

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.

¹The 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.

4

ELEMENTS OF A THEORY OF STRATEGY

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

6**ELEMENTS OF A THEORY OF STRATEGY**

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 deprecate 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

²There 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

8**ELEMENTS OF A THEORY OF STRATEGY**

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 *nomise* 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

³The 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).

10 ELEMENTS OF A THEORY OF STRATEGY

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

* 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

⁶Progress 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.

14 ELEMENTS OF A THEORY OF STRATEGY

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, communica-

tion 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

⁷In 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."

16 ELEMENTS OF A THEORY OF STRATEGY

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-

*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.

* Compare p. 239 below.

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).

24 ELEMENTS OF A THEORY OF STRATEGY

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 withdraws 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

* 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

26 ELEMENTS OF A THEORY OF STRATEGY

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

³The 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 know 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 cus-

tom 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.

34 ELEMENTS OF A THEORY OF STRATEGY

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-

*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.

36 ELEMENTS OF A THEORY OF STRATEGY

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.

38 ELEMENTS OF A THEORY OF STRATEGY

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.

40 ELEMENTS OF A THEORY OF STRATEGY

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 un-wisdom 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.

42 ELEMENTS OF A THEORY OF STRATEGY

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.

44 ELEMENTS OF A THEORY OF STRATEGY

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.

46 ELEMENTS OF A THEORY OF STRATEGY

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 , $A\beta$, αB , and $\alpha\beta$. 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.

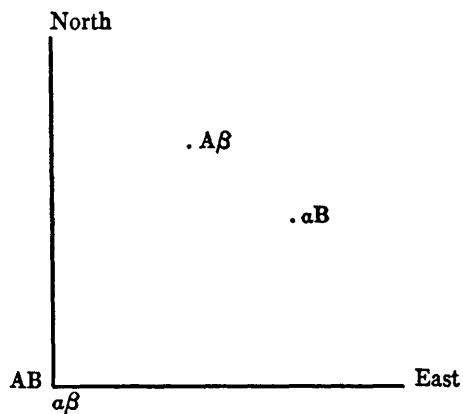


FIG. 1

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 aB can be thought of as alternative agreements that they may reach, while AB and $a\beta$, 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 aB or no agreement (that is, of aB or AB) and would have agreed to aB . 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 ad-

48 ELEMENTS OF A THEORY OF STRATEGY

vantage. 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 a (leading to aB) and East threatening a shift to β (resulting in $a\beta$) if he does. If North moves first, East can only lose by moving to β , and similarly if North can commit himself to a before East can make his threat;

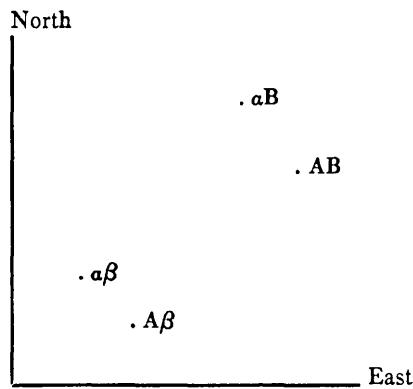


FIG. 2

but if East can effectively threaten the mutually undesirable $a\beta$, he leaves North only a choice of $a\beta$ 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 a . 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 aB regardless of who moved first. (East would choose B rather than β , to leave North a choice of aB or AB rather than of $a\beta$ or $A\beta$; and North would take aB . North, with first move, would choose a rather than A , leaving East $a\beta$ or aB rather than $A\beta$ or AB ; East would take aB .)

Figure 3 illustrates the promise. Whoever goes first, or even if moves are simultaneous, aB 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 aB ; but to reach $A\beta$

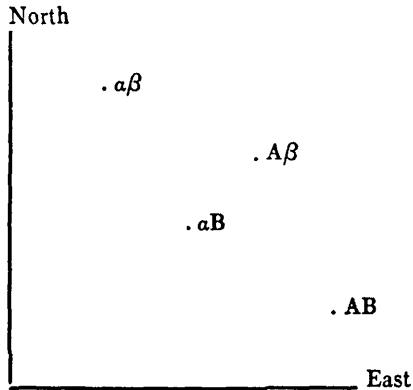


FIG. 3

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

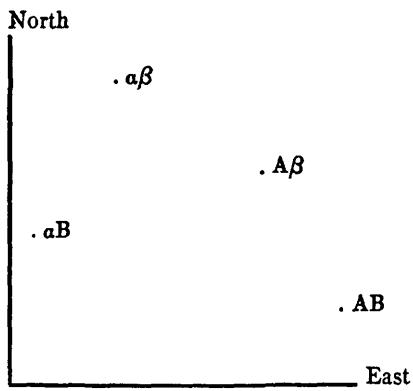


FIG. 4

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 .

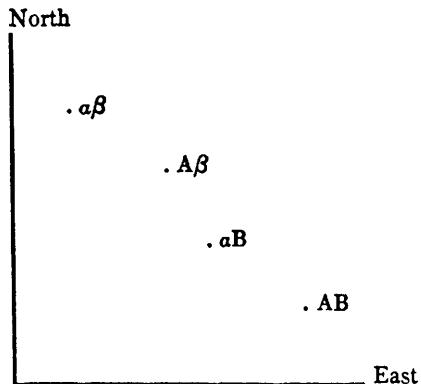


FIG. 5

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,

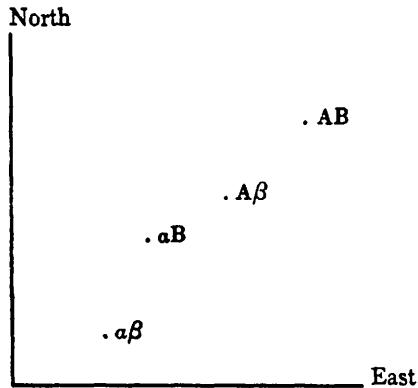


FIG. 6

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 $a\beta$ 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 $a\beta$ for North; if AB is made worse for him than $a\beta$ he becomes immune to the threat, which then is not made, and he wins at aB . This is an abstract example of the principle that, in bargaining, weakness may be strength.

BARGAINING, COMMUNICATION, AND LIMITED WAR

Limited war requires limits; so do strategic maneuvers if they are to be stabilized short of war. But limits require agreement or at least some kind of mutual recognition and acquiescence. And agreement on limits is difficult to reach, not only because of the uncertainties and the acute divergence of interests but because negotiation is severely inhibited both during war and before it begins and because communication becomes difficult between adversaries in time of war. Furthermore, it may seem to the advantage of one side to avoid agreement on limits, in order to enhance the other's fear of war; or one side or both may fear that even a show of willingness to negotiate will be interpreted as excessive eagerness.

The study of tacit bargaining — bargaining in which communication is incomplete or impossible — assumes importance, therefore, in connection with limited war, or, for that matter, with limited competition, jurisdictional maneuvers, jockeying in a traffic jam, or getting along with a neighbor that one does not speak to. The problem is to develop a modus vivendi when one or both parties either cannot or will not negotiate explicitly or when neither would trust the other with respect to any agreement explicitly reached. The present chapter will examine some of the concepts and principles that seem to underlie tacit bargaining and will attempt to draw a few illustrative conclusions about the problem of limited war or analogous situations. It will also suggest that these same principles may often provide a powerful clue to understanding even the logically dissimilar case of explicit bargaining with full communication and enforcement.

54 ELEMENTS OF A THEORY OF STRATEGY

The most interesting situations and the most important are those in which there is a conflict of interest between the parties involved. But it is instructive to begin with the special simplified case in which two or more parties have identical interests and face the problem not of reconciling interests but only of coordinating their actions for their mutual benefit, when communication is impossible. This special case brings out clearly the principle that will then serve to solve the problem of tacit "bargaining" over conflicting preferences.

TACIT COORDINATION (COMMON INTERESTS)

When a man loses his wife in a department store without any prior understanding on where to meet if they get separated, the chances are good that they will find each other. It is likely that each will think of some obvious place to meet, so obvious that each will be sure that the other is sure that it is "obvious" to both of them. One does not simply predict where the other will go, since the other will go where he predicts the first to go, which is wherever the first predicts the second to predict the first to go, and so ad infinitum. Not "What would I do if I were she?" but "What would I do if I were she wondering what she would do if she were I wondering what I would do if I were she . . . ?" What is necessary is to coordinate predictions, to read the same message in the common situation, to identify the one course of action that their expectations of each other can converge on. They must "mutually recognize" some unique signal that coordinates their expectations of each other. We cannot be sure they will meet, nor would all couples read the same signal; but the chances are certainly a great deal better than if they pursued a random course of search.

The reader may try the problem himself with the adjoining map (Fig. 7). Two people parachute unexpectedly into the area shown, each with a map and knowing the other has one, but neither knowing where the other has dropped nor able to communicate directly. They must get together quickly to be rescued. Can they study their maps and "coordinate" their behavior? Does the map suggest some particular meeting place so unambiguously

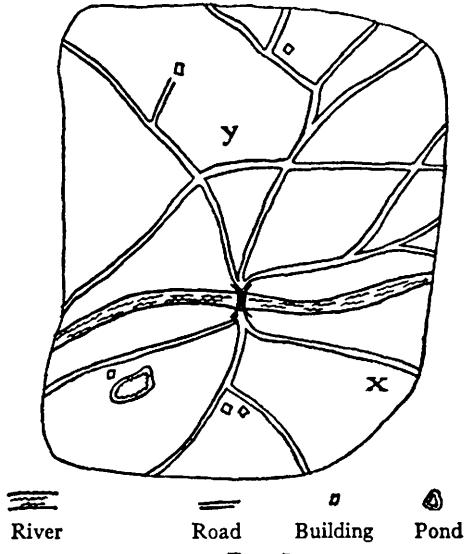


FIG. 7

that each will be confident that the other reads the same suggestion with confidence?

The writer has tried this and other analogous problems on an unscientific sample of respondents; and the conclusion is that people often can coordinate. The following abstract puzzles are typical of those that can be "solved" by a substantial proportion of those who try. The solutions are, of course, arbitrary to this extent: any solution is "correct" if enough people think so. The reader may wish to confirm his ability to concert in the following problems with those whose scores are given in a footnote.¹

¹ In the writer's sample, 36 persons concerted on "heads" in problem 1, and only 6 chose "tails." In problem 2, the first three numbers were given 37 votes out of a total of 41; the number 7 led 100 by a slight margin, with 13 in third place. The upper left corner in problem 3 received 24 votes out of a total of 41, and all but 3 of the remainder were distributed in the same diagonal line. Problem 4, which may reflect the location of the sample in New Haven, Connecticut, showed an absolute majority managing to get together at Grand Central Station (information booth), and virtually all of them succeeded in meeting at 12 noon. Problem 6 showed a variety of answers, but two-fifths of all persons succeeded inconcerting on the number 1; and in problem 7, out of 41 people, 12 got together on \$1,000,000, and only 3 entries consisted of numbers that were not a power of 10; of those 3, 2 were \$64 and, in the

56 ELEMENTS OF A THEORY OF STRATEGY

1. Name "heads" or "tails." If you and your partner name the same, you both win a prize.
2. Circle one of the numbers listed in the line below. You win if you all succeed in circling the same number.

7 100 13 261 99 555

3. Put a check mark in one of the sixteen squares. You win if you all succeed in checking the same square.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. You are to meet somebody in New York City. You have not been instructed where to meet; you have no prior understanding with the person on where to meet; and you cannot communicate with each other. You are simply told that you will have to guess where to meet and that he is being told the same thing and that you will just have to try to make your guesses coincide.

5. You were told the date but not the hour of the meeting in No. 4; the two of you must guess the exact minute of the day for meeting. At what time will you appear at the meeting place that you elected in No. 4?

6. Write some positive number. If you all write the same number, you win.

7. Name an amount of money. If you all name the same amount, you can have as much as you named.

8. You are to divide \$100 into two piles, labeled A and B.

more up-to-date version, \$64,000! Problem 8 caused no difficulty to 36 out of 41, who split the total fifty-fifty. Problem 9 secured a majority of 20 out of 22 for Robinson. An alternative formulation of it, in which Jones and Robinson were tied on the first ballot at 28 votes each, was intended by the author to demonstrate the difficulty of concerting in case of tie; but the respondents surmounted the difficulty and gave Jones 16 out of 18 votes (apparently on the basis of Jones's earlier position on the list), proving the main point but overwhelming the subsidiary point in the process. In the map most nearly like the one reproduced here (Fig. 1), 7 out of 8 respondents managed to meet at the bridge.

Your partner is to divide another \$100 into two piles labeled A and B. If you allot the same amounts to A and B, respectively, that your partner does, each of you gets \$100; if your amounts differ from his, neither of you gets anything.

9. On the first ballot, candidates polled as follows:

Smith	19	Robinson	29
Jones	28	White	9
Brown	15		

The second ballot is about to be taken. You have no interest in the outcome, except that you will be rewarded if someone gets a majority on the second ballot and you vote for the one who does. Similarly, all voters are interested only in voting with the majority, and everybody knows that this is everybody's interest. For whom do you vote on the second ballot?

These problems are artificial, but they illustrate the point. People *can* often concert their intentions or expectations with others if each knows that the other is trying to do the same. Most situations—perhaps every situation for people who are practiced at this kind of game—provide some clue for coordinating behavior, some focal point for each person's expectation of what the other expects him to expect to be expected to do. Finding the key, or rather finding *a* key—any key that is mutually recognized as the key becomes *the* key—may depend on imagination more than on logic, it may depend on analogy, precedent, accidental arrangement, symmetry, aesthetic or geometric configuration, casuistic reasoning, and who the parties are and what they know about each other. Whimsy may send the man and his wife to the "lost and found"; or logic may lead each to reflect and to expect the other to reflect on where they would have agreed to meet if they had had a prior agreement to cover the contingency. It is not being asserted that they will always find an obvious answer to the question; but the chances of their doing so are ever so much greater than the bare logic of abstract random probabilities would ever suggest.

A prime characteristic of most of these "solutions" to the problems, that is, of the clues or coordinators or focal points, is some kind of prominence or conspicuousness. But it is a promi-

58 ELEMENTS OF A THEORY OF STRATEGY

nence that depends on time and place and who the people are. Ordinary folk lost on a plane circular area may naturally go to the center to meet each other; but only one versed in mathematics would "naturally" expect to meet his partner at the center of gravity of an irregularly shaped area. Equally essential is some kind of uniqueness; the man and his wife cannot meet at the "lost and found" if the store has several. The writer's experiments with alternative maps indicated clearly that a map with many houses and a single crossroads sends people to the crossroads, while one with many crossroads and a single house sends most of them to the house. Partly this may reflect only that uniqueness conveys prominence; but it may be more important that uniqueness avoids ambiguousness. Houses may be intrinsically more prominent than anything else on the map; but if there are three of them, none more prominent than the others, there is but one chance in three of meeting at a house, and the recognition of this fact may lead to the rejection of houses as the "clue."²

But in the final analysis we are dealing with imagination as much as with logic; and the logic itself is of a fairly casuistic kind. Poets may do better than logicians at this game, which is perhaps more like "puns and anagrams" than like chess. Logic helps — the large plurality accorded to the number 1 in problem 6 seems to rest on logic — but usually not until imagination has selected some clue to work on from among the concrete details of the situation.

TACIT BARGAINING (DIVERGENT INTERESTS)

A conflict of interest enters our problem if the parachutists dislike walking. With communication, which is not allowed in our problem, they would have argued or bargained over where to meet, each favoring a spot close to himself or a resting place particularly to his liking. In the absence of communication, their overriding interest is to concert ideas; and if a particular spot

²That this would be "correct" reasoning, incidentally, is suggested by one of the author's map experiments. On a map with a single house and many crossroads, the eleven people who chose the house all met, while the four who chose crossroads all chose different crossroads and did not even meet one another.

commands attention as the "obvious" place to meet, the winner of the bargain is simply the one who happens to be closer to it. Even if the one who is farthest from the focal point knows that he is, he cannot withhold his acquiescence and argue for a fairer division of the walking; the "proposal" for the bargain that is provided by the map itself — if, in fact, it provides one — is the only extant offer; and without communication, there is no counterproposal that can be made. The conflict gets reconciled — or perhaps we should say ignored — as a by-product of the dominant need for coordination.

"Win" and "lose" may not be quite accurate, since both may lose by comparison with what they could have agreed on through communication. If the two are actually close together and far from the lone house on the map, they might have eliminated the long walk to the house if they could have identified their locations and concerted explicitly on a place to meet between them. Or it may be that one "wins" while the other loses more than the first wins: if both are on the same side of the house and walk to it, they walk together a greater distance than they needed to, but the closer one may still have come off better than if he had had to argue it out with the other.

This last case illustrates that it may be to the advantage of one to be unable to communicate. There is room here for a motive to destroy communication or to refuse to collaborate in advance on a method of meeting if one is aware of his advantage and confident of the "solution" he foresees. In one variant of the writer's test, A knew where B was, but B had no idea where A was (and each knew how much the other knew). Most of the recipients of the B-type questionnaire smugly sat tight, enjoying their ignorance, while virtually all the A-questionnaire respondents grimly acknowledged the inevitable and walked all the way to B. Better still may be to have the power to send but not to receive messages: if one can announce his position and state that his transmitter works but not his receiver, saying that he will wait where he is until the other arrives, the latter has no choice. He can make no effective counteroffer, since no counteroffer could be heard.⁸

⁸This is an instance of the general paradox, illustrated at length in Chap-

60 ELEMENTS OF A THEORY OF STRATEGY

The writer has tried a sample of conflicting-interest games on a number of people, including games that are biased in favor of one party or the other; and on the whole, the outcome suggests the same conclusion that was reached in the purely cooperative games. All these games require coordination; they also, however, provide several alternative choices over which the two parties' interests differ. Yet, among all the available options, some particular one usually seems to be the focal point for coordinated choice, and the party to whom it is a relatively unfavorable choice quite often takes it simply because he knows that the other will expect him to. The choices that cannot coordinate expectations are not really "available" without communication. The odd characteristic of all these games is that neither rival can gain by outsmarting the other. Each loses unless he does exactly what the other expects him to do. Each party is the prisoner or the beneficiary of their mutual expectations; no one can disavow his own expectation of what the other will expect him to expect to be expected to do. The need for agreement overrules the potential disagreement, and each must concert with the other or lose altogether. Some of these games are arrived at by slightly changing the problems given earlier, as we did for the map problem by supposing that walking is onerous.

1. A and B are to choose "heads" or "tails" without communicating. If both choose "heads," A gets \$3 and B gets \$2; if both choose "tails," A gets \$2 and B gets \$3. If they choose differently, neither gets anything. You are A (or B); which do you choose? (Note that if both choose at random, there is only a 50-50 chance of successful coincidence and an expected value of \$1.25 apiece — less than either \$3 or \$2.)

2. You and your two partners (or rivals) each have one of the letters A, B, and C. Each of you is to write these three letters, A, B, and C, in any order. If the order is the same on all three of your lists, you get prizes totaling \$6, of which \$3 goes to the one whose letter is first on all three lists, \$2 to the one whose letter is second, and \$1 to the person whose letter is third. If the letters are not in identical order on all three lists, none of

ter 2, that what is impotence by ordinary standards may, in bargaining, be a source of "strength."

you gets anything. Your letter is A (or B, or C); write here the three letters in the order you choose:

_____, _____, _____.

3. You and your partner (rival) are each given a piece of paper, one blank and the other with an "X" written on it. The one who gets the "X" has the choice of leaving it alone or erasing it; the one who gets the blank sheet has the choice of leaving it blank or writing an "X" on it. If, when you have made your choices without communicating, there is an "X" on only one of the sheets, the holder of the "X" gets \$3 and the holder of the blank sheet gets \$2. If both sheets have "X's" or both sheets are blank, neither gets anything. Your sheet of paper has the original "X" on it; do you leave it alone or erase it? (*Alternate*: your sheet of paper is the blank one; do you leave it blank or write an "X"?)

4. You and your partner (rival) are to be given \$100 if you can agree on how to divide it without communicating. Each of you is to write the amount of his claim on a sheet of paper; and if the two claims add to no more than \$100, each gets exactly what he claimed. If the two claims exceed \$100, neither of you gets anything. How much do you claim? \$_____.

5. You and your partner are each to pick one of the five letters, K, G, W, L, or R. If you pick the same letter, you get prizes; if you pick different letters, you get nothing. The prizes you get depend on the letter you both pick; but the prizes are not the same for each of you, and the letter that would yield you the highest prize may or may not be his most profitable letter. For you the prizes would be as follows:

K	\$4	L	\$2
G	\$3	R	\$5
W	\$1		

You have no idea what his schedule of prizes looks like. You begin by proposing to him the letter R, that being your best letter. Before he can reply, the master-of-ceremonies intervenes to say that you were not supposed to be allowed to communicate and that any further communication will disqualify you both. You must simply write down one of the letters, hoping that the other chooses the same letter. Which letter do you

62 ELEMENTS OF A THEORY OF STRATEGY

choose? (Alternate formulation for the second half of the sample shows schedule of K-\$3, G-\$1, W-\$4, L-\$5, R-\$2, and has the "other" party make the initial proposal of the letter R before communication is cut off.)

6. Two opposing forces are at the points marked *X* and *Y* in a map similar to the one in Fig. 7. The commander of each force wishes to occupy as much of the area as he can and knows the other does too. But each commander wishes to avoid an armed clash and knows the other does too. Each must send forth his troops with orders to take up a designated line and to fight if opposed. Once the troops are dispatched, the outcome depends only on the lines that the two commanders have ordered their troops to occupy. If the lines overlap, the troops will be assumed to meet and fight, to the disadvantage of both sides. If the troops take up positions that leave any appreciable space unoccupied between them, the situation will be assumed "unstable" and a clash inevitable. Only if the troops are ordered to occupy identical lines or lines that leave virtually no unoccupied space between them will a clash be avoided. In that case, each side obtains successfully the area it occupies, the advantage going to the side that has the most valuable area in terms of land and facilities. You command the forces located at the point marked *X* (*Y*). Draw on the map the line that you send your troops to occupy.

7. A and B have incomes of \$100 and \$150 per year, respectively. They are notified of each other's income and told that they must begin paying taxes totaling \$25 per year. If they can reach agreement on shares of this total, they may share the annual tax bill in whatever manner they agree on. But they must reach agreement without communication; each is to write down the share he proposes to pay, and if the shares total \$25 or more, each will pay exactly what he proposed. If the proposed shares fail to add up to \$25, however, each will individually be required to pay the full \$25, and the tax collectors will keep the surplus. You are A (B); how much do you propose to pay? \$_____.

8. A loses some money, and B finds it. Under the house rules, A cannot have his money back until he agrees with the finder on a suitable reward, and B cannot keep any except what A agrees to. If no agreement is reached, the money goes to the house. The

amount is \$16, and A offers \$2 as a reward. B refuses, demanding half the money for himself. An argument ensues, and the house intervenes, insisting that each write his claim, once and for all, without further communication. If the claims are consistent with the \$16 total, each will receive exactly what he claims; but if together they claim more than \$16, the funds will be confiscated by the house. As they sit pondering what claims to write, a well-known and respected mediator enters and offers to help. He cannot, he says, participate in any bargaining, but he can make a "fair" proposal. He approaches A and says, "I think a reasonable division under the circumstances would be a 2-1 split, the original owner getting two-thirds and the finder one-third, perhaps rounded off to \$11 and \$5, respectively. I shall make the same suggestion to him." Without waiting for any response, he approaches the finder, makes the same suggestion, and says that he made the same suggestion to the original owner. Again without waiting for any response, he departs. You are A (B); what claim do you write?

The outcomes in the writer's informal sample are given in the footnote.⁴ In those problems where there is some asymmetry between "you" and "him," that is, between A and B, the A formulations were matched with the B formulations in deriving

⁴In the first problem, 16 out of 22 A's and 15 out of 22 B's chose heads. Given what the A's did, heads was the best answer for B; given what the B's did, heads was the best answer for A. Together they did substantially better than at random; and, of course, if each had tried to win \$3, they would all have scored a perfect zero. Problem 2, however, which is logically similar to 1 but with a more compelling structure, showed 9 out of 12 A's, 10 out of 12 B's, and 14 out of 16 C's, successfully co-ordinating on ABC. (Of the remaining 7, incidentally, 5 discriminated against themselves in departing from alphabetical order, all to no avail.) Problem 3, which is structurally analogous to 1, showed 18 out of 22 A's concerting successfully with 14 out of 19 B's, giving A the \$3 prize. In problem 4, 36 out of 40 chose \$50. (Two of the remainder were \$49 and \$49.99.) In problem 5 the letter R won 5 out of 8 votes from those who had proposed it, and 8 out of 9 votes from those who were on the other side. In problem 6, 14 of 22 X's and 14 of 23 Y's drew their boundaries exactly along the river. The "correctness" of this solution is emphatically shown by the fact that the other 15, who eschewed the river, produced 14 different lines. Of 8×7 possible pairs among them, there were 55 failures and 1 success. Problem 7 showed 5 out of 6 of those with incomes of \$150 and 7 out of 10 of those with incomes of \$100 concerting on a 15-10 division of the tax. In problem 8 both those who lost money and those who found it, 8 and 7 persons respectively, unanimously concerting on the mediator's suggestion of an even \$5 reward.

64 ELEMENTS OF A THEORY OF STRATEGY

the "outcome." The general conclusion, as given in more detail in the footnote, is that the participants can "solve" their problem in a substantial proportion of the cases; they certainly do conspicuously better than any chance methods would have permitted, and even the disadvantaged party in the biased games permits himself to be disciplined by the message that the game provides for their coordination.

The "clues" in these games are diverse. Heads apparently beat tails through some kind of conventional priority, similar to the convention that dictates A, B, C, though not nearly so strong. The original X beats the blank sheet, apparently because the "status quo" is more obvious than change. The letter R wins because there is nothing to contradict the first offer. Roads might seem, in principle, as plausible as rivers, especially since their variety permits a less arbitrary choice. But, precisely because of their variety, the map cannot say *which* road; so roads must be discarded in favor of the unique and unambiguous river. (Perhaps in a symmetrical map of uniform terrain, the outcome would be more akin to the 50-50 split in the \$100 example—a diagonal division in half, perhaps—but the irregularity of the map rather precludes a geometrical solution.)

The tax problem illustrates a strong power of suggestion in the income figures. The abstract logic of this problem is identical with that of the \$100 division; in fact, it could be reworded as follows: each party pays \$25 in taxes, and a refund of \$25 is available to be divided among the two parties if they can agree on how to divide it. This formulation is logically equivalent to the one in problem 7, and, as such, it differs from problem 4 only in the amount of \$25 instead of \$100. Yet the inclusion of income figures, just by suggesting their relevance and making them prominent in the problem, shifts the focal point substantially to a 10-15 split rather than 12.5-12.5. And why, if incomes are relevant, is a perfectly proportional tax so obvious, when perhaps there are grounds for graduated rates? The answer must be that no particular graduation of rates is so obvious as to go without saying; and if speech is impossible, by default the uniquely simple and recognizable principle of proportionality has to be adopted. First the income figures take the initial plausibility away from a 50-50

split; then the simplicity of proportionality makes 10-15 the only one that could possibly be considered capable of tacit recognition. The same principle is displayed by an experiment in which question 7 was deliberately cluttered up with *additional* data — on family size, spending habits, and so on. Here the unique attraction of the income-proportionate split apparently became so diluted that the preponderant reply from both the high-income and the low-income respondents was a simple 50-50 division of the tax. The refined signal for the income proportionate split was drowned out by "noise," and the cruder signal for equality was all that came through.

Finally, problem 8 is again logically the same as problem 4, the amount being \$16 available for two people if they can write claims that do not exceed the amount. But the institutional arrangement is discriminatory; finder and loser do not have a compelling equality in any moralistic or legalistic sense, so the 50-50 split seems not quite obvious. The suggestion of the mediator provides the only other signal that is visible; its potency as a coordinator is seen even in the rounding to \$11 and \$5, which was universally accepted.

In each of these situations the outcome is determined by something that is fairly arbitrary. It is not a particularly "fair" outcome, from either an observer's point of view or the points of view of the participants. Even the 50-50 split is arbitrary in its reliance on a kind of recognizable mathematical purity; and if it is "fair," it is so only because we have no concrete data by which to judge its unfairness, such as the source of the funds, the relative need of the rival claimants, or any potential basis for moral or legal claims. Splitting the difference in an argument over kidnap ransom is not particularly "fair," but it has the mathematical qualities of problem 4.

If we ask what determines the outcome in these cases, the answer again is in the coordination problem. Each of these problems requires coordination for a common gain, even though there is rivalry among alternative lines of common action. But, among the various choices, there is usually one or only a few that can serve as coordinator. Take the case of the first offer in problem 5. The strongest argument in favor of R is the rhetorical question,

66 ELEMENTS OF A THEORY OF STRATEGY

"If not R, what then?" There is no answer so obvious as to give more than a random chance of concerting, even if both parties wanted to eschew the letter R after the first offer was made. To illustrate the force of this point, suppose that the master-of-ceremonies in that problem considered the first offer already to have spoiled the game and thought he might confuse the players by announcing the reversal of their prize schedules. A will get whatever prize B would have gotten, and B will get the prizes shown in A's schedule in problem 5. Does the original offerer of R have any reason to change his choice? Or suppose that the master-of-ceremonies announced that the prizes would be the same, no matter what letter were chosen, so long as they both picked the same letter. They will still rally to R as the only indicated means of coordinating choices. If we revert to the beginning of this game and suppose that the original proposal of R never got made, we might imagine a sign on the wall saying, "In case of doubt always choose R; this sign is visible to all players and constitutes a means of coordinating choices." Here we are back at the man and his wife in the department store, whose problems are over when they see a conspicuous sign that says, "The management suggests that all persons who become separated meet each other at the information booth in the center of the ground floor." Beggars cannot be choosers about the source of their signal, or about its attractiveness compared with others that they can only wish were as conspicuous.

The irony would be complete if, in game 5, your rival knew your prize schedule and you did not know his (as was the case in a variant of question 5 used in some questionnaires). Since you have no basis for guessing his preference and could not even do him a favor or make a "fair" compromise if you wished to, the only basis for concerting is to see what message you can both read in your schedule. Your own preferred letter seems the indicated choice; it is hard to see why to pick any other or which other to pick, since you have no basis for knowing what other letter is better for him than R itself. His knowledge of your preference, combined with your ignorance of his and the lack of any alternative basis for coordination, puts on him the responsibility of simply choosing in your favor. (This, in fact, was the preponderant

result among the small sample tested.) It is the same situation as when only one parachutist knew where the other was.⁵

EXPLICIT BARGAINING

The concept of "coordination" that has been developed here for tacit bargaining does not seem directly applicable to explicit bargaining. There is no apparent need for intuitive rapport when speech can be used; and the adventitious clues that coordinated thoughts and influenced the outcome in the tacit case revert to the status of incidental details.

Yet there is abundant evidence that some such influence is powerfully present even in explicit bargaining. In bargains that involve numerical magnitudes, for example, there seems to be a strong magnetism in mathematical simplicity. A trivial illustration is the tendency for the outcomes to be expressed in "round numbers"; the salesman who works out the arithmetic for his "rock-bottom" price on the automobile at \$2,507.63 is fairly pleading to be relieved of \$7.63. The frequency with which final agreement is precipitated by an offer to "split the difference" illustrates the same point, and the difference that is split is by no means always trivial. More impressive, perhaps, is the remarkable frequency with which long negotiations over complicated quantitative formulas or *ad hoc* shares in some costs or benefits converge ultimately on something as crudely simple as equal shares, shares proportionate to some common magnitude (gross national product, population, foreign-exchange deficit, and so forth), or the shares agreed on in some previous but logically irrelevant negotiation.⁶

Precedent seems to exercise an influence that greatly exceeds its logical importance or legal force. A strike settlement or an international debt settlement often sets a "pattern" that is fol-

⁵ And it is another example of the power that resides in "weakness," which was commented on in an earlier footnote.

⁶ From a great variety of formulas proposed for the contributions to UNRRA, the winner that emerged was a straight 1 per cent of gross national product—the simplest conceivable formula and the roundest conceivable number. This formula was, to be sure, the preferred position of the United States during the discussion; but that fact perhaps adds as much to the example as it detracts from it.

68 ELEMENTS OF A THEORY OF STRATEGY

lowed almost by default in subsequent negotiations. Sometimes, to be sure, there is a reason for a measure of uniformity, and sometimes there is enough similarity in the circumstances to explain similar outcomes; but more often it seems that there is simply no heart left in the bargaining when it takes place under the shadow of some dramatic and conspicuous precedent.⁷ In similar fashion, mediators often display a power to precipitate agreement and a power to determine the terms of agreement; their proposals often seem to be accepted less by reason of their inherent fairness or reasonableness than by a kind of resignation of both participants. "Fact-finding" reports may also tend to draw expectations to a focus, by providing a suggestion to fill the vacuum of indeterminacy that otherwise exists: it is not the facts themselves, but the creation of a specific suggestion, that seems to exercise the influence.

There is, in a similar vein, a strong attraction to the *status quo ante* as well as to natural boundaries. Even parallels of latitude have recently exhibited their longevity as focal points for agreement. Certainly there are reasons of convenience in using rivers as the agreed stopping place for troops or using old boundaries, whatever their current relevance; but often these features of the landscape seem less important for their practical convenience than for their power to crystallize agreement.

These observations would be trivial if they meant only that bargaining results were *expressed* in simple and qualitative terms or that minor accommodations were made to round off the last few cents or miles or people. But it often looks as though the ultimate focus for agreement did not just reflect the balance of bargaining powers but provided bargaining power to one side or the other. It often seems that a cynic could have predicted the outcome on the basis of some "obvious" focus for agreement, some strong suggestion contained in the situation itself, without much regard to the merits of the case, the arguments to be made, or the pressures to be applied during the bargaining. The "obvious" place to compromise frequently seems to win by some

⁷ This and the preceding paragraph are illustrated by the speed with which a number of Middle Eastern oil-royalty arrangements converged on the 50-50 formula a few years after World War II.

kind of default, as though there is simply no rationale for settling anywhere else. Or, if the "natural" outcome is taken to reflect the relative skills of the parties to the bargain, it may be important to identify that skill as the ability to set the stage in such a way as to give prominence to some particular outcome that would be favorable. The outcome may not be so much conspicuously fair or conspicuously in balance with estimated bargaining powers as just plain "conspicuous."

This conclusion may seem to reduce the scope for bargaining skill, if the outcome is already determined by the configuration of the problem itself and where the focal point lies. But perhaps what it does is shift the locus where skill is effective. The "obvious" outcome depends greatly on how the problem is formulated, on what analogies or precedents the definition of the bargaining issue calls to mind, on the kinds of data that may be available to bear on the question in dispute. When the committee begins to argue over how to divide the costs, it is already constrained by whether the terms of reference refer to the "dues" to be shared or the "taxes" to be paid, by whether a servicing committee is preparing national-income figures or balance-of-payments figures for their use, by whether the personnel of the committee brings certain precedents into prominence by having participated personally in earlier negotiations, by whether the inclusion of two separate issues on the same agenda will give special prominence and relevance to those particular features that they have in common. Much of the skill has already been applied when the formal negotiations begin.⁸

If all this is correct, as it seems frequently to the author to be, our analysis of tacit bargaining may help to provide an understanding of the influence at work; and perhaps the logic of tacit bargaining even provides a basis for believing it to be correct. The fundamental problem in tacit bargaining is that of *coordination*; we should inquire, then, what has to be coordinated in ex-

⁸Perhaps another role for skill is contained in this general approach. If one is unsuccessful in getting the problem so formulated that the "obvious" outcome is near his own preferred position, he can proceed to confuse the issue. Find multiple definitions for all the terms and add "noise" to drown out the strong signal contained in the original formulation. The technique may not succeed, but in the variant of our income-tax problem mentioned above it certainly did.

plicit bargaining. The answer may be that explicit bargaining requires, for an ultimate agreement, some coordination of the participants' expectations. The proposition might be as follows.

Most bargaining situations ultimately involve some range of possible outcomes within which each party would rather make a concession than fail to reach agreement at all. In such a situation any potential outcome is one from which at least one of the parties, and probably both, would have been willing to retreat for the sake of agreement, and very often the other party knows it. Any potential outcome is therefore one that either party could have improved by insisting; yet he may have no basis for insisting, since the other knows or suspects that he would rather concede than do without agreement. Each party's strategy is guided mainly by what he expects the other to accept or insist on; yet each knows that the other is guided by reciprocal thoughts. The final outcome must be a point from which neither expects the other to retreat; yet the main ingredient of this expectation is what one thinks the other expects the first to expect, and so on. Somehow, out of this fluid and indeterminate situation that seemingly provides no logical reason for anybody to expect anything except what he expects to be expected to expect, a decision is reached. These infinitely reflexive expectations must somehow converge on a single point, at which each expects the other not to expect to be expected to retreat.

If we then ask what it is that can bring their expectations into convergence and bring the negotiation to a close, we might propose that it is the intrinsic magnetism of particular outcomes, especially those that enjoy prominence, uniqueness, simplicity, precedent, or some rationale that makes them qualitatively differentiable from the continuum of possible alternatives. We could argue that expectations tend not to converge on outcomes that differ only by degree from alternative outcomes but that people have to dig in their heels at a groove in order to make any show of determination. One has to have a reason for standing firmly on a position; and along the continuum of qualitatively undifferentiable positions one finds no rationale. The rationale may not be strong at the arbitrary "focal point," but at least it can defend itself with the argument "If not here, where?"

There is perhaps a little more to this need for a mutually identifiable resting place. If one is about to make a concession, he needs to control his adversary's expectations; he needs a recognizable limit to his own retreat. If one is to make a finite concession that is not to be interpreted as capitulation, he needs an obvious place to stop. A mediator's suggestion may provide it; or any other element that qualitatively distinguishes the new position from surrounding positions. If one has been demanding 60 per cent and recedes to 50 per cent, he can get his heels in; if he recedes to 49 per cent, the other will assume that he has hit the skids and will keep sliding.

If some troops have retreated to the river in our map, they will expect to be expected to make a stand. This is the one spot to which they can retreat without necessarily being expected to retreat further, while, if they yield any further, there is no place left where they can be expected to make a determined stand. Similarly, the advancing party can expect to force the other to retreat to the river without having his advance interpreted as an insatiable demand for unlimited retreat. There is stability at the river — and perhaps nowhere else.

This proposition may seem intuitively plausible; it does to the writer, and in any event some kind of explanation is needed for the tendency to settle at focal points. But the proposition would remain vague and somewhat mystical if it were not for the somewhat more tangible logic of tacit bargaining. The latter provides not only an analogy but the demonstration that the necessary psychic phenomenon — tacit coordination of expectations — is a real possibility and in some contexts a remarkably reliable one. The "coordination" of expectations is analogous to the "coordination" of behavior when communication is cut off; and, in fact, they both involve nothing more nor less than intuitively perceived mutual expectations. Thus the empirically verifiable results of some of the tacit-bargaining games, as well as the more logical role of coordinated expectations in that case, prove that expectations can be coordinated and that some of the objective details of the situation can exercise a controlling influence when the coordination of expectations is essential. *Something* is perceived by both parties when communication is absent;

it must still be perceptible, though undoubtedly of lesser force, when communication is possible. The possibility of communication does not make 50-50 less symmetrical or the river less unique or A B C a less natural order for those letters.

If all we had to reason from were the logic of tacit bargaining, it would be only a guess and perhaps a wild one that the same kind of psychic attraction worked in explicit bargaining; and if all we had to generalize from were the observation of peculiarly "plausible" outcomes in actual bargains, we might be unwilling to admit the force of adventitious details. But the two lines of evidence so strongly reinforce each other that the analogy between tacit and explicit bargaining seems a potent one.

To illustrate with the problem of agreeing explicitly on how to divide \$100: 50-50 seems a plausible division, but it may seem so for too many reasons. It may seem "fair"; it may seem to balance bargaining powers; or it may, as suggested in this paper, simply have the power to communicate its own inevitability to the two parties in such fashion that each appreciates that they both appreciate it. What our analysis of tacit bargaining provides is evidence for the latter view. The evidence is simply that if they had to divide the \$100 without communicating, they could concert on 50-50. Instead of relying on intuition, then, we can point to the fact that in a slightly different context — the tacit-bargaining context — our argument has an objectively demonstrable interpretation.

To illustrate again: the ability of the two commanders in one of our problems to recognize the stabilizing power of the river — or, rather, their inability not to recognize it — is substantiated by the evidence that if their survival depended on some agreement about where to stabilize their lines *and communication were not allowed*, they probably could perceive and appreciate the qualities of the river as a focus for their tacit agreement. So the tacit analogy at least demonstrates that the idea of "coordinating expectations" is meaningful rather than mystical.

Perhaps we could push the argument further still. Even in those cases in which the only distinguishing characteristic of a bargaining result is its evident "fairness," by standards that the participants are known to appreciate, we might argue that the

moral force of fairness is greatly reinforced by the power of a "fair" result to focus attention, if it fills the vacuum of indeterminacy that would otherwise exist. Similarly, when the pressure of public opinion seems to force the participants to the obviously "fair" or "reasonable" solution, we may exaggerate the "pressure" or at least misunderstand the way it works on the participants unless we give credit to its power to coordinate the participants' expectations. It may, to put it differently, be the power of *suggestion*, working through the mechanism described in this paper, that makes public opinion or precedent or ethical standards so effective. Again, as evidence for this view, we need only to suppose that the participants had to reach ultimate agreement without communicating and visualize public opinion or some prominent ethical standard as providing a strong suggestion analogous to the suggestions contained in our earlier examples. The mediator in problem 7 is a close analogy. Finally, even if it is truly the force of moral responsibility or sensitivity to public opinion that constrains the participants, and not the "signal" they get, we must still look to the source of the public's own opinion; and there, the writer suggests, the need for a simple, qualitative rationale often reflects the mechanism discussed in this paper.

But, if this general line of reasoning is valid, any analysis of explicit bargaining must pay attention to what we might call the "communication" that is inherent in the bargaining situations, the signals that the participants read in the inanimate details of the case. And it means that tacit and explicit bargaining are not thoroughly separate concepts but that the various gradations from tacit bargaining up through types of incompleteness or faulty or limited communication to full communication all show some dependence on the need to coordinate expectations. Hence all show some degree of dependence of the participants themselves on their common inability to keep their eyes off certain outcomes.

This is not necessarily an argument for expecting explicit outcomes as a rule to lean toward exactly those that would have emerged if communication had been impossible; the focal points may certainly be different when speech is allowed, except in some of the artificial cases we have used in our illustrations. But

74 ELEMENTS OF A THEORY OF STRATEGY

what may be the *main* principle in tacit bargaining apparently may be at least *one* of the important principles in the analysis of explicit bargaining. And, since even much so-called "explicit" bargaining includes maneuver, indirect communication, jockeying for position, or speaking to be overheard, or is confused by a multitude of participants and divergent interests, the need for convergent expectations and the role of signals that have the power to coordinate expectations may be powerful.

Perhaps many kinds of social stability and the formation of interest groups reflect the same dependence on such coordinators as the terrain and the circumstances can provide: the band wagon at political conventions that often converts the slightest sign of plurality into an overwhelming majority; the power of constitutional legitimacy to command popular support in times of anarchy or political vacuum; the legendary power of an old gang leader to bring order into the underworld, simply because obedience depends on the expectation that others will be obedient in punishing disobedience. The often expressed idea of a "rallying point" in social action seems to reflect the same concept. In economics the phenomena of price leadership, various kinds of nonprice competition, and perhaps even price stability itself appear amenable to an analysis that stresses the importance of tacit communication and its dependence on qualitatively identifiable and fairly unambiguous signals that can be read in the situation itself. "Spontaneous" revolt may reflect similar principles: when leaders can easily be destroyed, people require some signal for their coordination, a signal so unmistakably comprehensible and so potent in its suggestion for action that everyone can be sure that everyone else reads the same signal with enough confidence to act on it, thus providing one another with the immunity that goes with action in large numbers. (There is even the possibility that such a signal might be provided from outside, even by an agent whose only claim to leadership was its capacity to signal the instructions required for concerted action.)

TACIT NEGOTIATION AND LIMITED WAR

What useful insight does this line of analysis provide into the practical problems of tacit bargaining that usually confront us,

particularly the problems of strategic maneuver and limited war? It certainly suggests that it is *possible* to find limits to war—real war, jurisdictional war, or whatever—with overt negotiation. But it gives us no new strong sense of *probability*. War was limited in Korea, and gas was not used in World War II; on the possibility of limited war these two facts are more persuasive than all the suggestions contained in the foregoing discussion. If the analysis provides anything, then, it is not a judgment of the probability of successfully reaching tacit agreement but a better understanding of where to look for the terms of agreement.

If there are important conclusions to be drawn, they are probably these: (1) tacit agreements or agreements arrived at through partial or haphazard negotiation require terms that are qualitatively distinguishable from the alternatives and cannot simply be a matter of degree; (2) when agreement must be reached with incomplete communication, the participants must be ready to allow the situation itself to exercise substantial constraint over the outcome; specifically, a solution that discriminates against one party or the other or even involves “unnecessary” nuisance to both of them may be the only one on which their expectations can be coordinated.

Gas was not used in World War II. The agreement, though not without antecedents, was largely a tacit one. It is interesting to speculate on whether any alternative agreement concerning poison gas could have been arrived at without formal communication (or even, for that matter, with communication). “Some gas” raises complicated questions of how much, where, under what circumstances: “no gas” is simple and unambiguous. Gas only on military personnel; gas used only by defending forces; gas only when carried by vehicle or projectile; no gas without warning—a variety of limits is conceivable; some may make sense, and many might have been more impartial to the outcome of the war. But there is a simplicity to “no gas” that makes it almost uniquely a focus for agreement when each side can only conjecture at what rules the other side would propose and when failure at coordination on the first try may spoil the chances for acquiescence in any limits at all.

The physical configuration of Korea must have helped in de-

fining the limits to war and in making geographical limits possible. The area was surrounded by water, and the principal northern political boundary was marked dramatically and unmistakably by a river. The thirty-eighth parallel seems to have been a powerful focus for a stalemate; and the main alternative, the "waist," was a strong candidate not just because it provided a shorter defense line but because it would have been clear to both sides that an advance to the waist did not necessarily signal a determination to advance farther and that a retreat to the waist did not telegraph any intention to retreat farther.

The Formosan Straits made it possible to stabilize a line between the Communist and National government forces of China, not solely because water favored the defender and inhibited attack, but because an island is an integral unit and water is a conspicuous boundary. The sacrifice of any part of the island would have made the resulting line unstable; the retention of any part of the mainland would have been similarly unstable. Except at the water's edge, all movement is a matter of degree; an attack across water is a declaration that the "agreement" has been terminated.

In Korea, weapons were limited by the qualitative distinction between atomic and all other; it would surely have been much more difficult to stabilize a tacit acceptance of any limit on size of atomic weapons or selection of targets.⁹ No definition of size or target is so obvious and natural that it goes without saying, except for "no size, on any target." American assistance to the French forces in Indochina was persuasively limited to material, not people; and it was appreciated that an enlargement to include, say, air participation could be recognized as limited to air, while it would not be possible to establish a limited *amount* of air or ground participation. One's intentions to abstain from ground intervention can be conveyed by the complete withholding of ground forces; one cannot nearly so easily commit *some* forces and communicate a persuasive limit to the *amount* that one intends to commit.

The strategy of retaliation is affected by the need to communicate or coordinate on limits. Local aggression defines a place;

* This point is developed at length in Appendix A.

with luck and natural boundaries, there may be tacit acceptance of geographical limits or limits on types of targets. One side or both may be willing to accept limited defeat rather than take the initiative in breaching the rules, and to act in a manner that reassures the other of such willingness. The "rules" may be respected because, if they are once broken, there is no assurance that any new ones can be found and jointly recognized in time to check the widening of the conflict. But if retaliation is left to the method and place of the retaliator's own choosing, it may be much more difficult to convey to the victim what the proposed limits are, so that he has a chance to accept them in his counter-retaliation. In fact, the initial departure of retaliation from the locality that provokes it may be a kind of declaration of independence that is not conducive to the creation of stable mutual expectations. Thus the problem of finding mutually recognized limits on war is doubly difficult if the definition implicit in the aggressor's own act is not tolerable.

In sum, the problem of limiting warfare involves not a continuous range of possibilities from most favorable to least favorable for either side; it is a lumpy, discrete world that is better able to recognize qualitative than quantitative differences, that is embarrassed by the multiplicity of choices, and that forces both sides to accept some dictation from the elements themselves. The writer suggests that the same is true of restrained competition in every field in which it occurs.

PRIOR ARRANGEMENTS

While the main burden of this paper has been that tacit bargaining is possible and is susceptible of systematic analysis, there is no assurance that it will succeed in any particular case or that, when it succeeds, it will yield to either party a particularly favorable outcome compared with alternatives that might have been available if full communication had been allowed. There is no assurance that the next war, if it comes, will find mutually observed limits in time and of a sort to afford protection, unless explicit negotiation can take place. There is reason, therefore, to consider what steps can be taken before the time for tacit

bargaining occurs, to enhance the likelihood of a successful outcome.

Keeping communication channels open seems to be one obvious point. (At a minimum, this might mean assuring that a surrender offer could be heard and responded to by either side.) The technical side of this principle would be identification of who would send and receive messages, upon what authority, over what facilities, using what intermediaries if intermediaries were used, and who stood in line to do the job in what fashion if the indicated parties and facilities were destroyed. In the event of an effort to fight a restrained nuclear war, there may be only a brief and busy instant in which each side must decide whether limited war is in full swing or full war has just begun; and twelve hours' confusion over how to make contact might spoil some of the chances for stabilizing the action within limits.

Thought should be given to the possible usefulness of mediators or referees. To settle on influential mediators usually requires some prior understanding, or at least a precedent or a tradition or a sign of welcome. Even if we rule out overt arrangements for the contingency, evidences by each side of an appreciation of the role of referees and mediators, even a little practice in their use, might help to prepare an instrument of the most extreme value in an awful contingency.

But all such efforts may suffer from the unwillingness of an adversary to engage in any preparatory steps. Not only may an adversary balk at giving signs of eagerness to come to agreement; it is even possible that one side in a potential war may have a tactical interest in keeping that war unrestrained and aggravating the likelihood of mutual destruction in case it comes. Why? Because of the strategy of threats, bluffs, and deterrents. The willingness to start a war or take steps that may lead to war, whether aggression or retaliation to aggression, may depend on the confidence with which a nation's leaders think a war could be kept within limits. To be specific, the willingness of America to retaliate against local aggression with atomic attack depends — and the Russians know that it depends — on how likely we consider it that such retaliation could itself remain limited. That is, it

depends on how likely it is in our judgment that we and the Russians, when we both desperately need to recognize limits within which either of us is willing to lose the war without enlarging those limits, will find such limits and come to mutually recognized acquiescence in them. If, then, Russian refusal to engage in any activity that might lead to the possibility of limited war deters our own resolution to act, they might risk forgoing such limits for the sake of reducing the threat of American action. One parachutist in our example may know that the other will be careless with the plane if he is sure they can meet and save themselves; so if the first abstains from discussing the contingency, the other will have to ride quietly for fear of precipitating a fatal separation in the terrain below.

Whether this consideration or just the usual inhibitions on serious negotiation make prior discussion impossible, there is still a useful idea that emerges from one of our earlier games. It is that negotiation or communication for the purpose of coordinating expectations need not be reciprocal: unilateral negotiation may provide the coordination that will save both parties. Furthermore, even an unwilling member cannot necessarily make himself unavailable for the receipt of messages. Recall the man who proposed the letter R in one of the bargaining games: as long as the partner heard — and it is obvious that he heard — the letter R is the only extant proposal, and, being unchallenged, it may coordinate in default of any counterproposal nearly as well as if it had been explicitly accepted. (Even *denial* of it by the other party might not manage to dislodge its claim to prominence but rather simply prove his awareness of it, as long as no rival claim was made that created ambiguousness.) If one of our parachutists, just before the plane failed and while neither of them dreamed of having to jump, idly said, "If I ever had to meet somebody down there, I'd just head for the highest hill in sight," the other would probably recall and know that the first would be sure he recalled and would go there, even though it had been on the tip of his tongue to say, "How stupid," or "Not me, climbing hurts my legs," when the plane failed. When some signal is desperately needed by *both* parties and both parties

80 ELEMENTS OF A THEORY OF STRATEGY

know it, even a poor signal and a discriminatory one may command recognition, in default of any other. Once the contingency is upon them, their interests, which originally diverged in the play of threats and deterrents, substantially coincide in the desperate need for a focus of agreement.

PART II

A REORIENTATION OF

GAME THEORY

TOWARD A THEORY OF INTERDEPENDENT DECISION

On the strategy of pure conflict — the zero-sum games — *game theory* has yielded important insight and advice. But on the strategy of action where conflict is mixed with mutual dependence — the nonzero-sum games involved in wars and threats of war, strikes, negotiations, criminal deterrence, class war, race war, price war, and blackmail; maneuvering in a bureaucracy or in a traffic jam; and the coercion of one's own children — traditional game theory has not yielded comparable insight or advice. These are the "games" in which, though the element of conflict provides the dramatic interest, mutual dependence is part of the logical structure and demands some kind of collaboration or mutual accommodation — tacit, if not explicit — even if only in the avoidance of mutual disaster. These are also games in which, though secrecy may play a strategic role, there is some essential need for the signaling of intentions and the meeting of minds. Finally, they are games in which what one player *can* do to avert mutual damage affects what another player *will* do to avert it, so that it is not always an advantage to possess initiative, knowledge, or freedom of choice.

Traditional game theory has, for the most part, applied to these mutual-dependence games (nonzero-sum games) the methods and concepts that proved successful in studying the strategy of pure conflict. The present chapter and the one to follow attempt to enlarge the scope of game theory, taking the zero-sum game to be a limiting case rather than a point of departure. The proposed extension of the theory will be mainly along two lines. One is to identify the perceptual and suggestive element in the

formation of mutually consistent expectations. The other (in the following chapter) is to identify some of the basic "moves" that may occur in actual games of strategy, and the structural elements that the moves depend on; it involves such concepts as "threat," "enforcement," and the capacity to communicate or to destroy communication.

That game theory is underdeveloped along these two lines may reflect its preoccupation with the zero-sum game. Suggestions and inferences, threats and promises, are of no consequence in the accepted theory of zero-sum games. They are of no consequence because they imply a relation between the two players that, unless perfectly innocuous, must be to the disadvantage of one player; and he can destroy it by adopting a minimax strategy, based, if necessary, on a randomizing mechanism. So the "rational strategies" pursued by two players in a situation of pure conflict — as typified by pursuit and evasion — should not be expected to reveal what kind of behavior is conducive to mutual accommodation, or how mutual dependence can be exploited for unilateral gain.

If the zero-sum game is the limiting case of pure conflict, what is the other extreme? It must be the "pure-collaboration" game in which the players win or lose together, having identical preferences regarding the outcome. Whether they win fixed shares of the total or shares that vary with the joint total, they must rank all possible outcomes identically, in their separate preference scales. (And, to avoid any initial conflict, it has to be evident to the players that the preferences are identical, so that there is no conflict of interest in the information or misinformation that they try to convey to each other.)

What is there about pure collaboration that relates it to game theory or to bargaining? A partial answer, just to establish that this game is not trivial, is that it may contain problems of perception and communication of a kind that quite generally occur in nonzero-sum games. Whenever the communication structure does not permit players to divide the task ahead of time according to an explicit plan, it may not be easy to coordinate behavior in the course of the game. Players have to understand each other, to dis-

cover patterns of individual behavior that make each player's actions predictable to the other; they have to test each other for a shared sense of pattern or regularity and to exploit clichés, conventions, and impromptu codes for signaling their intentions and responding to each other's signals. They must communicate by hint and by suggestive behavior. Two vehicles trying to avoid collision, two people dancing together to unfamiliar music, or members of a guerrilla force that become separated in combat have to concert their intentions in this fashion, as do the applauding members of a concert audience, who must at some point "agree" on whether to press for an encore or taper off together.

If *chess* is the standard example of a zero-sum game, *charades* may typify the game of pure coordination; if *pursuit* epitomizes the zero-sum game, *rendezvous* may do the same for the coordination game.

An experiment of O. K. Moore and M. I. Berkowitz provides a nice mixture in which the two limiting cases are both visible.¹ It involves a zero-sum game between two teams, each team consisting of three people. The three members of the team have identical interests but, because of a special feature of the game, cannot behave as a single entity. The special feature is that the three members of each team are separated and can communicate only by telephone and that all six telephones are connected on the same line so that everyone can hear both the other team and his own teammates. No rearrangement of codes is permitted. Between teams we have here a pure-conflict game; among the members of the team we have a pure-coordination game.

If in this game we suppress the "other team" and if the three players simply try to coordinate a winning strategy in a game of skill or chance in the face of communication difficulty, we have a three-person pure-coordination game. Several "games" of this sort have been studied, both experimentally and formally; in fact, there is substantial overlap at this point between the nonzero-sum game and organization or communication theory.²

¹O. K. Moore and M. I. Berkowitz, *Game Theory and Social Interaction*, Office of Naval Research, Technical Report, Contract No. SAR/NONR-609 (16) (New Haven, November, 1956).

²An extensive formal analysis of the coordination problem is developed by Jacob Marschak, "Elements for a Theory of Teams," and, "Toward an Eco-

The experiments reported in Chapter 3 showed that coordinated choice is possible even in the complete absence of communication. Further, they showed that there are tacit bargaining situations in which the *conflict* of interest in the choice of action may be overwhelmed by the sheer need for concerting on *some* action; in those situations, the limiting case of pure coordination isolates the essential feature of the corresponding nonzero-sum game.

So we do have, in this *coordinated problem-solving*, with its dependence on the conveyance and perception of intentions or plans, a phenomenon that brings out an essential aspect of the nonzero-sum game; and it stands in much the same relation to it as the zero-sum game, namely, that of "limiting case." One is the mixed conflict-cooperation game with all scope for cooperation eliminated; the other is the mixed conflict-cooperation game with the conflict eliminated. In one the premium is on secrecy, in the other on revelation.

It is to be stressed that the pure-coordination game is a *game of strategy* in the strict technical sense. It is a behavior situation in which each player's best choice of action depends on the action he expects the other to take, which he knows depends, in turn, on the other's expectations of his own. This interdependence of expectations is precisely what distinguishes a game of strategy from a game of chance or a game of skill. In the pure-coordination game the interests are convergent; in the pure-conflict game the interests are divergent; but in neither case can a choice of action be made wisely without regard to the dependence of the outcome on the mutual expectations of the players.³

nomic Theory of Organization and Information," *Cowles Foundation Discussion Papers*, Nos. 94 and 95 (New Series), and, with Roy Radner, "Structural and Operational Communication Problems in Teams," *Cowles Foundation Discussion Papers, Economics*, No. 2076. Examples of relevant empirical work can be found in Alex Bavelas, "Communication Patterns in Task-oriented Groups," in D. Cartwright and A. F. Zander, *Group Dynamics* (Evanston, 1953), G. A. Heise and G. A. Miller, "Problem Solving by Small Groups Using Various Communication Nets," in P. A. Hare, E. F. Borgatta, and R. F. Bales, *Small Groups* (New York, 1955), H. J. Leavitt and R. A. H. Mueller, "Some Effects of Feedback on Communication," in *Small Groups*, and L. Carmichael, H. P. Hogan, and A. A. Walter, "An Experimental Study of the Effects of Language on the Reproduction of Visually Perceived Form," *Journal of Experimental Psychology*, 15, 73-86 (February, 1932).

³Concerning this point, Carl Kaysen in his review of Von Neumann and

Recall the famous case of Holmes and Moriarty on separate trains, neither directly in touch with the other, each having to choose whether to get off at the next station. We can consider three kinds of payoff. In one, Holmes wins a prize if they get off at different stations, Moriarty wins it if they get off at the same station; this is the zero-sum game, in which the preferences of the two players are perfectly correlated inversely. In the second case, Holmes and Moriarty will both be rewarded if they succeed in getting off at the same station, whatever station that may be; this is the pure-coordination game, in which the preferences of the players are perfectly correlated positively. The third payoff would show Holmes and Moriarty both being rewarded if they succeed in getting off at the same station, but Holmes gaining more if both he and Moriarty get off at one particular station, Moriarty gaining more if both get off at some other particular station, both losing unless they get off at the same station. This is the usual nonzero-sum game, or "imperfect-correlation-of-preferences" game. This is the mixture of conflict and mutual dependence that epitomizes bargaining situations. By specifying particular communication and intelligence systems for the players, we can enrich the game or make it trivial or provide an advantage to one of the two players in the first and third variants.

The essential game-of-strategy element is present in all three cases: the best choice for either depends on what he expects the other to do, knowing that the other is similarly guided, so that each is aware that each must try to guess what the second guesses the first will guess the second to guess and so on, in the familiar spiral of reciprocal expectations.

Morgenstern's *Theory of Games and Economic Behavior* says: "The theory of such games of strategy deals precisely with the actions of several agents, in a situation in which all actions are interdependent, and where, in general, there is no possibility of what we called parametrization that would enable each agent (player) to behave as if the actions of the others were given. In fact, it is this very lack of parametrization which is the essence of a game." Similar language is used by R. Duncan Luce and Howard Raiffa in *Games and Decisions* (New York, 1957): "Intuitively, the problem of conflict of interest is, for each participant, a problem of individual decision making under a mixture of risk and uncertainty, the uncertainty arising from his ignorance as to what the others will do" (p. 14). Their preoccupation is with the conflict, however; the case of coincident preferences they dispose of as trivial (pp. 59, 88), and they deal with such players as a single individual (p. 13).

A RECLASSIFICATION OF GAMES

Before going further, we can usefully reclassify game situations. The twofold division into zero-sum and nonzero-sum lacks the symmetry that we need and fails to identify the limiting case that stands opposite to the zero-sum game. The essentials of a classification scheme for a two-person game could be represented on a two-dimensional diagram. The values of any particular outcome of the game, for the two players, would be represented by the two coordinates of a point. All possible outcomes of a pure-conflict game would be represented by some or all of the points on a negatively inclined line, those of a pure common-interest game by some or all of the points on a positively inclined line. In the mixed game, or bargaining situation, at least one pair of points would denote a negative slope and at least one pair a positive slope.⁴

*"If the nature of the game makes it desirable for a player to use a random device in the choice of his strategy, or feasible for the players to negotiate an enforceable agreement that, like a drawing of lots, depends on a chance mechanism, there may be room for cooperation in the choice of *strategies* even when there is perfect disagreement over the ranking of *outcomes*. In that case the points representing the pure-conflict game must meet the tighter restriction of lying on a straight line, with the two axes measuring the players' "utilities" in the sense now familiar in game theory. This restriction also applies to the pure common-interest game, since players who agree perfectly on the ranking of *outcomes* may not agree on the desirability of, say, one particular point over a fifty-fifty chance between the two points immediately above and below it. Thus "strictly pure" conflict and common-interest games, providing no scope for collaboration in the one case and no scope for disagreement in the other, would have to show the *expected values* of all pertinent mixed (random) strategies lying along the downward-sloping and upward-sloping lines, respectively, with axes measured in "utility units" of the kind mentioned; this in turn means that the points denoting *outcomes* must lie on a straight line.*

Also, the pure games cannot admit "side payments." If one of the partners in a pure common-interest game threatens to sabotage the effect unless he is paid — assuming that the communication and enforcement structure of the game makes this possible — a conflict of interest is introduced; in effect, the point denoting the payment of a bribe would appear to the upper left or lower right of another point or points on the upward-sloping line, producing the configuration of a mixed game. And if one of the players in a pure-conflict game can threaten damage or offer compensation to induce his opponent to yield in this game, there is scope for bargaining; there is no longer a relation of pure conflict, and the points denoting the threatened damage or promised compensation would lie off the downward-sloping line. In other words, *all* pertinent potential outcomes must be allowed for. (Two simultaneous pure-conflict games, even if

We could stay close to traditional terminology, with respect to the strictly pure games, by calling them *fixed-sum* and *fixed-proportions* games, getting the unwieldy *variable-sum-variable-proportions* as the name for all games except the limiting cases. We could also call them perfect-negative-correlation games and perfect-positive-correlation games, referring to the correlation of their preferences with respect to outcomes, leaving for the richer mixed game the rather dull title of "imperfect-correlation game."

The difficulty is in finding a sufficiently rich name for the mixed game in which there is both conflict and mutual dependence. It is interesting that we have no very good word for the *relation* between the players: in the common-interest game we can refer to them as "partners" and in the pure-conflict game as "opponents" or "adversaries"; but the mixed relation that is involved in wars, strikes, negotiations, and so forth, requires a more ambivalent term.⁵ In the rest of this book I shall refer to the mixed game as a *bargaining game* or *mixed-motive game*, since these terms seem to catch the spirit. "Mixed-motive" refers not, of course, to an individual's lack of clarity about his own preferences but rather to the ambivalence of his relation to the other player — the mixture of mutual dependence and conflict, of partnership and competition. "Nonzero-sum" refers to the mixed game together with the pure common-interest game. And, because it characterizes the problem and the activity involved, *coordination game* seems a good name for the perfect sharing of interests.

GAMES OF COORDINATION

While most of this book will be about the mixed game, a brief discussion of the pure coordination game, beyond that of Chapter

they meet the restriction of straight lines, provide room for negotiation unless the slopes of the two lines happen to be identical.)

⁵ It deserves to be emphasized that nonzero-sum games can as properly be classed under theory of partnership as under theory of conflict; and for providing insight into problems like that of limiting war, there is merit in using words that bring out the common interest of the adversaries and the "bargaining process" involved in the military maneuvers themselves. As will be seen in Chapter 9, even the problem of surprise attack is logically equivalent to a problem in partnership discipline. If *theory of games* has become endowed with a too conflict-oriented connotation, perhaps something like *theory of interdependent decision* would be a neutral term that equally covers the two limiting cases as well as the mixed case.

3, will help to show that this is an important game in its own right and will identify certain qualities of the mixed game that appear most clearly in the limiting case of pure coordination.

Recall the various pure coordination problems of Chapter 3. Each of them evidently provided some focal point for a concerted choice, some clue to coordination, some rationale for the convergence of the participants' mutual expectations. It was argued there that the same kind of coordinating clue might be a potent force not only in pure coordination but in the mixed situation that includes conflict; and, in fact, the experiments demonstrated that, in the complete absence of communication, this is certainly true. But there are a number of instances in which pure coordination itself — the *tacit* procedure of identifying partners andconcerting plans with them — is a significant phenomenon. A good example is the formation of riotous mobs.

It is usually the essence of mob formation that the potential members have to know not only where and when to meet but just when to act so that they act in concert. Overt leadership solves the problem; but leadership can often be identified and eliminated by the authority trying to prevent mob action. In this case the mob's problem is to act in unison without overt leadership, to find some common signal that makes everyone confident that, if he acts on it, he will not be acting alone. The role of "incidents" can thus be seen as a coordinating role; it is a substitute for overt leadership and communication. Without something like an incident, it may be difficult to get action at all, since immunity requires that all know when to act together. Similarly, the city that provides no "obvious" central point or dramatic site may be one in which mobs find it difficult to congregate spontaneously; there is no place so "obvious" that it is evident to everyone that it is obvious to everyone else. Bandwagon behavior, in the selection of leadership or in voting behavior, may also depend on "mutually perceived" signals, when a part of each person's preference is a desire to be in a majority or, at least, to see some majority coalesce.⁸

Excessively polarized behavior may be the unhappy result of

* A closely related phenomenon is appreciated by the person who tries to blend into the crowd to avoid being called on to recite, picked on by a bully, or singled out for "election" to some post that everybody wants to escape.

dependence on tacit coordination and maneuver. When whites and Negroes see that an area will "inevitably" become occupied exclusively by Negroes, the "inevitability" is a feature of convergent expectation.⁷ What is most directly perceived as inevitable is not the final result but the *expectation* of it, which, in turn, makes the result inevitable. Everyone expects everyone else to expect everyone else to expect the result; and everyone is powerless to deny it. There is no stable focal point except at the extremes. Nobody can expect the tacit process to stop at 10, 30, or 60 per cent; no *particular* percentage commands agreement or provides a rallying point. If tradition suggests 100 per cent, tradition could be contradicted only by explicit agreement; if coordination has to be tacit, compromise may be impossible. People are at the mercy of a faulty communication system that makes it easy to "agree" (tacitly) to move but impossible to agree to stay. Quota systems in housing developments, schools, and so forth, can be viewed as efforts to substitute an explicit game with communication and enforcement for a tacit game that has an undesirably extreme "solution."

The coordination game probably lies behind the stability of institutions and traditions and perhaps the phenomenon of leadership itself. Among the possible sets of rules that might govern a conflict, tradition points to the particular set that everyone can expect everyone else to be conscious of as a conspicuous candidate for adoption; it wins by default over those that cannot readily be identified by tacit consent. The force of many rules of etiquette and social restraint, including some (like the rule against ending a sentence with a preposition) that have been divested of their relevance or authority, seems to depend on their having become "solutions" to a coordination game: everyone expects everyone to expect everyone to expect observance, so that non-observance carries the pain of conspicuousness. Clothing styles and motorcar fads may also reflect a game in which people do not wish to be left out of any majority that forms and are not

⁷The phenomenon, called "tipping," is analyzed by M. Grodzins, "Metropolitan Segregation," *Scientific American*, 197:33-41 (October, 1957). A more innocuous example of explosively convergent expectations, based on tacit communication that has an almost electric quality, is the snicker that ignites an outburst of uncontrollable laughter in a nervous crowd. An important example was the collapse of the Batista regime, or of the Fourth Republic.

organized to keep majorities from forming. The concept of *role* in sociology, which explicitly involves the expectations that others have about one's behavior, as well as one's expectations about how others will behave toward him, can in part be interpreted in terms of the stability of "convergent expectations," of the same type that are involved in the coordination game. One is trapped in a particular role, or by another's role, because it is the only role that in the circumstances can be identified by a process of tacit consent.

A good example might be the *esprit de corps* (or lack of it) of an army unit or naval vessel or the value system of a particular college or fraternity. These are social organisms that are subject to a substantial rate of replacement but that maintain their own peculiar identities to an extent that does not seem to be accounted for by selective or biased recruitment. The individual character of one of these units seems to be largely a matter of convergent expectations — everyone's expectation of what everyone expects of everyone — with the new arrivals' expectations being molded in time to help mold the expectations of subsequent arrivals. There is a sense of "social contract," the particular terms of which are sensed and accepted by each incoming generation. I am told that this persistence of a tradition in a social entity is one of the reasons why the legal identity of an army division or regiment — its name and number and history — is often deliberately preserved when its strength has fallen to where abolition might seem indicated: the tradition that goes with the legal identity of the group is an asset worth preserving for a future buildup. It may be the same phenomenon that makes it possible to collect income tax in some countries and not in others: if appropriate mutual expectations exist, people will expect evasion to be on a scale small enough not to overwhelm the authorities and may consequently pay up either out of a sense of reciprocated honesty or out of fear of apprehension, thus together justifying their own expectations.

Nature of the intellectual process in coordination. It should be emphasized that coordination is not a matter of guessing what the "average man" will do. One is not, in tacit coordination, trying to

guess what another will do in an objective situation; one is trying to guess what the other will guess one's self to guess the other to guess, and so on ad infinitum. ("Meeting" someone in the personal column of a newspaper is a good example.⁸) The reasoning becomes disconnected from the objective situation, except insofar as the objective situation may provide some clue for a concerted choice. The analogy is not just trying to vote with the majority but trying to vote with a majority when everyone wants to be in a

*So is meeting on the same radio frequency with whoever may be signaling to us from outer space. "At what frequency shall we look? A long spectrum search for a weak signal of unknown frequency is difficult. But, just in the most favored radio region there lies a unique, objective standard of frequency, which must be known to every observer in the universe: the outstanding radio emission line at 1420 megacycles of neutral hydrogen" (Giuseppe Cocconi and Philip Morrison, *Nature*, Sept. 19, 1959, pp. 844-846). The reasoning is amplified by John Lear: "Any astronomer on earth would say, 'Why, 1420 megacycles of course! That's the characteristic radio emission line of neutral hydrogen. Hydrogen being the most plentiful element beyond the earth, our neighbors would expect it to be looked for even by tyros in astronomy'" ("The Search for Intelligent Life on Other Planets," *Saturday Review*, Jan. 2, 1960, pp. 39-43). What signal to look for? Cocconi and Morrison suggest a sequence of small prime numbers of pulses, or simple arithmetic sums.

And this suggests an alternative orientation of those experiments in which subjects are instructed to make guesses, throughout a long random sequence of red or green lights, whether red or green will come up next. Subjects apparently persist in guessing on the basis of some pattern they think they perceive, an "irrational" mode of behavior given their knowledge that the sequence is generated by a random device. But, as Herbert Simon points out, "Man is not only a learning animal; he is a pattern-finding and concept-forming animal" ("Theories of Decision-Making in Economics and Behavioral Science," *American Economic Review*, 44:272). Why not, then, add to the experiment a cooperating pattern-maker, who generates the signals subject to various constraints and random interferences, and let the persistent pattern-seeking subject use his skill in finding the pattern planted by a cooperative partner rather than spend it futilely on random series? If, to make it tax the communicators' ingenuity, we add a third party whose reward is inversely related to that of the cooperating partners, who is allowed to intercept the message and within limits to alter it, we have something akin to the game of Moore and Berkowitz described earlier. Enriching the materials available beyond the binary choice of red and green might provide scope for genuinely creative pattern forming, of the kind that is interesting for Gestalt psychology, esthetics, and even higher-order problem solving. Simon notes in the same article (p. 426) that even a computer can be programmed "to use something akin to imagery or metaphor in planning its proofs" of geometrical theorems. This is pattern seeking of real interest. (It reminds us that the assumption of "malevolent nature" by the zero-sum game theorist is not applicable to, say, mathematical invention. Nature gives hints; she presents her secrets in patterns that make them infinitely easier to guess than if an exhaustive scanning were required to find them.)

majority and everyone knows it — not to predict Miss Rheingold of 1960 but to buy the stock or real estate that everyone expects everyone to expect everyone to buy. Investment in diamonds may be a perfect example; the greatest of all may be the monetary role of gold, which can perhaps be explained only as the “solution” of a coordination game. (A common household version of the coordination game occurs when two people are cut off in a telephone conversation; if they both call back, they only get busy signals.)

Consider the game of “name a positive number.” Experiments like those of Chapter 3 demonstrate that most people, asked just to pick a number, will pick numbers like 3, 7, 13, 100, and 1. But when asked to pick the same number the others will pick when the others are equally interested in picking the same number, and everyone knows that everyone else is trying, the motivation is different. The preponderant choice is the number 1. And there seems to be good logic in this: there is no unique “favored number”; the variety of candidates like 3, 7, and so forth, is embarrassingly large, and there is no good way of picking the “most favorite” or most conspicuous. If one then asks what number, among all positive numbers, is most clearly unique, or *what rule of selection would lead to unambiguous results*, one may be struck with the fact that the universe of all positive numbers has a “first” or “smallest” number.⁹

⁹There is a widely quoted passage in Keynes (p. 156) that may be worth repeating in order to point out that, while it deals with exactly the problem dealt with here, its conception of the “solution” is *not* at all the same: “Professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one’s judgment, are really the prettiest, nor even those which average opinion genuinely thinks prettiest. We have reached the third degree where we devote our intelligence to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees” (J. M. Keynes, *The General Theory of Employment, Interest and Money* [New York, 1936], p. 156). This class of games demonstrates, incidentally, that the usual correlation between parametric behavior and large numbers does not hold for tacit

Game-theory formulation of the coordination problem. The payoff matrix for a pure coordination problem would look something like that in Fig. 8. One player chooses a row, the other a column;

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

FIG. 8

and they receive the rewards denoted by the numbers contained in the cell where their choices intersect. If to each choice of one player there corresponds a single choice for the other that "wins" for both of them, we can arrange columns so that all the winning cells lie along the diagonal. In those cells there are positive payoffs to both players, in the rest we can put zeros. (For our present purpose there is nothing lost by letting a single number stand in each cell for the payoff to both players.)

But we must rule out a possible axiom that might seem to be suggested by analogy with other game theories, namely, that (to use the term of Luce and Raiffa) the "labeling" of rows, columns, and players should make no difference to the outcome.¹⁰ It is pre-

play with multiple equilibria. To adapt "parametrically" to the behavior of others requires in this case that their behavior be observable, not conjectural; the nonparametric character of tacit coordination remains, no matter how large the number of players.

¹⁰ Labeling of the *players* is explicitly ruled out by Luce and Raiffa (pp. 123-127) in discussing cooperative games and in effect is ruled out by Nash in his symmetry assumption (J. F. Nash, "The Bargaining Problem," *Econometrica*, 18:155-162 [1950], and "Two Person Cooperative Games," *Econometrica*, 21:128-140 [1953]). Labeling of *strategies* for tacit or explicit nonzero-sum games is implicitly precluded by dealing only with games in normal form, that is, the abstract version of them as represented by a payoff matrix (which is itself an *analytical* device, not part of the game, and hence provides no left-right, upper-lower, or numerical ordering of the actual strategies). A good example in which the labeling of *players* is the controlling factor is the interrupted

cisely because strategies are "labeled" in some sense — that is, have symbolic or connotative characteristics that transcend the mathematical structure of the game — that players can rise above sheer chance and "win" these games; and it is for that same reason that these games are interesting and important.

Even the game portrayed in Fig. 8 which might seem to have a minimum of symbolic significance attached to rows and columns, is not a hard one to "win," that is, for players to do substantially better on than chance would suggest, if it is portrayed in a matrix as shown. (If we give that same game an infinite series of rows and columns, it seems to become easier rather than harder. In that case it is formally identical with the game mentioned earlier, "Pick a positive number," but, because the "labeling" is different, there is less tendency for minorities to congregate at 3, 7, 13, and so forth.) Just forming the matrix prejudices the choice, since it focuses attention on "first," "middle," "last," and so forth.¹¹ If strategies are not given sequential labels, that is, labels that can be ordered like numbers and alphabets, but are given individual names, and these are not presented in any particular order, it is the names that must coordinate choice.

And here it becomes emphatically clear that the intellectual processes of choosing a strategy in pure conflict and choosing a strategy of coordination are of wholly different sorts. At least this is so if one admits the "minimax" solution, randomized if necessary, in the zero-sum game. In the pure-coordination game, the player's objective is to make contact with the other player through some imaginative process of introspection, of searching for shared clues; in the minimax strategy of a zero-sum game — most strikingly so with randomized choice — one's whole objective is to avoid any meeting of minds, even an inadvertent one.¹²

telephone call mentioned earlier, with the problem of who should call back and who should wait for the call.

¹¹ This point is typical of a number of demonstrations in the author's experiments reported earlier, to the effect that the postulate regarding the "independence of irrelevant alternatives" cannot be credited in the tacit game and, for analogous reasons, should not be expected to hold in the explicit bargaining game. Potential outcomes can be relevant to the coordination of choice though not themselves near to being chosen. For a statement and discussion of this postulate see Luce and Raiffa, p. 127.

¹² Randomized strategies may nevertheless be useful to achieve a coordinated

To illustrate, suppose that I am to name one card in an ordinary deck of fifty-two and you are to guess which one I name. Traditional game theory gives guidance on how to make my choice on the assumption that I do not want you to outguess me; I can select at random and defy you to have a better than random chance of guessing what I name. But if the game is that I *do* want you to guess correctly and you know that I will try to pick one that facilitates your guess, the random device can only guarantee to make tacit cooperation impossible. Holmes can *destroy* the labeling of the stations by flipping a coin to decide where to get off the train; and Moriarty has only a fifty-fifty chance of guessing a coin. But in the common-interest version they must somehow *use* the labeling of the stations in order to do better than pure chance; and how to use it may depend more on imagination than on logic, more on poetry or humor than on mathematics. It is noteworthy that traditional game theory does not assign a "value" to this game: how well people can concert in this fashion is something that, though hopefully amenable to systematic analysis,

distribution of votes, say, among a panel of candidates. If a 55 per cent majority exists and knows that it does, among a hundred voters; if two out of six candidates are congenial to it; and if the three candidates polling the largest numbers of votes become the board of directors, there is danger that uncoordinated polling may concentrate too many votes on the first (or second) majority choice, leaving the minority two winning candidates with 22 votes apiece. But if each member of the majority flips a coin to cast his vote for one of his party's men, the likelihood of one's getting as few as 22 votes is only one chance in six. If the minority, too, lacks an overt means of collaborating and relies on a chance device, the majority's chances are excellent.

A partial randomized strategy may also be used to reduce an area of conflict. Suppose two people, seated at North and East sides of a card table, are to move to another card table adjacent that is identically oriented, must choose without communication what seats they will take at the other table, and will win prizes of \$1 apiece if they pick adjacent seats. This is an easy coordination problem; but let us subvert the incentives, by giving an additional \$2 premium to the player who is on the other's right in the event they succeed in sitting next to each other. This game has no equilibrium point; interests do not converge; there is no seating arrangement that would not give one an incentive to move. (Each may wish that he could promise to sit on the other's left, but cannot.) A random strategy yields each player a minimax value of \$1. But, if each decides where he would sit in the pure common-interest game, then flips a coin to see whether he does sit there or sits opposite, the players guarantee that they neither choose the same seat nor sit opposite each other and share equal chances of winning the premium. This is an equilibrium pair of (mixed) strategies, worth an expected value of \$2 apiece.

cannot be discovered by reasoning a priori. This corner of game theory is *inherently* dependent on empirical evidence.¹³

It should particularly be noted that to assert the influence of "labels" (that is, of the symbolic and connotative details of the game) and the dependence of the theory on empirical evidence does not involve the question of whether game theory is predictive or normative — concerned with generalizations about actual choice or the strategy of correct choice. The assertion here is *not* that people simply *are* affected by symbolic details but that they *should* be for the purpose of correct play. A normative theory must produce strategies that are at least as good as what people can do without them. More, it must not deny or expunge details of the game that can demonstrably benefit two or more players and that the players, consequently, should not expunge or ignore in their mutual interest. Two couples jockeying for space on a dance floor or two armies jockeying for a truce line may jointly suffer

¹³ In cases like this we need only to consider the question of what *price* players would pay for a bit of coordinating information, and what different information patterns yield what chances of coordinating, to find ourselves in the middle of Marschak's *theory of teams*.

There is, incidentally, a version of "prisoners' dilemma" for this game: two accomplices, apprehended before their alibi is prepared and interrogated separately, must concoct the alibis they invent or be revealed in their guilt. A tantalizing variant can be built by supposing that confession carries a lighter sentence than unconfessed guilt; each player has a "minimax" strategy of confession and must not only consider which particular alibi constitutes the *best* alibi strategy but *how good it is* (in terms of likely coincidence with his partner's) and whether they share the decision to try it. The matrix might be:

.5	.5	0	.5	0
0	.5	1	0	0
0	.5	0	1	0
0	.5	0	0	1

(Lower left entry in each cell is payoff to player choosing row, upper right to player choosing column.)

from decision processes that are limited to the abstract properties of the situation.

A particular implication of this general point is that the game in "normal" (mathematically abstract) form is not logically equivalent to the game in "extensive" (particular) form, once we admit the logic by which rational players concert their expectations of each other. As pointed out in Chapter 3, these same considerations seem to be powerfully present in explicit bargaining as well. A terminological implication of these considerations is that "noncooperative" is a poor name for the game of tacit coordination; it is desperately cooperative in its own peculiar way and is still so when we add conflict and form the tacit mixed-motive game (In Appendix C it is argued that certain solution concepts familiar in game theory can be given an interpretation in terms of the coordination concept.)

SUGGESTION AND MUTUAL PERCEPTION IN THE MIXED-MOTIVE GAME

Coordination-game theory, while interesting in its own right, is interesting mainly for the light that it sheds on the nature of the mixed-motive game. The coordination element shows up most strikingly in a purely tacit game, in which there is neither communication nor any sequence of moves by which the two players accommodate themselves to each other. An example, similar to problem 6 on page 62, would be the following.

One player is "located" in Cincinnati, the other in San Francisco; they have identical maps of the United States and are to divide the country between them. Each is to draw a line dividing the United States into two parts; the line may be straight or curved, related or unrelated to physical or political landmarks. If the two of them divide the map differently, neither gets anything; but if they draw identical division lines on their maps, they are both rewarded. The reward for each player depends on what is contained in his piece after the division, that is, the piece that contains the city in which he is located. Let us leave these rewards vague; they may depend partly on area, partly on population, partly on industrial wealth and agricultural resources,

and so forth, and may differ somewhat for the two players. In other words, while all terrain is valuable, not all parts of the country are equally valuable, and there is no clear specification of the valuation formula. (There is consequently no means of selecting a perfectly symmetrical division of values between the two players.)

In this game there is a compelling problem of coordination; each player can win only if he does exactly what the other expects him to, knowing that the other is similarly trying to do exactly what is expected of him. They must jointly find a line that in some fashion suggests itself to both of them or appeals to both of them. Neither can "outsmart" the other without outsmarting himself.

The experiments of Chapter 3 suggest that players are by no means helpless when faced with this kind of game. The game is nowhere near so "infinitely" difficult as the infinity of possible division lines might suggest; some variants of the game are not difficult at all. But a successful outcome does depend on the kinds of factors that are controlling in the pure-coordination game; in fact, some games of this sort are "won" by the two players' choosing exactly the same outcome as they would have chosen if the reward system gave them identical, instead of conflicting, interests. The problem is to find some signal or clue or rationalization that both can perceive as the "right" one, with each party prepared to be disciplined by that signal or clue in the event that it appears to discriminate against him. They must find their clues where they can. (If the map they are using happens, for example, to contain an embarrassing richness of clues, making it difficult to single out any particular one, a fairly arbitrary line drawn as a suggestion by the referee, identical on both maps, might have to be accepted as a "mediator," even if it is substantially biased toward one of the players.)

But this coordination element, especially in the case without conflict, appears to be essentially related to a *communication problem*. The pure-coordination game not only ceases to be interesting but virtually ceases to be a "game" if the players can concert with certainty, without difficulty, and without cost. The question arises, then, how important the coordination element can

be in mixed-motive games generally, since many of these take the form of overt bargaining with uninhibited speech.

The pervasiveness of the coordination principle arises from two separate considerations. One, which was discussed in Chapter 3 is that tacit bargaining provides an analytical model — perhaps only an analogy but perhaps an identification of the actual psychic and intellectual phenomenon — of the "rational" process of finding agreement in pure bargaining situations, those in which both parties recognize that there is a wide range of outcomes preferable to both of them over no agreement at all. The psychic phenomenon of "mutual perception" that can be verified as real and important in the tacit case has a role to play in the analysis of explicit bargaining. *Coordination of expectations* is the role.

Second, many of the bargaining processes or game situations that we want to analyze are at least partly tacit. In some cases, like maneuvering a car in a traffic jam, speech is physically precluded; in others, like developing a modus vivendi with a neighbor, speech is inhibited in the interests of privacy. Illicit bargaining, or diplomatic bargaining that would be embarrassing to both sides if overheard by other countries, may be less than fully articulate. If the number of players in a game is large, as it is in the bargaining process that determines the racial border lines between residential areas and professions, there may be no institutional provision for explicit negotiation. In these cases, while speech may be part of the bargaining process, actions are also part of it, and the game is one of "maneuver" rather than just talk.

Furthermore, if there are *moves* available to the players, so that it is an advantage to get on with the maneuver even while negotiating, and particularly if some maneuvers become visible to the other player only after a time lag, there is no reason to suppose that an instantaneous moratorium on maneuver will reign from the outset; in that case, the game progresses while the talk is going on. If the moves had only symbolic significance, we could include them in the communication process along with speech; but, typically, moves have a tactical significance, leaving the game irreversibly different from what it was before, and typically also their tactical significance raises them above the level of pure

speech even in their communication content. One may say and say that a gun is loaded without being able to prove it until he actually shoots; one may say and say that he considers an area strategically important and not be believed until he incurs expense or risk in its protection. Thus moves can reveal information about a player's value system or about the choices of action available to him; moves can commit him to certain actions when speech often cannot; and moves can often progress at a speed that is determined unilaterally, not dependent on formalities of agreement at a conference.

In other words, bargaining games quite typically involve a dynamic process of mutual accommodation rather than pure communication culminating in a crystallized agreement. The jockeying for limits in limited war is a perfect example, and we might illustrate it by modifying the parlor game described above.

An illustrative tacit game. Suppose our two players with their maps of the United States before them are each given 100 chips and told to play a game as follows.¹⁴ At each "move," each player will distribute five chips among states on his map. The moves are compared, and if the two players have put a chip apiece in the same state, those two chips are removed; if one player has put a chip and the other player three chips in the same state, a chip apiece is removed leaving only two chips representing the one player; and so forth. They do the same at the next move, again with five chips; this time they have the option of placing their chips on states that are yet uncovered or of placing them on states where there are already chips. If A puts two chips on a state in

¹⁴ Since it will be proposed in Chapter 6 that such games have, in fact, a research value, as well as an illustrative value, it should be observed at the outset that there is a special problem of motivating the players in an experimental nonzero-sum game. In a zero-sum game, winning is measured relative to one's immediate adversary, and the intellectual challenge and bilateral competition motivate the player toward the correct (and only) type of winning. But for a mixed-motive game, "winning" must be made to involve one's absolute score, not his score relative to that of the person he plays with; the incentives are distorted if the play is dominated by strictly bilateral competition. So, unless real rewards are given, the game has to be organized as a round robin or some such schedule that involves more than two players in a series of two-person plays, with the final outcome decided by the relative position of one's absolute score. (This is why there are no two-person nonzero-sum parlor games.)