatening party leaves himself, if they are visible to the threatened party, weaken the visible commitment and hence reduce the credibility of the threat. (An example may be the ambiguous treatment of Quemoy in the Formosa Resolution and Treaty.)

It is essential, therefore, for maximum credibility, to leave as little room as possible for judgment or discretion in carrying out the threat. If one is committed to punish a certain type of behavior when it reaches certain limits, but the limits are not carefully and objectively defined, the party threatened will realize that when the time comes to decide whether the threat must be enforced or not, his interest and that of the threatening party will coincide in an attempt to avoid the mutually unpleasant consequences.

In order to make a threat precise, so that its terms are visible both to the threatened party and to any third parties whose reaction to the whole affair is of value to the adversaries, it may be necessary to introduce some arbitrary elements. The threat must involve overt acts rather than intentions; it must be attached to the visible deeds, not invisible ones; it may have to attach itself to certain ancillary actions that are of no consequence in themselves to the threatening party. It may, for example, have to put a penalty on the carrying of weapons rather than their use; on suspicious behavior rather than observed misdemeanors; on proximity to a crime rather than the crime itself. And, finally, the act of punishment must be one whose effect or influence is clearly discernible.[13]

In order that one be able to pledge his reputation behind a threat, there must be continuity between the present and subsequent issues that will arise. This need for continuity suggests a means of making the original threat more effective; if it can be decomposed into a series of consecutive smaller threats, there is an opportunity to demonstrate on the first few transgressions that the threat will be carried out on the rest. Even the first few become more plausible, since there is a more obvious incentive to fulfill them as a "lesson."

This principle is perhaps most relevant to acts that are inherently a matter of degree. In foreign aid programs the overt act of terminating assistance may be so obviously painful to both sides as not to be taken seriously by the recipient, but if each small misuse of funds is to be accompanied by a small reduction in assistance, never so large as to leave the recipient helpless nor to provoke a diplomatic breach, the willingness to carry it out will receive more credulity; or if it does not at first, a few lessons may be persuasive without too much damage.[14]

The threatening party may not, of course, be able to divide the act into steps. (Both the act to be deterred and the punishment must be divisible.) But the principle at least suggests the unwisdom of defining aggression, or transgression, in terms of some

[13] During 1950, the Economic Cooperation Administration declared its intention to reward Marshall Plan countries that followed especially sound policies, and to penalize those that did not, through the device of larger or smaller aid allotments. But since the base figures had not been determined, and since their determination would ultimately involve judgment rather than formulas, there would be no way afterwards to see whether in fact the additions and subtractions were made, and the plan suffered from implausibility.

[14] Perhaps the common requirement for amortization of loans at frequent intervals, rather than in a lump sum at the end of the loan period, reflects an analogous principle, as does the custom of giving frequent examinations in a college course to avoid letting a student's failure hinge exclusively on a single grading decision after the course is finished.

critical degree or amount that will be deemed intolerable. When the act to be deterred is inherently a sequence of steps whose cumulative effect is what matters, a threat geared to the increments may be more credible than one that must be carried out either all at once or not at all when some particular point has been reached. It may even be impossible to define a "critical point" with sufficient clarity to be persuasive.

To make the threatened acts divisible, the acts themselves may have to be modified. Parts of an act that cannot be decomposed may have to be left out; ancillary acts that go with the event, though of no interest in themselves, may be objects to which a threat can effectively be attached. For example, actions that are only preparatory to the main act, and by themselves do no damage, may be susceptible of chronological division and thus be effective objects of the threat. The man who would kick a dog should be threatened with modest punishment for each step toward the dog, even though his proximity is of no interest in itself.

Similar to decomposing a threat into a series is starting a threat with a punitive act that grows in severity with the passage of time. Where a threat of death by violence might not be credited, cutting off the food supply might bring submission. For moral or public relations purposes, this device may in fact leave the "last clear chance" to the other, whose demise is then blamed on his stubbornness if the threat fails. But in any case the threatener gets his overt act out of the way while it is still preliminary and minor, rather than letting it stand as a final, dreadful, and visible obstacle to his resolution. And if the suffering party is the only one in a position to know, from moment to moment, how near to catastrophe they have progressed, his is the last clear chance in a real sense. Furthermore, the threatener may be embarrassed by his adversary's collapse but not by his discomfort; and the device may therefore transform a dangerous once-for-all threat into a less costly continuous one. Tenants are less easily removed by threat of forcible eviction than by simply shutting off the utilities.[15]

[15] This seems to be the tactic that avoided an explosion and induced de Gaulle's forces to vacate a province they had occupied in Northern Italy in June 1945, after they had announced that any effort of their allies to dislodge them would

A piecemeal approach may also be used by the threatened person. If he cannot obviate the threat by hastening the entire act, he may hasten some initial stage that clearly commits him to eventual completion. Or, if his act is divisible while the threatener's retaliation comes only in the large economy size, performing it as a series of increments may deny the threatener the dramatic overt act that would trigger his response.

THE PROMISE

Among the legal privileges of corporations, two that are mentioned in textbooks are the right to sue and the "right" to be sued. Who wants to be sued! But the right to be sued is the power to make a promise: to borrow money, to enter a contract, to do business with someone who might be damaged. If suit does arise, the "right" seems a liability in retrospect; beforehand it was a prerequisite to doing business.

In brief, the right to be sued is the power to accept a commitment. In the commitments discussed up to this point, it was essential that one's adversary (or "partner," however we wish to describe him) not have the power to release one from the commitment; the commitment was, in effect, to some third party, real or fictitious. The promise is a commitment to the second party in the bargain and is required whenever the final action of one or of each is outside the other's control. It is required whenever an agreement leaves any incentive to cheat.[16]

This need for promises is more than incidental; it has an institutional importance of its own. It is not always easy to make a convincing, self-binding, promise. Both the kidnapper who would like to release his prisoner, and the prisoner, may search desperately for a way to commit the latter against informing on his captor, without finding one. If the victim has committed an act

---

be treated as a hostile act. See Harry S Truman, *Year of Decisions* (New York, 1955), pp. 239–42; and Winston S. Churchill, *Triumph and Tragedy*, vol. VI of *The Second World War* (Boston, 1953), pp. 566–68.

[16] The threat may seem to be a promise if the pledge behind it is only one's reputation with his adversary; but it is not a promise from which the second party can unilaterally release the threatener, since he cannot convincingly dissociate his own future estimate of the threatener from the latter's performance.

whose disclosure could lead to blackmail, he may confess it; if not, he might commit one in the presence of his captor, to create the bond that will ensure his silence. But these extreme possibilities illustrate how difficult, as well as important, it may be to assume a promise. If the law will not enforce price agreements; or if the union is unable to obligate itself to a no-strike pledge; or if a contractor has no assets to pay damages if he loses a suit, and the law will not imprison debtors; or if there is no "audience" to which one can pledge his reputation; it may not be possible to strike a bargain, or at least the same bargain that would otherwise be struck.

Bargaining may have to concern itself with an "incentive" system as well as the division of gains. Oligopolists may lobby for a "fair-trade" law; or exchange shares of stocks. An agreement to stay out of each other's market may require an agreement to redesign the products to be unsuitable in each other's area. Two countries that wish to agree not to make military use of an island may have to destroy the usefulness of the island itself. (In effect, a "third-party commitment" has to be assumed when an effective "second-party commitment" cannot be devised.) [17]

Fulfillment is not always observable. If one sells his vote in a secret election, or a government agrees to recommend an act to its parliament, or an employee agrees not to steal from inventory, or a teacher agrees to keep his political opinions out of class, or a country agrees to stimulate exports "as much as possible," there is no reliable way to observe or measure compliance. The observable outcome is subject to a number of influences, only one of which is covered by the agreement. The bargain may therefore have to be expressed in terms of something observable, even though what is observable is not the intended object of the bargain. One may have to pay the bribed voter if the election is won, not on how he voted; to pay a salesman a commission on sales, rather than on skill and effort; to reward policemen according to statistics on crime rather than on attention to duty; or to punish all employees for the transgressions of one. And, where performance is a matter of degree, the bargain may have to define arbitrary limits distinguishing performance from nonperformance; a speci-

[17] In an earlier age, hostages were exchanged.

fied loss of inventory treated as evidence of theft; a specified increase in exports considered an "adequate" effort; specified samples of performance taken as representative of total performance.[18]

The tactic of decomposition applies to promises as well as to threats. What makes many agreements enforceable is only the recognition of future opportunities for agreement that will be eliminated if mutual trust is not created and maintained, and whose value outweighs the momentary gain from cheating in the present instance. Each party must be confident that the other will not jeopardize future opportunities by destroying trust at the outset. This confidence does not always exist; and one of the purposes of piecemeal bargains is to cultivate the necessary mutual expectations. Neither may be willing to trust the other's prudence (or the other's confidence in the first's prudence, and so forth) on a large issue. But, if a number of preparatory bargains can be struck on a small scale, each may be willing to risk a small investment to create a tradition of trust. The purpose is to let each party demonstrate that he appreciates the need for trust and that he knows the other does too. So, if a major issue has to be negotiated, it may be necessary to seek out and negotiate some minor items for "practice," to establish the necessary confidence in each other's awareness of the long-term value of good faith.

Even if the future will bring no recurrence, it may be possible to create the equivalence of continuity by dividing the bargaining issue into consecutive parts. If each party agrees to send a million dollars to the Red Cross on condition the other does, each may be tempted to cheat if the other contributes first, and each one's anticipation of the other's cheating will inhibit agreement. But if the contribution is divided into consecutive small contributions, each can try the other's good faith for a small price. Furthermore, since each can keep the other on short tether to the finish, no one ever need risk more than one small contribution at a time. Finally, this change in the incentive structure itself takes most

[18] Inability to assume an enforceable promise, like inability to perform the activity demanded, may protect one from an extortionate threat. The mandatory secret ballot is a nuisance to the voter who would like to sell his vote, but protection to the one who would fear coercion.

of the risk out of the initial contribution; the value of established trust is made obviously visible to both.

Preparatory bargains serve another purpose. Bargaining can only occur when at least one party takes initiative in proposing a bargain. A deterrent to initiative is the information it yields, or may seem to yield, about one's eagerness. But if each has visible reason to expect the other to meet him half way, because of a history of successful bargaining, that very history provides protection against the inference of overeagerness.[19]

AN ILLUSTRATIVE GAME

Various bargaining situations involving commitments, threats, promises, and communication problems, can be illustrated by variants of a game in which each of two persons has a pair of alternatives from which to choose. North chooses either A or α; East chooses either B or β. Each person's gain depends on the choices of both. Each of the four possible combined choices, AB, $A\beta$, αB, or $\alpha\beta$, yields a particular gain or loss for North and a particular gain or loss for East. No compensation is payable between North and East. In general, each person's preference may depend on the choice the other makes.

Each such game can be quantitatively represented in a two-dimensional graph, with North's gain measured vertically and East's horizontally, and the values of the four combined choices denoted by points labeled AB, $A\beta$, $\alpha\beta$, and αB. In spite of the simplicity of the game there is actually a large number of qualitatively different variants, depending not only on the relative positions of the four points in the plane but also on the "rules" about order of moves, possibility of communication, availability of means of commitment, enforceability of promises, and whether

[19] Perhaps two adversaries who look forward to some large negotiated settlement would do well to keep avenues open for negotiation of minor issues. If, for example, the number of loose ends in dispute between East and West should narrow down so much that nothing remains to be negotiated but the "ultimate issue" (some final, permanent disposition of all territories and armaments) the possibility of even opening negotiations on the latter might be jeopardized. Or, if the minor issues are not disposed of, but become so attached to the "big" issue that willingness to negotiate on them would be construed as overeagerness on the whole settlement, the possibility of preparatory bargains might disappear.

two or more games between two persons can be joined together. The variations can be multiplied almost without limit by selecting different hypotheses about what each player knows or guesses about the "values" of the four outcomes for the other player, and what he guesses the other party guesses about himself. For convenience we assume here that the eight "values" are obvious in an obvious way to both persons. And, just as we have ruled out compensation, we rule out also threats of actions that lie outside the game. A very small sample of such games is presented.

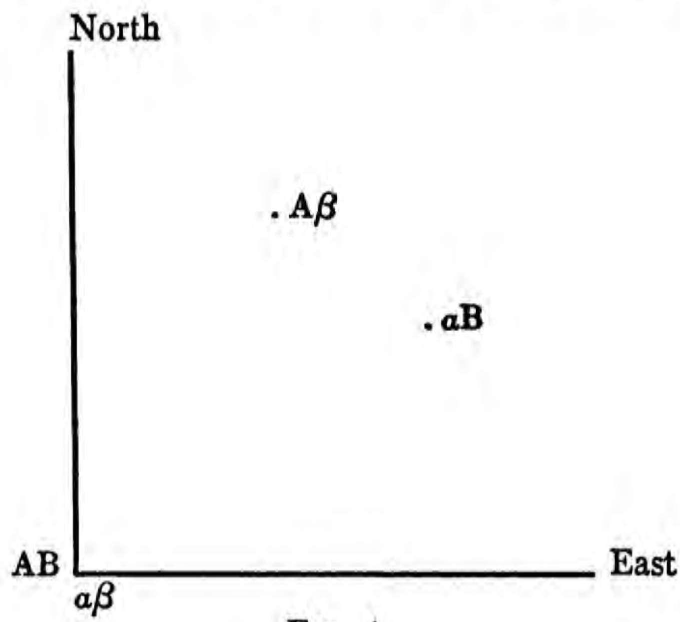

FIG. 1

Figure 1 represents an "ordinary" bargaining situation if we adopt the rule that North and East must reach explicit agreement before they choose. $A\beta$ and αB can be thought of as alternative agreements that they may reach, while AB and $\alpha\beta$, with zero values for both persons, can be interpreted as the bargaining equivalent of "no sale." Whoever can first commit himself wins. If North can commit himself to A he will secure $A\beta$, since he leaves East a choice between $A\beta$ and AB and the former is obviously East's choice under the circumstances. If East could have committed himself first to B, however, North would have been restricted to a choice of αB or no agreement (that is, of αB or AB) and would have agreed to αB. As a matter of fact, first commitment is a kind of "first move"; and in a game with the same numbers but with moves in turn, first move would be an advantage. If, by mistake, both parties get committed, North to A and East to B, they lock themselves in stalemate at AB.

Figure 2 illustrates a deterrent threat if we interpret AB as the *status quo,* with North planning a shift to α (leading to αB) and East threatening a shift to β (resulting in $\alpha\beta$) if he does. If North moves first, East can only lose by moving to β, and similarly if North can commit himself to α before East can make his threat;

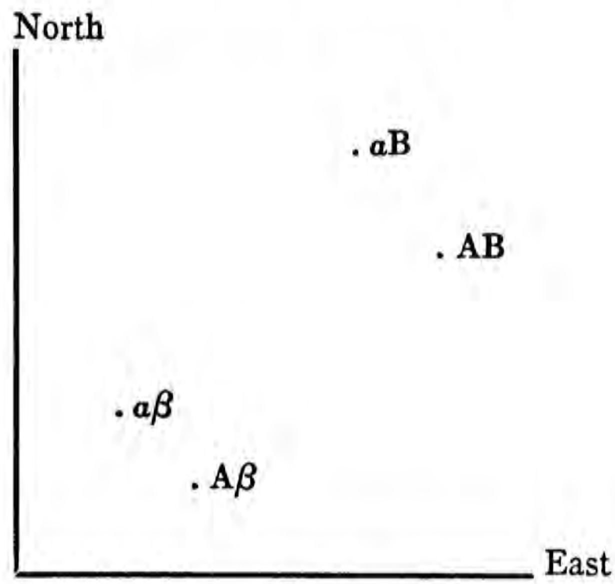

FIG. 2

but if East can effectively threaten the mutually undesirable $\alpha\beta$, he leaves North only a choice of $\alpha\beta$ or AB and North chooses the latter. Note that it is not sufficient for East to commit his *choice* in advance, as it was in Figure 1; he must commit himself to a *conditional* choice, B or β depending on whether North chooses A or α. If East committed his choice he would obtain only the advantage of "first move"; and in the present game, if moves were in turn, North would win at αB regardless of who moved first. (East would choose B rather than β, to leave North a choice of αB or AB rather than of $\alpha\beta$ or $A\beta$; and North would take αB. North, with first move, would choose α rather than A, leaving East $\alpha\beta$ or αB rather than $A\beta$ or AB; East would take αB.)

Figure 3 illustrates the promise. Whoever goes first, or even if moves are simultaneous, αB is a "minimax"; either can achieve it by himself, and neither can threaten the other with anything worse. Both would, however, prefer $A\beta$ to αB; but to reach $A\beta$

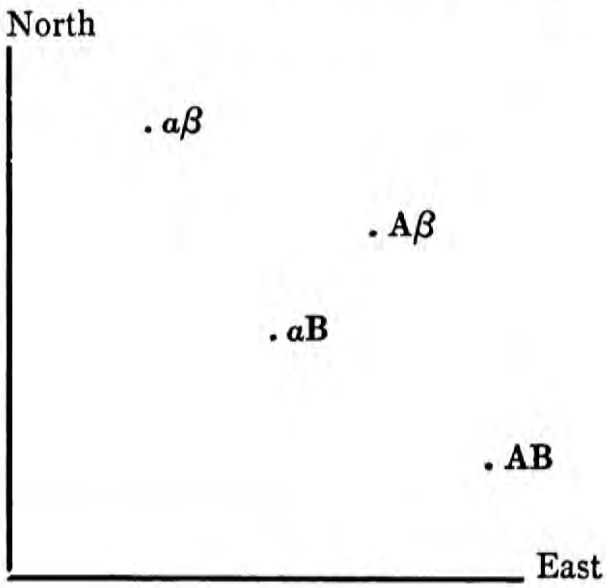

FIG. 3

they must trust each other or be able to make enforceable promises. Whoever goes first, the other has an incentive to cheat; if North chooses A, East can take AB, and if East chooses β first, North can choose $a\beta$. If moves are simultaneous each has an incentive to cheat, and each may expect the other to cheat; and either deliberate cheating, or self-protection against the other's incentive to cheat, indicates choices of a and B. At least one party

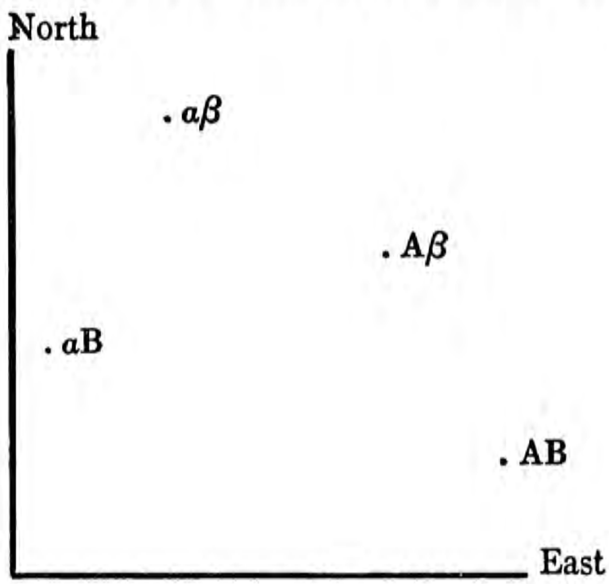

FIG. 4

must be able to commit himself to abstention; then the other can move first. If both must move simultaneously, both must be able to make enforceable promises.

Figure 4 is the same as Fig. 3 except that αB has been moved leftward. Here, in the absence of communication, North wins at $\alpha\beta$ regardless of whether he or East moves first or moves are simultaneous. If, however, East can communicate a *conditional* commitment, he can force North to choose A and an outcome of $A\beta$. But this commitment is something more than either a promise or a threat; it is both a promise and a threat. He must threaten αB if North chooses α; and he must promise "not AB" if North chooses A. The threat alone will not induce North to avoid α; αB is better than AB for North, and AB is what he gets with A if East is free to choose B. East must commit himself to do, for either α or A, the opposite of what he would do if he were not committed: abstention from AB or immolation at αB.

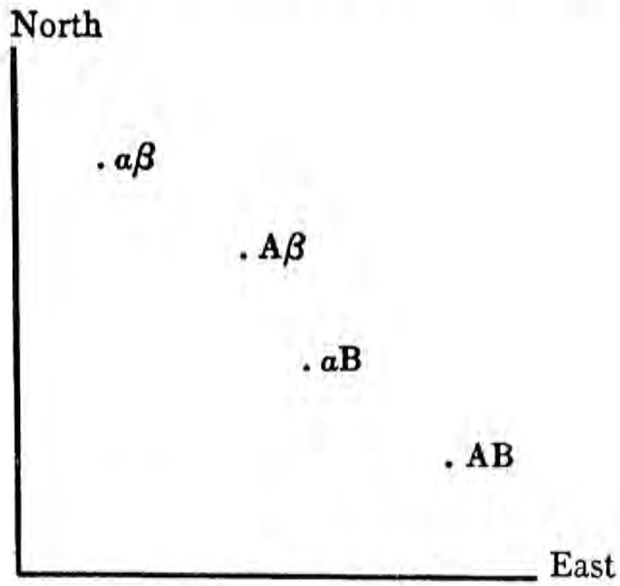

FIG. 5

Finally, Figs. 5 and 6 show two games that separately contain nothing of interest but together make possible an extortionate threat. Figure 5 has a minimax solution at αB; either can achieve αB, neither can enforce anything better, no collaboration is possible, no threat can be made. Figure 6, though contrasting with Fig. 5 in the identity of interest between the two parties, is similarly devoid of any need for collaboration or communication or any possible threat to exploit. With or without communication, with or without an order of moves, the outcome is at AB.

But suppose the two games are simultaneously up for decision,

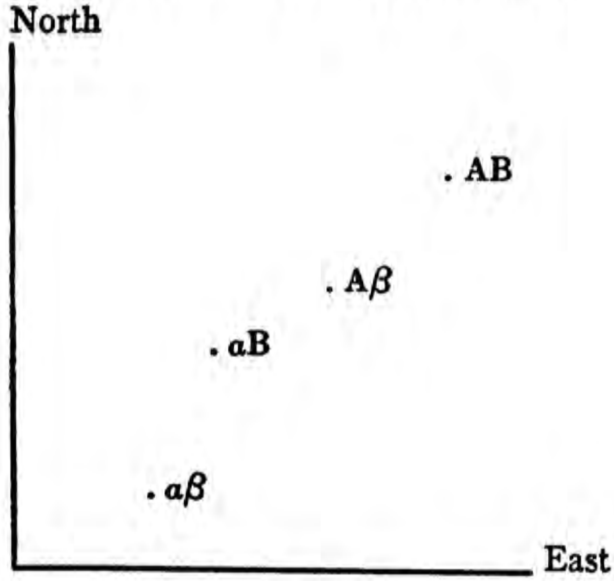

FIG. 6

and the same two parties are involved in both. If either party can commit himself to a threat he may improve his position. East, for example, could threaten to choose β rather than B in game 6, unless North chose A rather than α in game 5; alternatively, North could threaten α in game 6 unless East chose β in game 5. Assuming the intervals large enough in game 6, and the threat persuasively committed and communicated, the threatener gains in game 5 at no cost in game 6. Because his threat succeeds he does not carry it out; so he gets AB in 6 as well as his preferred choice in game 5. To express this result differently, game 6 supplies what was ruled out earlier, namely, the threat of an act "outside the game." From the point of view of game 5, game 6 is an extraneous act, and East might as well threaten to burn North's house down if he does not choose A in 5. But such purely extortionate threats are not always easy to make; they often require an occasion, an object, and a means of communication, and additionally often suffer from illegality, immorality, or resistance out of sheer stubbornness. The joining of two negotiations on the same agenda may thus succeed where a purely gratuitous threat would be impracticable.

If North cannot commit himself to a threat, and consequently desires only to prevent a threat by East, it is in his interest that communication be impossible; or if communication occurs, it is in his interest that the two games not be placed on the same agenda; or if he cannot prevent their being discussed together by East, it is in his interest to turn each game over to a different agent whose compensation depends only on the outcome of his own game. If North can force game 6 to be played first, and is unable to commit himself in response to a threat, the threat is obviated. If he can commit his choice in game 5 before the threat is made, he is safe. But if he can commit himself in game 5, and game 6 is to be played first, East could threaten to choose β in game 6 unless North assumed a prior commitment to A in game 5; in this case North's ability to commit himself is a disadvantage, since it permits him to be forced into "playing" game 5 ahead of 6.

Incidentally, dropping AB vertically in Fig. 2 to below the level $\alpha\beta$ would illustrate an important principle, namely, that moving one point in a manner "unfavorable" to North may actually improve the outcome for him. The threat that kept him from winning in Fig. 2 depends on the comparative attractiveness of AB over $\alpha\beta$ for North; if AB is made worse for him than $\alpha\beta$ he becomes immune to the threat, which then is not made, and he wins at αB. This is an abstract example of the principle that, in bargaining, weakness may be strength.