Strategy and Politics

Emerson M.S. Niou
Duke University

Peter C. Ordeshook
California Institute of Technology

Chapter 1: Politics as a Game

1.1 Decision versus Game Theoretic Decision Making

Over twenty five hundred years ago the Chinese scholar Sun Tzu, in *The Art of War*, proposed a codification of the general strategic character of armed conflict and, in the process, offered practical advice for securing military victory. His advice is credited, for example, with having greatly influenced Mao Zedong's approach to conflict and the subtle tactics of revolution and the ways in which North Vietnam and the Viet Cong thwarted America's military advantages. The formulation of general strategic principles -- whether applied to war, parlor games such as Go, or politics -- has long fascinated scholars. And regardless of context, the study of strategic principles is of interest because it grapples with fundamental facts of human existence -- first, people's fates are interdependent; second, this interdependence is characterized generally by conflicting goals; and, finally as a consequence of the first two facts, conflicts such as war are not accidental but are the purposeful extension of a state's or an individual's motives and actions and must be studied in a rational way.

The Art of War is, insofar as we know, our first written record of the attempt to understand strategy and conflict in a coherent and general way. It is important, moreover, to recall that it was written at a time of prolonged conflict within an emerging China whereby the leaders of competing kingdoms possessed considerable experience not only in the explicit conduct of war, but also in diplomacy and strategic maneuver. As such, then, we should presume that it codifies the insights of an era skilled at strategy and tactics, including those of planning, deception and maneuver. This assumption, though, occasions a question: Although *The Art of War* was ostensibly written for the leader of a specific kingdom, what if all sides to a conflict have a copy of the book (or, equivalently, an advisor no less insightful than Sun Tzu)? How might our reading of Sun Tzu change if it is *common knowledge* that everyone studied *The Art of War* or its equivalent – where by 'common knowledge' we mean that everyone knows that everyone has a copy of the book, everyone knows that everyone knows that everyone ... and so on, ad infinitum. The assumption of

common knowledge presumes that not only is each decision maker aware of the situation, but each is aware that the other is aware, each knows that the other knows, and so on and so forth, and after being told by Sun Tzu himself that the great trap to be avoided is underestimating the capabilities of one's opponents, it seems imperative that the implementation of his advice proceed with the presumption that common knowledge applies.

In the case of *The Art of War*, taking account of the possibility that both sides of a conflict have a copy of the book differentiates the social from the natural sciences. In physics or chemistry, including their practical applications, one does not assume that the scientist or engineer confronts a benevolent or malevolent nature that acts strategically to deliberately assist or thwart one's research or the application of natural laws as we understand them. Things might not function as designed, but only because our understanding or application of nature's laws is imperfect. In the social sciences, on the other hand, especially in the domain of politics, it is typically the case that individuals must choose and act under the assumption that others are choosing and acting in reaction to one's decisions or in anticipation of them, where those reactions can be either benevolent or malevolent.

Despite this fact, it is our experience that most readers of *The Art of War* implicitly, or unconsciously at least initially, take the view that the reader is the sole beneficiary of Sun Tzu's advice … that one's opponent is much like nature, a 'fixed target'. This might have been a valid assumption in 225 BC China in the absence of the internet, printing presses and Xerox machines, but it is no longer valid given the worldwide distribution of the book, including having it as required reading in business schools and military war colleges. So a more sophisticated student of Sun Tzu's writings might suppose that one's opponents have read the book as well, and might then reasonably assume that their opponents' tactics and strategies conform to Sun Tzu's guidance. But suppose we take things a step further and try to interpret an opponent's actions not simply with the assumption that they've read the same books with which we are familiar, but that they know we are familiar with those books and that we are not only attempting to assess their tactics and strategies in light of the advice contained in *The Art of War*, but they also know we are attempting to take into

account the fact that they are attempting to take into account our familiarity with that advice.

If all of this sounds confusing, then referencing *The Art of War* as an introduction to this volume has served its purpose. Specifically, there are two general modes of decision making: *Decision Making Under Risk* and *Game Theoretic Decision Making.* In decision making under risk one assumes, in effect, that although there may be inherent uncertainties associated with the consequences of one's actions due to chance events and the actions of others, probabilities can be associated with those events and actions and the right choice is the one that yields the greatest expected return, where that return can be expressed in monetary terms, as psychological satisfaction or whatever. In this world one assumes that other decisions makers, including those who might be opposed to your goals, have, in effect, a limited view and do not respond to the assumption of common knowledge. In other words, just as the engineer or natural scientist does not assume that nature has the capacity for logical thought, the notion of common knowledge plays little to no role in decision making under risk since here one sees opponents as non-strategic 'fixed targets' whose likely actions can be guessed at on the basis, say, past patterns of behavior, bureaucratic rigidities or simple stupidity.

In game theoretic decision making, in contrast, one assumes that one's opponents and other decision makers, in pursuit of their goals, take into account their knowledge of you, including the fact that you know that they know, etc. Other decision makers are no longer fixed targets. Now you must be concerned that, since they know you are aware of their past behaviors, they might try to confound your calculations by defecting in some way from whatever patterns their earlier decisions exhibit. And there is, moreover, the additional complication. Since in a game-theoretic analysis we can also assume that they know you know their past history, they also know you know they might have an incentive to thwart your calculations by not changing past patterns of behavior at all. But, since you also know that they know that you know they might consider sticking to past patterns … and so on, ad infinitum once again.

Game theoretic decision making attempts to untangle such seemingly endless and convoluted thinking and in the process to define what it means to be rational in interactive decision contexts. This volume, then, attempts to lay out the rudiments of game theoretic analysis as it can be applied to situations we label 'political'. Our specific objective, however, is not to provide a text on the mathematics of game theory. There are any number of excellent books available for that purpose, and the subject itself can be as dense as any branch of mathematics. Rather, our goal is to show that a game theoretic approach to understanding individual action is an essential component not only of being skillful at war, but also to understanding the less violent aspect of politics. However, rather than try to argue this point here, let us consider a series of examples that perhaps more clearly illustrates the difference between decision theoretic and game theoretic reasoning.

> **The Atomic Bomb and Japan**: On the morning of August 6, 1945, a single plane (preceded by two weather reconnaissance aircraft), the Enola Gay, flew to and dropped its bomb on the city of Hiroshima. Ignoring the debate over whether this act was warranted or unwarranted with respect to the goal of ending a war, the question that concerns us here is: Why only one plane, which so easily could have been intercepted? The answer is that America's strategic planners assumed that if the Enola Gay had been part of a fleet of bombers, the Japanese would have attempted to intercept the raid with its ground based fighters. A single plane, on the other hand, would be far less threatening and draw far less attention. That calculation turned out to be correct – based on earlier bomb raids over its cities, strategic planners correctly assessed Japan's approach to air defense and when the two scout planes turned back to the Pacific, city sirens sounded the "all clear" on the ground. The logic behind sending a single plane, then, on its deadly mission seems straightforward. But then, three days later, another solitary plane, Bockscar, flew to Nagasaki and dropped America's second atomic bomb, and the question for us is: Did the same strategic calculation in choosing between a lone plane versus a plane imbedded in a fleet apply to Bockscar?

We do not know precisely what calculations were made in deciding to deliver the second bomb via another lone aircraft as opposed to a fleet. But certainly the calculation this time had to be different from the one that sent the Enola Gay on its way. In the case of the Enola Gay, America's strategic planners could reasonably assume that the Japanese had no idea as to the destructiveness of its cargo and, thus, no reason to fear it any more than any previous lone aircraft over Japan. The response of Japan's air defense command could be predicted with near certainty. But circumstances changed markedly once the Enola Gay delivered its payload. Now, presumably, there were those in Japan who knew the potential of a lone bomber, and the American decision to proceed as before had to be justified by a different calculation -- one that took into account what the Japanese might now assume about lone bombers and how they might weight that danger against the costs of scrambling interceptors against it. Might the Japanese assume that the United States wouldn't be bold enough to again send a single bomber to drop any additional atomic bombs and instead would now try to disguise any subsequent use of its atomic arsenal by imbedding the plane carrying it in a fleet of bombers? In other words, America's strategic planners now had to concern themselves with the possibility that Japan's approach to air defense had changed in a complex way dictated by its best guess as to America's guess about Japan's response to the first bomb.

The decision to send a single plane to Hiroshima, then, was decision-theoretic: Japan's likely response to one plane versus many could be determined by its previous actions. All a strategic planner needs to do is to calculate the probability that Japan would try to intercept a single plane versus the likelihood that, if imbedded in a fleet, it would intercept the fleet and successfully shoot down the specific plane carrying a bomb. The decision to send only one plane to Nagasaki, in contrast, required an assessment of what Japan might have learned about the potential lethality of a single plane, whether Japan might assume the Americans would employ the same tactic a second time, how that tentative assessment might impact America's tactical calculations, and how in turn Japan should respond to

what it thinks America's response would be to Japan's reassessment of things. The decision to use a single bomber versus a fleet over Nagasaki, then, was a game theoretic one.

The Boston Marathon 2013: For a more contemporary example of decision- versus game-theoretic choice, consider the FBI's move to release a department store security camera video of the two brothers who planted bombs at Boston's 2013 marathon. As portrayed by the media, that decision was intended to elicit the public's help in identifying the terrorists, and indeed the video was soon plastered across the internet's social media. From this perspective, then, the FBI's action appears to be a strictly decision theoretic move to increase the likelihood that their suspects would be recognized and identified. But suppose we give the FBI's personnel more credit in assessing motives. Suppose they anticipated the released video going viral on the internet and knew the suspects would soon realize that their identities could not be kept hidden. Thus, if they planned any additional terrorist acts, both men knew they had better act quickly with little opportunity to plan carefully. In other words, suppose the FBI intended to 'smoke out' their Russian suspects and induce them to be less careful than they might otherwise be if they assumed their identities could remain hidden for a time. It might have been the case, of course, that the two brothers knew the FBI was trying to smoke them out, but as committed jihadists, what choice did they have? Thus, by anticipating the terrorists' response, the FBI can be said to have acted with a game-theoretic understanding of things. And this is precisely what happened. A day or two after the bombing, at least one brother, seemingly oblivious to the fact that his identity would soon be known, was seen partying at the college he'd been attending. But following the video's release, the two brothers, with bomb parts still unassembled in their apartment, tipped their hand by hijacking a car that led to a shoot-out with the police wherein one brother was killed and the other injured and captured soon thereafter.

Voters and Interest Groups in Three Candidate Elections: It isn't always easy to decide how to vote in a three-candidate plurality rule (first-past-the-post) election. The problem here is the possibility of wasting one's vote by casting a ballot for a candidate, however strongly preferred, who stands no chance of winning. If the candidates' chances are unequal, it might be wise to vote for one's second preference. In making this decision, then, a voter might, after perhaps talking things over with family and friends, consult the polls to determine whether his or her preferred candidate is competitive. But now suppose our voter is not an ordinary citizen but heads some highly visible interest group … a labor union or citizen action committee … and that he or she must decide who that group should endorse. The endorsement decision is similar to that of an ordinary voter in that the relative competitiveness of the candidates should be taken into consideration; but it is different in that any decision should also take into account the likelihood that the endorsement will not only influence more than a mere handful of voters but also perhaps the actions of other interest groups. If their endorsement carries some weight and impacts the election's competitiveness, then presumably it will impact the calculations of others who might attempt to influence the election's outcome. Some of that influence might benefit the candidate in question if it leads other groups to endorse the same candidate. But it might also work against that group's interests if it results in any increase in the endorsements received by other candidates. Thus, your decision as a solitary voter made under the assumption of a "fixed electorate" as reflected in the polls is decision-theoretic since your decision hardly affects anyone. But leaders of influential interest groups must not only concern themselves with their immediate impact on the electorate, but also with the responses of other interest groups. The decisions of each such group, then, should be made on the basis of a game theoretic analysis that attempts to take into account the reactions and counter-reactions of other groups and, ultimately, of the electorate as a whole.

The Electoral College and Bloc Voting: The US Constitution leaves the door open to any number of schemes for translating individual votes into Electoral College votes and the ultimate determination of who wins a presidential election. Presently, nearly all Electoral College votes are determined by a winner-take-all system wherein whoever receives a plurality of popular votes in a state wins all of that state's electors. That, however, is not how it has always been. So suppose we step back in time to when individual states, as in the late 18th and early 19th centuries, employed a variety of schemes for allocating Electoral votes among competing candidates, including selecting them proportionally or electing them by pre-defined districts. Suppose further that you are an advisor to some state legislature and are trying to convince them that they ought to use today's winner-take-all scheme (perhaps the state in question is Virginia, perhaps it is the election of 1800 and perhaps you are Thomas Jefferson). If, for whatever reason, you assume that no other state is likely to change its system for selecting electors, your argument is likely to be a simple one that focuses on the added weight and attention your state might enjoy by not splitting its vote among a multitude of candidates. It also seems an essential step to forestall (again if you are Jefferson) the reelection of your rival, John Adams, since otherwise some of Virginia's vote will go, if not to Adams, then perhaps to some third candidate. Suppose, on the other hand, that you think it's possible (as in fact happened), that one or more states will respond to Virginia's actions. No doubt, your calculations will differ since now you must concern yourself with guessing which states are likely to change their method of selecting electors and who those changes will benefit. Thus, if you take the myopic view of supposing that your advice can treat the electoral schemes of all other states as fixed, your analysis is a decision theoretic one. But if, as actually occurred, a state such as Massachusetts responds by altering its method of choosing electors in order to aid its favorite son, John Adams, you best consider a game theoretic analysis that takes into account the likely responses of all other states.

Crime Control; Police Patrols and Crime Voting: Although we may not be experts, we all pretty much know the parameters with which a professional burglar or car thief must deal when they set out to ply their craft so as to minimize the likelihood of getting caught and convicted. Successfully implementing a crime may be difficult, especially if one is not a professional, but the decisions one makes seem straightforward and include such rules of thumb as "work fast, work at night, wear gloves, work quietly and discreetly." In this case, crime prevention requires an effective police force, a competent staff of prosecutors and perhaps an education process whereby ordinary law abiding citizens learn how not to make themselves a criminal's easy targets. Now consider a different system as practiced in feudal Japanese villages. If the culprit of some crime could not be identified, the villagers themselves voted on who they thought was guilty, wherein anyone receiving more than some pre-established threshold of votes was summarily banished without compensation or trial. For example, then, in Fuse village (currently Chiba prefecture) in 1696, three bales of rice stored for tribute were stolen. After 10 days of searching, the thief could not be identified with any certainty. The village chief, section leaders, and 131 peasants thus agreed to hold an election to identify the thief. As a consequence, the two highest vote-getters were banished from the village and three others who received one or two votes each were sentenced to house arrest.

In predicting the actions of a potential criminal in the usual case, a decision-theoretic analysis would most likely suffice. Using experience and common sense, we can suppose that all but the stupidest criminals can calculate the approximate likelihood of detection and apprehension under varied circumstances. This calculation, in combination with an assessment of the value of the crime in the event one is not apprehended, should suffice in providing a criminal with a good idea as to whether and/or where to strike. Correspondingly, those who have no intention of being criminals but who also do not wish to be victims can make the same calculations and take some simple measures to protect themselves. Similar calculations might apply to the example of Japan's crime voting system, but here

things are more complicated. Not only must potential criminals be concerned with the likelihood of being discovered, but people generally must worry about what might happen to them if the culprit is not identified. A person might be subjected to banishment simply because their neighbors and acquaintances don't like them or seek retribution for some otherwise long forgotten slight. It seems only reasonable to suppose that hatreds and grudges were often reflected in the ballots. One might anticipate, then, that one's general social behavior is likely to entail a good deal of concern about how one is viewed by neighbors and acquaintances. Indeed, one can readily imagine society evolving to exhibit a great deal of deferential and overly courteous behaviors, including seeing those prone to commit crimes acting with extreme deference in their everyday lives. In other words, this somewhat strange judicial system will most likely induce a variety of strategic calculations of the sort "Do I appear too deferential? Am I deferential enough?" To understand what if any equilibrium of social behaviors is likely to emerge in this case requires something more than a simple decision-theoretic analysis.

Anti-Missile Deployment: In the mid 1990's the United States set itself upon a course of convincing Poland and the Czech Republic that it was in their interest to allow the US to install anti-ballistic missile (ABM) technology on their territory. The argument offered by American strategists seemed straightforward: There are those in the Middle East intent on developing and deploying offensive missiles that could target Europe – Poland and the Czech Republic included. Armed with statistics on costs and the assessments of the likelihood that states such as Iran were pursuing the development of long range offensive systems, the argument for ABM seemed simple and incontrovertible. What that initial argument lacked, however, was an assessment of Russia's response, not only to a blunting of the capabilities of its client states, but of its own missile system since it seemed evident that a European-based ABM system could be directed at them as well as Iran. The Russians, unsurprisingly, were strongly opposed to the installation of any ABM system close to its borders, especially one controlled by its post-World

War II foe, NATO. They thereafter initiated a contentious negotiation with whoever occupied the White House that included the threat to reignite the arms race if an AMB system were installed. Ultimately the White House capitulated, ostensibly because it was attempting to secure agreements with Russia on other issues, while both Poland and the Czech Republic were left in the lurch after having committed to supporting American policy.

American policy here then illustrates the consequences of making foreign policy decisions by ignoring or by not being fully cognizant of the reactions of other relevant actors. We appreciate that a detailed historical analysis might tell us that one Presidential administration was fully cognizant of those reactions and preferred to ignore them while a subsequent administration was naïve in placing a different value on the threat of Russian retaliation and/or cooperation on other issues. Nevertheless, this example reveals how a decision-theoretic approach can yield one policy while an analysis that makes even a minimal attempt at anticipating the responses of others might yield something wholly different.

Grading on a Curve: When administering a final exam, an instructor generally has two choices –- to grade in absolute terms (i.e, an A requires a final grade of 90 to 100, a B requires 80 to 89, and so on) or on a curve wherein the class average grade is set at, say, B even if the class, in the instructor's judgment, does poorly. Suppose you are a student in some class wherein everyone has, by some miracle, received an identical grade of B on the midterm exam (or where, perhaps, the final grade is determined solely by one's performance on the final). If the instructor grades on an absolute basis, how hard you study for the final will, we can assume, depend on the things that might serve as distractions, on how well you think you've mastered the subject and on your personal motivation to strive for an A versus the possibility of your final grade slipping to a C. Alternatively, suppose the instructor grades on a curve. If no one studies and everyone again does equally well on the final, you and everyone else will maintain a grade of B. But if a number of other students study and you do not, they will raise the curve

and some mid-term B's, including yours, will become C's (or worse). Thus, how hard you choose to study for the final will depend not only on personal motivation and distractions, but also on how hard you think your classmates will study. But of course, how hard they study will depend on how hard they think others will study, including you.

Thus, while decision theoretic reasoning is most likely sufficient to predict the study habits of a student when the instructor grades on a fixed basis, a game theoretic analysis is required to account for behavior when grades are curved on a relative basis. In the case of an instructor who grades on an absolute basis, the study habits and performance of one's classmates is irrelevant to the ultimate determination of one's grade. But in the case of grading on a curve, not only is your final grade a function of the performance of one's classmates, but the effort one puts into studying for the final depends on your assessment of their actions, and by logical extension, your assessment of their assessment of your actions, and so on. This is precisely the sort of circumstance addressed by game theory.

Presidential power: If we look at the formal constitutionally proscribed powers of the presidency in the United States we see a position with few powers that cannot be checked by other political actors. An American president plays no formal role in amending the Constitution, his veto over legislative acts can be over-ridden by the legislature, he cannot make formal appointments without the approval of the legislature, he cannot implement treaties with foreign powers without Senate approval, there is no constitutional provision that the legislature must consider any legislative proposal he might offer, he has no authority over state and local level offices, and he is now precluded from serving more than two terms of office. Yet, the assertion that an American president holds one of the most domestically powerful offices in the world would seem self-evident. This view, though, seems to fly in the face of the fact that presidents of countries elsewhere hold far greater formal constitutional powers, including the authority to veto regional laws and to appoint and discharge regional executives. The

supposition, though, that by granting a chief executive strong constitutional powers necessarily renders that office powerful commits the error of confusing decision-theoretic with game-theoretic reasoning. We might conclude that wide ranging constitutional authority renders that office powerful, but only if we impose a strong *ceteris paribus* condition on the responses of all other political actors. On the contrary, constitutionally strong powers might merely energize opponents to resist those powers while at the same time leading anyone who holds that office to rely solely on those powers and nothing else. Weak constitutional authority, on the other hand, might lessen the natural opposition of others while simultaneously inducing those who hold that office to develop less formal avenues of authority. In the case of the American presidency, for example, those weak powers have encouraged presidents to cultivate political parties and the non-constitutionally proscribed ways in which they can exert power thru persuasion and the leadership of a party. Thus, to understand the implications of alternative political institutional designs requires a game-theoretic treatment rather than a decision-theoretic one – a treatment that examines how individual motives and choices influence each other as opposed to one that assumes the motives and choices of people are somehow fixed.

West Point Honor Code & Chinese self-reporting: The Honor Code as it is practiced in America's military academies such as West Point requires that, among other things, students report any observed instances of cheating. The code provides for consequences, moreover, in the event that someone observes cheating but fails to report it. Thus, this implementation of the code parallels a Chinese version that dates back to the Zhou dynasty (1088-221 B.C.), wherein a person failing to report a violation of the code is punished more harshly than if he or she had themselves committed the violation. In this scheme, we not only prosecute anyone who commits a crime, but we prosecute in a doubly harsh way anyone who had knowledge of the crime but fails to report it to the authorities. And to make this system even more interesting (and akin to "turning state's evidence"), suppose the perpetrator of a crime, after being caught, identifies

others who knew of his illegal actions and in so doing either receives a more lenient sentence or none at all.

With the distinction between decision and game theoretic reasoning in mind, we can perhaps see more clearly the difference between an honor code system that prosecutes only a person who commits a violation versus one that also prosecutes a person who fails to report a violation. Aside from the agonizing one might experience if a code's violator were a close friend, in the first case deciding whether to report a violation might hinge on one's assessment of the violation's severity. But in the second case, one also has to be concerned that the violator, in seeking to reduce their penalty, will report things on their own, in which case if you fail to report, you will be punished … possibly even more heavily than if you had been the one who originally violated the code. In the first case, then, your choice is a decision theoretic one whereas in the second it is game theoretic because you must anticipate the actions of another person who is, at the same time, attempting to assess the likelihood that you will turn them in.

It is also interesting here to compare the Japanese system of crime voting with China's self-reporting system. In the Japanese case a person's probability of being ostracized by his neighbors as a criminal depends only in part on whether or not they are guilty of the crime under investigation. It is not unreasonable to suppose that a good many persons were "wrongly" convicted merely because those around them deemed some aspect of their personality distasteful or disreputable. In response, one can readily imagine a system of social norms arising whereby acting in accordance with those norms avoids having such descriptive words as "arrogant," "unfriendly," "intemperate," "mean," "boisterous" and "combative" appended to one's character. However, the more fully those norms take hold, the more difficult it is to sort people by their degree of conformity to them, in which case, signaling one's conformity may require overt and accentuated actions such as ritualized bowing as if one were being presented to a monarch. The important point here, however, is that the evolution of such norms and their ultimate

manifestation must be viewed as the consequences of people's strategic interactions. If, for instance, one bows not enough, then that might be taken as a violation of the norm and a potential basis for people to vote for you in some criminal investigation. On the other hand, bowing too deeply might be viewed as a sign that one is indeed over-compensating for some prior criminal actions. There follows, then, a complex evolutionary process wherein people, across generations, learn and then codify 'proper' methods of social greeting. In the case of Maoist China, in contrast, a different pattern of social behavior is likely to emerge: Since the innocent must be concerned that a criminal might attempt to implicate them when caught, the best approach is to isolate oneself from society to minimize the chance of being associated with anyone who might be accused of criminal activity. Thus, in both Japan and China people must make game-theoretic decisions in assessing the reactions of their acquaintances to their everyday actions: How much deferential behavior is too much because it raises suspicions versus how much is too little and marks me as an ungracious and disliked member of the community? Or, do I dare make any friends at all since almost anyone might be a reader of pornography or of banned literature and likely to try to save his own skin by fingering me as an accomplice should they be discovered?

Fighting a war with allies: It might seem that in confronting Japan in WWII, America and Britain simply had to ensure the effective coordination of their actions and the efficient allocation of their resources. If so, then whatever was to be decided could be decided by the generals (or admirals), with perhaps the assistance of a staff skilled in organizing each country's industrial capacity. Aside from various inter-service rivalries, a decision-theoretic approach aided by such tools as operations research would appear to be adequate to the task of directing the actions of the two allies. Things, however, were a bit more complicated and only partially influenced by the shared goal of Japan's unconditional surrender. Britain (or at least Churchill) was also concerned about maintaining (or resurrecting) its empire and thus favored military actions and an allocation of

resources that facilitated the recapture of Malaya and Singapore, moving the Japanese out of Burma, and maintaining its control of India. The United States, in contrast (or at least Roosevelt) was wholly unsympathetic with this goal, and simple logistics seemed to dictate focusing its resources on a Pacific campaign. It was well understood, of course, that, with Britain focused on the German threat to its homeland that the main burden of the war against Japan would be borne by the United States. Nevertheless, cooperation was essential and to sustain it at an efficient level often required negotiation and anticipating the likely responses of one's ally. Churchill, of course, had to make certain it pursued a strategy that kept the US committed to its "Germany first" policy and that it didn't pursue its Asian and Southeast Asian goals in a way that left the American public to view it as simply another imperialistic power. And as America's input into the overall war effort increased and then surpassed Britain's, Churchill sought a strategy whereby it would remain a great power after the war. The US for its part needed whatever assistance Britain could supply, especially in airlifting supplies to China, along with the unflagging commitment of the other Commonwealth countries of Australia and New Zealand. And it understood as well that the reconstruction of Asia after the war would benefit from Britain's input. Thus, Anglo-American relations during the war could not be modeled in simple decision theoretic terms but were more akin to the give and take that often describe legislative coalitions and the trading of votes across legislation – processes that cooperative game theory seeks to address.

Although the political content of some of the preceding examples is minimal, each suggests that if all of politics entailed simple decision-theoretic reasoning, politics most likely would be utterly boring. But politics and the processes that characterize it entail, virtually by definition, the interactions of people wherein the consequences of their choices depend on what others do, and what everyone does depends on what everyone else does or is expected to do. Which candidate wins an election depends on the character and actions of his or her opponents; which bills pass a legislature depend on what vote trades individual legislators might make across even disparate legislation; what

international alliances form depend, at least in part, on an assessment of what counter-coalitions are likely to form and the actions of states absent from those alliances. In other words, individual decisions we might label political do not arise in a vacuum and are rarely predicated on the assumption that only one decision maker's actions are relevant. Politics, then, is inherently game theoretic and understanding political processes either from the perspective of explaining what has happened or from that of predicting what will happen necessarily requires understanding how participants perceive (or misperceive) the game(s) they are playing. And to do that requires that we understand how to represent and analyze those games, and here our examples give us some idea as to the components of that representation. Specifically, a careful description of each of the above scenarios requires at least the following:

1. The identities of relevant decision makers;
2. the choices confronting decision makers, including the order in which their decisions (choices) must be made;
3. a specification of outcomes and the linkage between choices and outcomes;
4. each decision maker's preferences over the set of possible outcomes; and
5. the perceptions of each decision maker about the components of the game that concern them.

In the case of grading on a curve, for instance, the relevant decision makers are the students, the choices confronting each is how much or how hard to study, the outcomes are final grades, the linkage between choices and outcomes is dictated by the instructor's grading scheme, and preferences are, presumably "a higher grade is preferred to a lower grade and, ceteris paribus, less effort devoted to studying is preferred to more work studying". And since we are ostensibly speaking of students who are at least semi-conscious of their educational environment, we can assume that their perceptions of things correspond to our description of them.

1.2 Preferences, Risk and Utility

In expanding on the preceding list of the things that comprise a potential game-theoretic representation, the easiest place to begin is with individual preferences. So suppose we start with an abstract list of outcomes $\boldsymbol{o} = (o_1, o_2, o_3, \ldots, o_n)$. In fact, to begin with the simplest quantifiable possibility, suppose the o's correspond to different amounts of money, where o_1 denotes a greater amount than o_2, that o_2 corresponds to a greater amount than o_3, and so on. It seems reasonable to suppose now that a person, ceteris paribus, will prefer more money to less so that o_1 is preferred to o_2, o_2 is preferred to o_3, etc. Moreover, given this preference, we can also say that o_i is preferred to o_j provided only that $j > i$. In this instance, then, a person's preferences are *complete* (i.e., between any two outcomes, o_i and o_j, o_i is preferred to o_j, o_j is preferred to o_i or indifference holds between them) and *transitive* (i.e., if the person prefers o_i to o_j and prefers o_j to o_k, then he or she prefers o_i to o_k).

To this point nothing seems exceptional and if there was nothing else to consider when abstractly describing preferences the reader could legitimately claim we have introduced the idea of complete and transitive preferences merely to add some academic jargon to the discussion. Unfortunately (or fortunately, depending on one's perspective), things can quite readily become more complicated. Consider, for example, what is likely to happen if one of the authors of this book were taken to an art museum and asked to state a preference between successive pairs of paintings. Given our somewhat pedestrian understanding of art, when shown paintings #1 and #2, we might say we prefer #1 because it has more blue in it. Then when shown paintings #2 and #3 we might indicate a preference for #2 because, while the intent of each artists is unintelligible to our eyes, we find #2's frame more appealing. Finally, when asked to choose between paintings #1 and #3 we cannot preclude the possibility that we would state a preference for #3 because we have yet to be exposed to the current self-proclaimed purveyors of fashion and lack an appreciation, as art, for a painting of a blue soup can.

One could write this example off as aberrant and assert that our models and theories of politics can be limited to those situations where people know their preferences. Of

course, excepting the tautological assertion that people are said to know their preferences only when those preferences match our assumptions, we are left with the question as to how and when we know what other people's preferences might be. Matters grow even more confusing, though, when we try to be anthropomorphic about things and attribute preferences or goals to groups such as when we seek to explain a state's foreign policies while treating a state as a unified entity. Consider for instance the problems one encounters with attempting to assess Britain's goals prior to the outbreak of WWI. It seems easy to focus on its treaty commitments with France, its commitment to Belgian sovereignty and its longstanding policy of working against any one country becoming predominant on the continent in explaining its commitment of troops to the defense of France. But there were confounding matters. First, an equally salient issue for Britain at the time was that of home rule for Ireland and the conflict between Northern Ireland and the South. Policy makers in London could not discount the possibility that maintaining peace there would require whatever military resources it might otherwise allocate to the Continent. Second, there was a simmering diplomatic conflict with Russia over Iran. Britain was converting its navy from coal to oil, which required Iran's oil resources. But Russia was also attempting to extend its influence there, and, if one looked at its history with respect to the expansion of its territory, perhaps its sovereignty as well. So why join in an alliance, via France, with Russia against Germany? Indeed, Germany could be viewed as a counterweight to Russia in the rapidly decaying Ottoman empire and, in particular, in helping forestall Russia's longstanding designs on Constantinople. It was anything but clear at the time, both to outside observers and to some within Britain's government, how these concerns would play out in dictating policy. At a minimum, attributing coherent transitive preferences to Britain then was fraught with difficulty.

We will, in fact, have other more theoretically exact reasons for questioning the advisability of attributing goals to groups, but setting such things aside for the moment, consider another problem with the preceding representation of preference, which concerns the possibility that outcomes arise only up to some probability. To see the

problem here, suppose you are asked how much you are willing to pay to play the following "game" denoted The St. Petersburg Paradox, named after Daniel Bernoulli's presentation of the problem and his solution in 1738 in *The Commentaries of The Imperial Academy of Science of St. Petersburg* (although the problem was first stated by his cousin, Nicholas, in 1713): A fair coin will be tossed and if it comes up heads, you will be paid $2 and the game ends. But if it comes up tails, the coin will be tossed again and if it comes up heads on that second toss, you will be paid $4, and the game will then end. But if it comes up tails twice in a row, the coin will be tossed yet a third time, and so on until a heads finally appears, so that if a heads first appears on the *n*th toss, you will be paid 2^n dollars. Usually, now, when people are asked how much they are willing to pay to play this game, few give an answer in excess of $20. Consider, though, the game's *expected* dollar value. The probability of earning only $2 is ½ (the probability that a heads appears on the first toss); the probability of earning $4 is ¼ (the probability of a tails on the first toss times the probability of a heads on the second); … the probability of earning $\$2^n$ is $1/2^n$ (the probability of n-1 tails followed by a heads), and so on. Thus, the *expected dollar return* is

$$\$2(1/2) + \$4(1/4) + \ldots + \$2^n(1/2^n) + \ldots = \$1 + \$1 + \ldots + \$1 + \ldots = \infty$$

That is, the expected payoff from this game expressed in dollars is infinite -- an infinite summation of 1's. We seriously doubt, however, that most people who initially said they wouldn't pay more than $20 to play this game would, after shown this calculation, increase their willingness to pay by more than a few dollars (if anything at all).

Now consider a second observation about human behavior: The vast majority of homeowners buy insurance that protects them against the possibility of their homes burning down or of someone tripping on their basement stairs and suing for bodily injury. We also know that the big prize from state lotteries commonly achieve a value of upwards of eight or nine digits and that hundreds of thousands if not millions of people buy lottery tickets in the hopes of winning that mega-prize. It seems safe to assume, then, that there are a considerable number of people who buy both insurance and lottery

tickets. But in one instance (buying insurance) a person is exhibiting *risk averse* behavior with respect to money, while in the second instance (buying a lottery ticket) that same person is exhibiting *risk acceptant* behavior. In the case of insurance, people are spending money to avoid risk while in the case of a state lottery people are spending money in pursuit of risk. And in both cases the expected return on their "investments" are negative since neither insurance companies nor state lotteries are in the business of losing money. More formally, suppose a person is presented with a lottery that affords them a probability p of receiving \$$X$ and (1-p) of \$$Y$, where $X < Y$ and where the expected dollar value of the lottery is $pX + (1-p)Y =$ \$$Z$. If given a choice, now, between \$$Z$ with certainty versus playing the lottery, a risk acceptant person prefers the lottery while a risk averse person prefers the certainty of \$$Z$. Thus, if given a 50-50 chance of winning \$100 versus nothing, a risk acceptant person would choose the lottery to an offer of being given \$50 with certainty whereas a risk averse person would take the fifty dollars.

It is important to note that nothing said here contradicts the reasonable assumption that people prefer more money to less or negates the assumptions of transitivity and completeness since our discussion merely introduces a new consideration into people's choices; their assessments of risk. In the case of the coin toss, it is surely true that $\$2^n$ is a considerable amount of money when n is large, and it doubtlessly remains true that $\$2^n$ is preferred to $\$2^{n-1}$. But it is also true that the number paired with $\$2^n$, the probability of winning that amount ($1/2^n$), is quite small for large n –- so small in fact that a person might reasonably choose to ignore the term entirely as a feasible possibility. Alternatively, while buying insurance and lottery tickets may also entail small probabilities and considerable sums of money, the choices here are qualitatively different. In the case of insurance, one is trading the certainty of an insurance premium for a guarantee against the threat of a disagreeable lifestyle-changing loss whereas in the case of the lottery ticket one is trading the certainty of a small loss (the cost of the ticket) for a potentially wondrously lifestyle changing gain. And we should not be surprised that people will somehow treat risk differently, depending on whether we are speaking of significant gains versus significant losses.

What we require, then, is a way of representing preferences that parsimoniously summarizes people's attitudes toward risk. That thing is the concept of *utility*. So suppose instead of our previous coin toss calculation we instead, for the left hand side of the equation, write

$$U(\$2)(1/2) + U(\$4)(1/4) + U(\$8)(1/8) + \ldots + U(\$2^n)(1/2^n) + \ldots$$

with the assumption that $U(\$2) < U(\$4) < U(\$8) < \ldots < U(\$2^n) \ldots$ That is, suppose we define the function $U(x)$ such that it increases monotonically as x increases and require that $U(x) > U(y)$ if and only if the person prefers x to y. Then surely we have not violated the assumption that a person prefers more money to less. But we have instead substituted for that statement the requirement that "the greater the amount of money, the greater is that person's *utility*."

If one asks now about the form of the function $U(.)$ it is here that we gain our handle on representing preferences over choices that entail risk or uncertainty. First, notice that there is no reason to suppose that $U(\$)$ is a *linear* function of money – that the utility of a $1 increase in wealth is invariant with the amount of money a person currently has in their wallet. Indeed, speaking for ourselves, we can honestly say that the utility of, say, ten million dollars given our current salaries would surely outweigh the utility or pleasure we'd likely derive had we already been in possession of a hundred million dollars. At least for the authors of this volume, when speaking of substantial sums, money exhibits diminishing marginal value (and we are open to anyone who might wish to test this hypothesis). At the same time, differences in the value of various sums of money will vary depending on the range over which those differences will apply. In the case of insurance and lottery tickets, suppose the potential loss of one's home from a natural disaster or the amount we can be sued equals $\$X$, and that an insurance policy that protects us against such a possibility costs $\$Y << \$X$. Thus, if the perceived probability of incurring that loss is p, we are then choosing $U(-\$Y)$ over the lottery $pU(-\$X) + (1-p)U(0)$. At the same time, suppose we are one of those people who, when the potential winnings from a state run lottery reach, say, $Z, we run out and immediately spend $W on lottery tickets. If the probability of winning the lottery is q, our actions reveal a

preference $qU(\$Z) + (1\text{-}q)U(\text{-}\$W)$ over $U(0)$. Regardless of how small p and q might be and regardless of how large Y and W are, there is nothing in the definition of preferences or utility that renders these two preferences necessarily inconsistent. Indeed, we would be surprised to learn that the average person is anything but risk averse when confronting lotteries that entail large potential losses and risk acceptant when dealing with lotteries that open the door to large potential gains.

With respect to the St. Petersburg Paradox, to represent the idea that increasing amounts of money exhibit diminishing marginal value, suppose for purposes of a numerical example that $U(\$X) = X/(X+1)$. With this assumption the value of tossing a fair coin becomes, in expected utility terms, the sum of the infinite series

$$(2/3)(1/2) + (4/5)(1/4) + (8/9)(1/8) \ldots$$

which sums to approximately 0.77. If we now set $X/(X+1) = 0.77$, we find that $X \approx 3.35$. Thus, if a person's utility for money abided by the admittedly ad hoc function $X/(X+1)$, he or she should be willing to pay no more than \$3.35 to play our coin toss game (as opposed to infinity).

As a side note, we emphasize that no one has ever seen a utility function (aside from those which academics postulate on paper). Utility is a contrived concept developed for the purpose of representing people's preferences over risky alternatives. Thus, they serve much the same function as did the concept of the electron in the 19th century. No one had ever seen or at the time hoped to see an electron, but positing its existence (and here credit is due to Benjamin Franklin) explained the observable phenomena associated with positive and negative charge. This isn't to say that someday we will not find a better and more theoretically satisfying way to deal with the complexities of individual choice. It may be that the concept of a utility function will have a half-life no greater than that of the ether, which scientists once thought necessary to explain the transmission of light.

So restating our assumptions about individual preferences, the requirement that preference is a complete relation is akin to supposing that between any two outcomes, o_1 and o_2, either $U(o_1) > U(o_2)$ or $U(o_1) < U(o_2)$ or $U(o_1) = U(o_2)$. Transitivity, in turn,

requires that if $U(o_1) > U(o_2)$ and $U(o_2) > U(o_3)$, then $U(o_1) > U(o_3)$. In other words, we require that U act much like the natural number system. There is, though, one additional requirement. Suppose $\boldsymbol{p} = (p_1, 0, 1 - p_1)$ is a lottery that assigns o_1 the probability p_1, o_2 the probability 0, and o_3 the probability $1 - p_1$, suppose $\boldsymbol{q} = (0, 1, 0)$ is a "lottery" that assigns the probability 0 to both o_1 and o_3, and certainty to o_2, and suppose $U(o_1) > U(o_2) > U(o_3)$. Then a person is said to prefer $\boldsymbol{p}$ to $\boldsymbol{q}$ (or equivalently, $U(\boldsymbol{p}) > U(\boldsymbol{q})$) if and only if $p_1U(o_1) + (1 - p_1)U(o_3) > U(o_2)$. In other words, we assume that a person's utility function can be defined so that it not only represents a person's *ordinal* preferences over outcomes, but allows us to represent that person's preferences over risky prospects in terms of his or her preferences across the specific outcomes over which the risk is spread.

Before we elaborate on the concept of a utility function and some problems with it, let us first consider some examples to better appreciate the role risk plays in individual decisions making:

> **Risk, Traffic Control and China's Media:** People's attitudes toward risk can sometimes go a long way in explaining government policies or in understanding how governments might manipulate individual choice by manipulating risk. Consider, for example, China's policy with respect to its mass media. If one questions newspaper editors, columnists, and so on there, you will quickly learn that Beijing's policy seems at times mercurial – sometimes it is harsh and at other times lenient, with no apparent pattern to it forbearance of criticism. It is, of course, entirely possible that this ebb and flow merely reflects the shifting fortunes of individuals in authority within the PRC hierarchy. But consider the possibility that a mercurial policy is wholly intentional and intended to keep publishers, commentators, newspaper editors and the like in line. Here the argument would be that with no clearly delineated and seemingly coherent policy, the PRC leadership is essentially making the likelihood of punishment a lottery – and if, as is likely, those publishers, etc are risk averse with respect to their careers, they will adhere to a more docile and constrained agenda than if the regime established a hard and fast rule. Under a stable and well-defined rule we

can expect that publishers will "walk up to the line" as closely as possible and even, in a few cases, cross over it for short periods of time, knowing precisely when they are in compliance with the government's policy. But under an uncertain or unclear rule, individuals will make risk-avoidant choices and adhere more carefully to Beijing's ultimate (but imperfectly publicly stated) goal. To see what we mean here in a different context, consider normal behavior on a Los Angeles freeway unencumbered by gridlock (yes, that happens on occasion). With a posted speed limit of 65mph and a general understanding that the police rarely tickets anyone driving within 10mph of that limit, most traffic will move along at 75mph and a few drivers will push the envelope a bit. Suppose instead that the state highway patrol adopts the publicly stated policy of choosing a number at random between 70 and 80 every day at midnight, and, without publicly revealing that number, tickets everyone who exceeds it on that day. Now we would expect the same average speed limit – 75mph – that was the de facto limit before, but the question is: How might the behavior of drivers change? If drivers are risk averse with respect to receiving speeding tickets and the time lost spent by the side of the road while the officer writes the ticket, our answer should be a decrease in average driving speeds to something below 75.

A Crime Control Proposal: In the attempt to insure that people and convicted criminals in particular are not unduly penalized merely because of their race, ethnicity or economic status, state and local governments in the US have, over time, instituted an admittedly varied system of sentencing guidelines for judges wherein two people convicted of a similar offense receive the same or approximately equivalent sentences. Such guidelines, then, like a posted speed limit, define one's sentence for, say, a first, second and third conviction of automobile theft or shop-lifting. But suppose instead of penalties being defined in terms of fines or length of time incarcerated in a prison we instead formulate those guidelines as a probability – a probability of being put to death. Thus, when convicted of some minor offense (e.g., failing to stop at a stop sign) the assigned probability will be small (hopefully VERY VERY small). But when convicted of

murder, that probability will be significant, perhaps even 0.99 or 1.0. Following conviction, a lottery will be conducted in accordance with the assigned (sentencing) probabilities with the outcome of the lottery dictating whether that person will be immediately set free or put to death. We don't know about the readers of this volume, but we do know that in such a system, the authors herein would most definitely be very careful about stopping at every stop sign we encounter when driving.

China, Taiwan, the United States and Strategic Deterrence thru Risk: The case of the United States policy of strategic ambiguity toward the dispute between China and Taiwan serves as an additional illustration of the strategic manipulation of risk. China believes that Taiwan is but a renegade province, that the island's reunification with the Mainland is a domestic issue, and that force may legitimately be used to compel reunification. There is widespread agreement, however, that China at the present time prefers the status quo to entering into a military conflict with the United States over Taiwan. Taiwan, on the other hand, refuses to acknowledge the People's Republic of China as the legitimate representative government for all of China, and seeks increased international autonomy. It is also commonly believed that Taiwan prefers to be de jure independent from the PRC regime but prefers its de facto political independent status to fighting China without American assistance. Most strategic analysts agree that the US prefers the status quo to all other feasible outcomes. The US, then, faces a standard dual deterrence dilemma: Announcing a policy of under-commitment to Taiwan raises the incentive for China to secure reunification by force, whereas a policy that over-commits to Taiwan's defense risks emboldening Taiwan to move recklessly toward independence and, thereby, compelling China to upset the military status quo. Beginning with President Eisenhower in the early 1950s, the US policy response has, therefore, been to be strategically ambiguous about the conditions under which it will defend Taiwan. Specifically, the policy of strategic ambiguity, which derives formally today from the Taiwan Relations Act, acknowledges that there is only one China, that Taiwan is part of China, that

resolution of the Taiwan issue is a domestic matter, but at the same time regards any security threat to Taiwan as a "grave concern" to the US. This seemingly contradictory policy has the effect of signaling that the US has a definite stake in the outcome of the conflict but prefers to abdicate any "first move" to China or Taiwan while leaving both sides uncertain as to its ultimate response to any change in the status quo. Uncertain about how the US will respond, neither China nor Taiwan has chosen to take decisive provocative action, and as long as the US enjoys an asymmetrical power advantage over both China and Taiwan, strategic ambiguity offers a better shot at maintaining things as they are than does strategic clarity.

The preceding examples demonstrate that the sources of risk need not derive from the things we don't know or cannot predict about "nature" such as the weather, but also include those risks deliberately contrived as an element of individual strategy. A good part of this volume, then, will consider the manipulation of risk as a strategy in human interactions. But before we do so, we need to confront the fact that when it comes to the analysis of risk and our treatment of preferences, neither life nor the study of politics is ever simple. To wit, consider the following three outcomes:

$$o_1 = \$5 \text{ million}$$

$$o_2 = \$1 \text{ million}$$

$$o_3 = 0$$

Now we would like the reader to carefully consider these two lotteries over the outcomes:

$$\boldsymbol{p} = (0.10, 0.89, 0.01) \text{ versus } \boldsymbol{q} = (0, 1, 0)$$

After making the bold attempt at putting yourself in a situation where you might actually get to make such a choice, which would you choose? Done thinking? Now give some serious thought to the following two alternative lotteries:

$$\boldsymbol{p'} = (0.10, 0, 0.90) \text{ versus } \boldsymbol{q'} = (0, 0.11, 0.89)$$

It has been our experience now that when students (and most everyone else, including ourselves) are asked to choose between $\boldsymbol{p}$ and $\boldsymbol{q}$, a good share chooses $\boldsymbol{q}$ after reasoning that "a bird in the hand is worth two in the bush." Or, "a lot can be done with one million dollars, and think of the regret if $\boldsymbol{p}$ were chosen instead and I ended up with nothing."

Now, when asked to choose between ***p'*** and ***q'*** a reasonable share of people who initially chose ***q***, would choose ***p'*** over ***q'*** with the rationalization that "there isn't much difference between the likelihood of getting five million with ***p'*** as opposed to one million with ***q'*** so why not go for the big bucks."

We would hardly label these two choices – ***q*** over ***p*** and ***p'*** over ***q'*** – as irrational or illogical since they might be the ones we ourselves would make. The problem here, though, is that no utility function is consistent with them. By indicating a preference for ***q*** over ***p***, it must be that

$$0.10U(\$5 \text{ million}) + 0.89U(\$1 \text{ million}) + 0.01U(\text{nothing}) < U(\$1 \text{ million})$$

or equivalently,

$$0.10U(\$5 \text{ million}) + 0.01U(\text{nothing}) < 0.11U(\$1 \text{ million})$$

However, the choice of ***p'*** over ***q'*** requires,

$$0.10U(\$5 \text{ million}) + 0.90U(\text{nothing}) > 0.11U(\$1 \text{ million}) + 0.89U(\text{nothing})$$

or , after rearranging terms

$$0.10U(\$5 \text{ million}) + 0.01U(\text{nothing}) > 0.11U(\$1 \text{ million}),$$

which directly contradicts the implication of a choice of ***q*** over ***p***. What's going on here? There are, we suppose, any number of possible explanations for such seemingly inconsistent choices, but the one that especially appeals to us is that the 0.01 difference in the likelihood of getting nothing between ***p*** and ***q*** is not being evaluated in the same way as is the difference in these likelihoods between ***p'*** and ***q'***. In the first case moving from ***q*** to ***p*** renders something that is impossible (getting nothing) possible, whereas in the second case, moving from ***q'*** to ***p'***, something that is likely merely becomes a bit more likely. In other words, the 0.01 difference between the pairs of lotteries of coming away empty handed, while treated identically in an algebraic manipulation, has a different psychological impact in the two sets of decisions.

It would seem, then, that not only is the value we place on objects wholly subjective and dependent on context (e.g. how we value a million dollars depends on whether or not we are already rich), but the probabilities we associate with risky choices are subjective as well and dependent on context. Unsurprisingly, this fact is widely recognized by decision

theorists and considerable effort has been given to seeing what generalizations can be devised about *subjective probability* – probabilities that do not necessarily adhere to the rules we impose on them in mathematics and statistics. In this volume, however, we will make little use of that research since it only complicates our attempt to lay out the fundamentals of game theory as applied to politics and since very little of that research has yet been applied to the study of politics. Thus, throughout this volume we will treat probabilities in much the same way a statistician might by assuming that they obey the laws of algebra, that they do not fall outside of the range [0, 1], and that when summed across all feasible outcomes for a particular problem, that sum equals 1.0. Once again, though, we realize that individual behavior will often violate this assumption and it is important that we keep this fact in mind before we draw too strong a conclusion from any analytical exercise.

Why Vote?: To perhaps better appreciate the role subjective probabilities might play in politics, consider the simple act of voting in mass elections. At least in a democracy there is perhaps no more fundamental act of citizenship than that of casting one's ballot for or against a candidate, a party or some proposition on a referendum. But suppose we ask why people vote. This might seem a question with a simple answer – people vote because they want to increase the likelihood their preferred outcome prevails. Presumably, however, voting is a costly act. Even if one ignores the costs of becoming informed about the alternatives on the ballot, it requires an allocation of time to simply get to the polling station and in important elections people have been known to stand in line for hours waiting for their turn to enter the voting booth. So, proceeding to some minimal formalism, let P be the probability that your favored candidate in a 2-candidate contest wins if you do not vote, P' be that probability if you do vote, U be the value you associate with seeing that candidate victorious, U' be the value associated with that candidate loosing, and C the cost of voting. Then ignoring any algebraic complexities occasioned by the possibility of making or breaking ties between the candidates, the expected utility of not voting and not incurring the cost C is

$$E(\text{NV}) = PU + (1\text{-}P)U'$$

while the expected utility of voting is

$$E(\mathrm{V}) = P'U + (1\text{-}P')U' - C.$$

Presumably, then, a person should vote if and only if $E(\mathrm{V}) > E(\mathrm{NV})$, or equivalently,

$$(P' - P)(U' - U) - C > 0.$$

Admittedly, now, for people who intensely prefer a candidate, the difference $U' - U$ may be considerable. But consider $P' - P$, which is the probably of being pivotal in the election in terms of making or breaking ties. Such a probability might not be small if we are considering an election in some village with 100 or so voters. But what of a national election with millions of voters? Surely the probability of being pivotal then fades to insignificance. Indeed, to say that your favored candidate is more likely to win if you vote for him rather than abstain is akin to saying you are more likely to hit your head on the moon by standing on a chair. But if $(P' - P)$ is essentially zero, and if C is consequential, then few people should vote. Since this prediction is clearly at odds with the data, we must ask again why people take the time to cast ballots in mass elections.

There are, in fact, two alternative explanations for non-zero turnout in mass elections (aside from those countries that fine people for not voting). The first hypothesis is that voting gives people a sense of fulfilled citizen duty – a warm feeling in the tummy you might say. That is, we might suppose that people derive utility from the mere act of voting regardless of what impact their vote has on the election outcome. Equivalently, we might suppose that failing to vote is costly. Anyone living with a 12 year old daughter or granddaughter who, on the basis of what she has been taught in school, regards her parents or grandparents as beneath contempt if they do not vote understands this cost. An alternative hypothesis (which does not preclude the first explanation from applying) is to suppose that people, subjected with mass media reports of how close an election might be, subjectively over-estimate $(P' - P)$. In fact, it is possible that people partake of a rather strange form of backwards causality, reasoning that "there are millions of people like me, and if I decide not to vote, they most likely would reach the same

decision. But if I decide to vote, they will as well because their thinking will be the same as mine. Thus, my decision isn't merely impacting one vote but millions." Such thinking, of course, inflates the probability that one's vote is pivotal, and far be it for us to say that such reasoning cannot describe the inner workings of the mind.

Some academics object to the idea that people vote because of a sense of citizen duty, arguing that such a supposition merely makes voting rational by assumption and thus tautological. However, it is no more tautological to suppose that people vote because they have been socialized to value the simple act of voting for its own sake any more than to say a person buys a red as opposed to blue car because he or she likes red. Similarly, to suppose that people partake of a seemingly perverse view of causality when voting might seem strange, but in modeling people we best be prepared to learn that the human brain can entertain or seemingly employ forms of logic that defy logic. Be that as it may, there is one final qualification we need to add to our presentation of the concept of utility.

To this point we've made the assumption, when speaking of money, that $U(\$X) > U(\$Y)$ if and only if $X > Y$. However, suppose to the description of outcomes we append the date at which the money is received. Specifically, suppose $\$X$ is "One hundred dollars a month from today" and $\$Y$ is "Fifty dollars today." In other words, even when speaking of a simple thing like money we suppose outcomes are multidimensional and their descriptions include not only the quantity of money but also when its is received. Here we know that people's preferences vary. Some will prefer receiving the fifty dollars immediately whereas others will prefer to postpone things provided they are compensated by a larger amount. In other words, people's preferences are defined not only over monetary amounts but also over time. Such possibilities require a representation, and perhaps the simplest is to add a discount to the timing of an outcome, where that discount is calibrated by the units of time under consideration. For example, for $\$X$ received next month we might write δX, where $0 < \delta < 1$ since presumably people will prefer $\$X$ today to $\$X$ next month. And for $\$X$ two months from now we can doubly discount and write $\delta^2 X$, and for three months from now $\delta^3 X$, and so on.

Time discounting applies to things other than monetary outcomes. For example, it is well-known that the behavior of drug addicts is only imperfectly impacted by a knowledge of the long term medical consequences of their addiction. Attempting to cure addiction by educating the addict about the harmful medical consequences of their problem will almost certainly fail. This is because addicts, nearly by definition, have an overly strong preference for immediate self-gratification as opposed to the long term benefits of abstention and recovery. Indeed, one might say that "getting hooked on drugs" is shorthand for saying that the drug itself alters a person's time discount. Time discounts can also be impacted by one's environment and the time discounts of others. Suppose you are contemplating an investment in a society rife with political corruption and where most persons, as a consequence (and as is arguably the case in many of the states of the former Soviet Union), act with very short-term horizons. Those short horizons derive from the fact that, in a truly corrupt state where there is no line between the criminal and the government, the government today might encourage your investments but tomorrow, after being bribed by your competitors, act to confiscate everything. Confronted with such a state, most people would naturally prefer, when making any investment, to "take what they can and run." But this means people will have few incentives to abide by long term contractual agreements, in which case, anyone entering that economy with a long term planning horizon will be akin to a small fish in a pool of sharks.

1.3 Economics versus Politics and Spatial Preferences

The notion of time discounting will bear substantive fruit later when, in addition to the matter of political corruption, we consider such things as how political constitutions survive or fail as well as how cooperation in any form emerges in a society. But before we proceed to modeling specific political processes or phenomena, we note that when attempting to theorize about politics or to construct a model of some particular political process it behooves us to use the weakest assumptions possible, if only to ensure the greatest generality of whatever insights we might establish. But while generality has a self-evident value, it is unfortunately also the case that the weaker our assumptions, the less substantively precise are our insights. Thus, theorizing about anything, be it physics,

chemistry, biology, economics or politics, entails maintaining a balance between generality and substantive specificity. Political science, though, is a discipline that stands relatively high on the food chain of our knowledge of and theories about social processes. Indeed, one might even draw a parallel between various fields of engineering and design versus the more fundamental fields of physics, chemistry and mathematics. Political science is (or at least should be) an applied field that takes what we know from statistics, from decision theory, from psychology and from game theory (as well as from the other social sciences) and applies what is known to the social processes we label political, ostensibly with the goal of improving the lot of our species. Thus, while the political scientist is not required to be a game theorist per se who goes about proving mathematical theorems about this or that, he or she is expected to be able to say something about such things as constitutional design, coalition formation in legislatures and parliaments, the imperatives of various forms of democratic governance, the sources of international peace versus war and the operation of alternative electoral processes.

The engineer who wishes to design a more efficient gas turbine or faster aircraft illustrates the parallel in the physical sciences. While the engineer is not expected to advance fundamental laws of physics and thermodynamics, he must nevertheless make use of those laws (or at least not presume that a design can violate them) in a creative way, filling in the blanks of abstract constructs with specific measurements or assumptions while at the same time making approximations that allow for the formulation of a substantively (physically) meaningful design proposal. The same is true in economics wherein those attempting to gauge trends in interest rates or the impact of some regulatory edict on firm behavior know that the laws of supply and demand will constrain events. And just as those elementary economic principles begin with highly abstract formal representations of consumer preferences and firm objectives, the political scientist, when modeling political processes, must often begin with abstract representations of preference and uninterpreted functions that denote utility, supplying them later with specific substantive meaning.

The Grocery Store: To see what we mean by all of this let us attempt to gain a better understanding of the differences between economics and political science (without presuming that these two disciplines are necessarily disjoint) with a somewhat fanciful scenario. Consider the simple act of purchasing groceries in a supermarket, but to make our life simple, suppose there are but two distinct commodities in that store, X and Y. Your decision, then, is to choose how many items of X to buy, denoted x, and how many of Y to buy, denoted y, where your decision is subject to a budget constraint, B. Thus, the most of X and Y you can purchase is $xp_X + yp_Y = B$, where p_X and p_Y are the per unit prices of X and Y respectively. Assuming that you prefer as much of X and of Y as possible (i.e., you don't confront a problem of storing either commodity and neither is perishable), we can assume you'll balance off your purchases of these two goods so as to maximize your overall utility.

This much, of course, is little more than the introductory chapter of Elementary Economics 101 and corresponds to the economist's classical representation of a trip to the grocery store. Now, however, imagine a somewhat modified scenario. Rather than visit the store whenever you feel the need to replenish your supply of X and Y, suppose you are assigned a specific day and time to go to the store and that you are also required to bring with you a certain amount of money. Upon arriving at the store you find that 100 other people have been assigned the same time as you to shop and have been told to bring the same amount of money with them. However, upon entering the store the door is locked behind all of you, you are all led into a back room and told, after your money has been collected, that what you purchase today will be determined by a majority vote among all 101 of you. More precisely, suppose two of you are chosen at random and labeled "candidates". Each candidate must then propose a package of X and Y whose cost equals the sum of money collected from you with the presumption that everyone's budget will be spent in an identical way. The 99 of you who are now designated "voters" must then vote for one of the two candidates, and the candidate receiving the most votes will be declared the winner. Each voter and

both candidates will then be given a shopping bag that matches the proposal of that winner with the winning candidate receiving an additional side payment of some sort so that both candidates have an incentive to win (as opposed to merely proposing their ideal allocation as their "campaign platform").

This might seem a truly strange way of organizing grocery shopping, but it does illustrate some of the differences between economics and politics, which in this case is simply the difference between two ways of allocating the goods and services people value. Now, though, consider the full implications of this difference. In the more regular way of allocating groceries each person is free to choose the combination of X and Y that best serves their tastes whereas in the second each person is, in effect, a prisoner of the tastes of a majority or of the two competing proposals offered by the candidates. In the economic realm, then, we might attempt to predict how many of X and Y will sell by a careful examination of individual consumer tastes with the understanding that ultimate demand will equal the sum of demands. In the more collective or political realm, on the other hand, ultimate demand will depend on learning what proposals the candidates are likely to make and how voters will vote when confronted by alternative proposals. In the economic realm we need only identify that specific combination of X and Y that maximizes a consumer's utility, given their budget constraint. In the political realm, in order to learn how they might vote between the proposals of the two competing candidates, we must also be concerned with what their preferences look like over combinations they might not choose were they dictator of their own budgets.

We warn that we should not draw too sharp a distinction between economics and politics, since often politics entails deciding how to organize our "shopping" – should, for example, the purchase of heath care insurance be a private or public matter, should people be free to discriminate against certain classes or races when selling their own home, and should even a long-established public retirement program be made a partially private affair with both public and private options? Surely, few would argue that the

answers to these questions are straightforward or without controversy. The same is true with our grocery store example. Suppose X and Y correspond to beer and baby food, and suppose that a clear majority of the 101 people sharing the back room of the supermarket are mothers with babies. But suppose you are an unmarried male. I suspect you would then hold a strong preference for the usual way of buying groceries (unless you have a perverse taste for crushed peas and strained carrots). Conversely, your preferences for how grocery shopping might best be organized could change if mothers with babies constituted only a minority of those present. Absent a concern for mothers with hungry and crying babies, you might see this as an opportunity to have parents subsidize your consumption of beer.

Politics is often a choice of how to allocate goods and resources – what to relegate to the private sector and what to allocate by some collective process. But in making that decision it is important to understand how different institutional forms -- different methods for making social decisions – perform. For example, in lieu of selecting two persons at random to play the role of candidates, suppose we simply let the 101 people in the room negotiate directly among themselves until a majority reach an agreement and terminate further discussion? Or suppose we divide them into three constituencies of 33, 34, and 34 people, let each of them in a manner of their own choosing select a representative who will then negotiate with the two representatives from the other constituencies until they reach an agreement? What difference, if any, will each of these schemes imply in terms of the agreements reached?

To answer such questions – to conduct a comparative analysis of institutional forms -- requires a common underlying structure for modeling the alternatives and individual preferences over the outcomes with which they deal. Returning, then, to our two commodities X and Y, and for a specific (albeit arbitrary) analytic example, let us suppose that the utility a person associates with a combination of X and Y is given by

$$U(x, y) = [5 - 5/(x+1)] + [4 - 4/(y+1)]$$

As complex as this expression might seem, it has a simple interpretation: If $x = y = 0$, then $U(0, 0) = 0$, but as either x or y increase, the subtractions in the expression decrease

at a decreasing rate. Thus, as x or y increase, $U(x, y)$ increases, but at a decreasing rate and approaches the upper bound of 9 as the amount of both commodities approaches infinity. The two commodities, though, are not perfect substitutes. For example, $U(2,0) = 10/3$ whereas $U(0,2) = 8/3$. In other words, if you have two units of X, you would require more than two units of Y to be compensated for the loss of your holdings of X. The relationship between X and Y in a person's preferences can be portrayed, then, as in Figure 1.1. The horizontal axis denotes units of X while the vertical axis denotes units of Y. The curves in turn correspond to *indifference curves* – combinations of X and Y that yield the same value for $U(x, y)$ and where combinations on curves further from the origin are preferred to combinations that fall on curves closer to the origin. Figure 1.1 also portrays a person's decision when choosing some combination of X and Y subject to a budget constraint. Here we assume that the per unit cost of X exceeds that of Y, so if a person spends their entire budget on one commodity, they can buy more units of Y than of X. Finally, the indifference curve that is tangent to this line represents the highest level of utility our decision maker can achieve, given their budget, so that x^* and y^* are the combination of good we can assume they will purchase if they are dictator over their purchases.

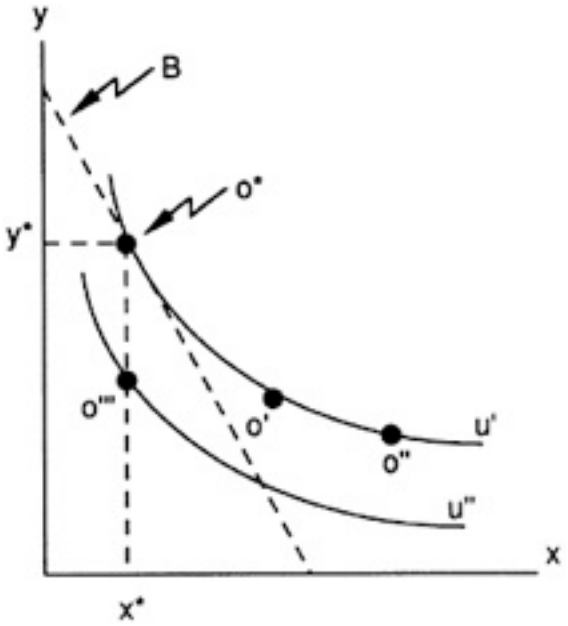

Figure 1.1: Traditional Economic Indifference Curves

Figure 1.1 is common to any introductory economics text and its discussion of consumer behavior in markets. But now let us again shift back to our peculiar (collective) method of grocery shopping. Here a person can no longer ensure that (x^*, y^*) is chosen, since the outcome depends on the preferences of other voters and the packages proposed by the candidates. In this instance, any point along the budget constraint line is a possibility (recall our assumption that each person brought the same sum of money to the store). But notice that the shape of the indifference curves in Figure 1.1 tells us something about the nature of this person's preferences across that line. Specifically, if we label the point o^*, which corresponds to the combination (x^*,y^*), the person's *ideal point*, then as we move away from that point in either direction, we move to lower and lower indifference curves. That is, the further we move from o^* the less our abstract person/voter likes it.

If we now lay out the budget constraint line horizontally, we can draw a *preference or utility curve* such as the one shown in Figure 1.2, which for obvious reasons we refer to as a *single peaked preference curve*. The horizontal axis now corresponds to different allocations of the person's budget between X and Y, while the vertical axis denotes the person's preference or utility – which we know decreases as we move from o^*, either to the left or right.

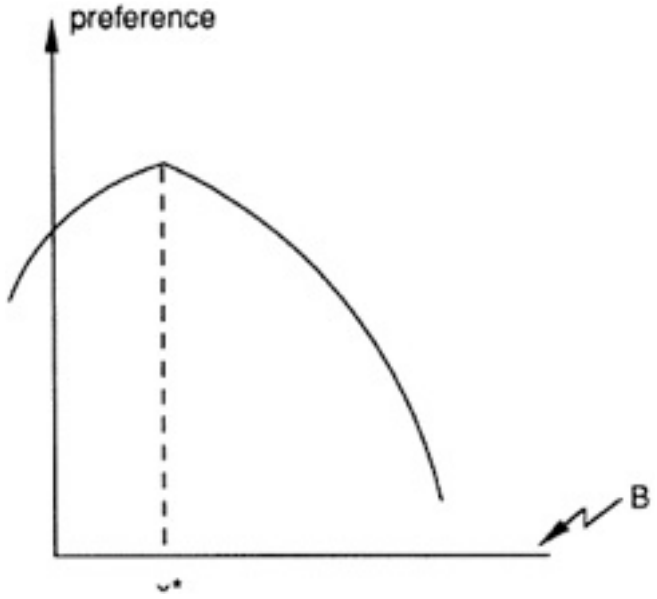

Figure 1.2: A Single Peaked Preference

While Figure 1.2 might offer information about preferences that we need not concern ourselves with when discussing choices in a supermarket when the usual rules apply, they may be critical for determining how a person votes when those store purchases are made using some collective mechanism. Suppose, for example, that X = beer and Y = baby food. Then an unmarried male might have the preferences denoted by the rightmost curve in Figure 1.3 (not setting that person's ideal point at Y = 0 allows for the possibility that he might be curious as to what crushed peas or strained carrots taste like or because he feels some degree of sympathy toward mothers with babies). In contrast, the left-most curve with an ideal point at A might correspond to one of those women with babies who, nevertheless, is willing to allocate a small part of the family budget to beer for her husband. Voter 2 with an ideal point at B, on the other hand, might correspond to a husband who knows he'd be in serious trouble at home were he to return from grocery shopping after spending the majority of the family's budget on beer.

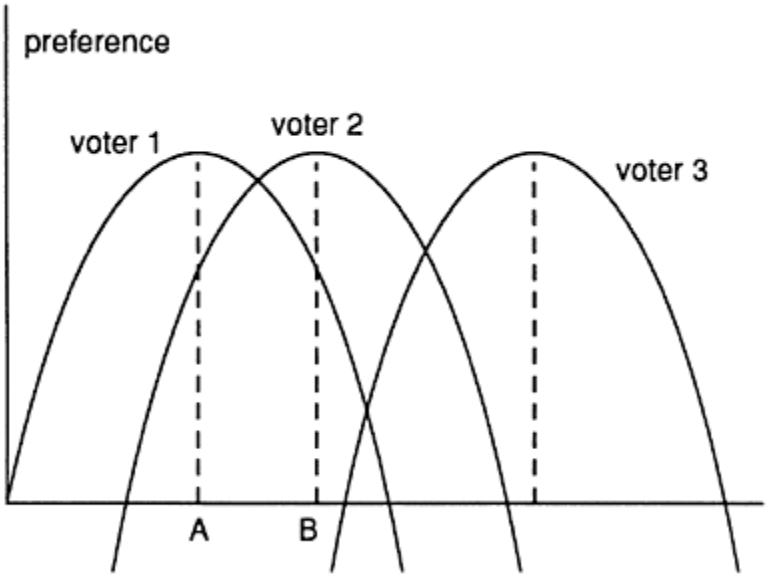

Figure 1.3: Three single peaked preferences for three different "consumers"

We will make a good use of preference curves such as those in Figure 1.3. But before we do so let us consider some extensions of this representation of preferences. Specifically, suppose there is a third commodity, Z, that can be purchased only outside of the supermarket. If we were now to attempt to represent a person's preferences over X, Y and Z simultaneously in a three-dimensional space by way of extending Figure 1.1, we'd

most likely imagine something like a set of nested mixing bowls with their bottoms aimed at the origin of the space, and each smaller or more distant bowl corresponding to a higher level of utility. We refrain from drawing such curves because doing so exceeds our graphic skills. But now imagine a person's budget constraint in this three dimensional space. Rather than a line, that constraint would be a triangle (a *budget simplex*) wherein each vertex of the triangle corresponds to all of the budget being spent on X or Y or Z. Finally, try to imagine what the surface of that triangle might look like as it cuts thru various mixing bowls. Some of those bowls will not, of course, touch the triangle since they represent combinations of the three goods that cannot be achieved, given one's budget. And some of them will inscribe circles or some such curve on the triangle as the triangle cuts thru them, thereby denoting budget-consuming mixes of X, Y and Z over which the person is indifferent. And unless the decision maker in question has preferences that yield a taste for spending their entire budget on only one of the three goods, we will find that one of the bowls just touches (is tangent to) the triangle. That point of tangency, then, corresponds to the person's ideal allocation of his or her budget and, as in Figures 1.2 and 1.3, the further from that point, the less that person will like it. Figure 1.4, then, illustrates these indifference curves on the budget simplex assuming perfectly round bowls after we lay out that simplex flat on the page.

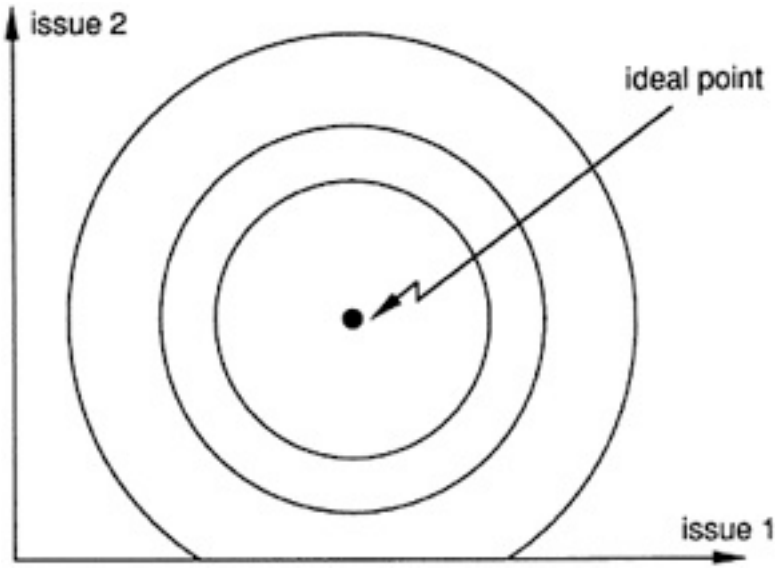

Figure 1.4: Two-dimensional Spatial Indifference Curves

The reader might ask why we've gone thru so much trouble to extend our 2-good model to three goods. Suppose, then, that instead of taking a fixed amount of money from our grocery shoppers when they enter the supermarket, we instead allow them to vote on how much of their budget will be spent on X and Y (and, thereby, how much they can spend subsequently, once released from the store, on Z) – or, more properly, we require that the two candidates take positions on how much will be spent in total on X and Y as well as on the allocation of monies between X and Y. Suppose we also eliminate any reference to beer and baby food and instead substitute such words as "social welfare" and "national defense". And instead of the abstract labeling of the third dimension as good Z, we instead think of it as the negative of a tax rate. Thus, we have arrived at a model – admittedly simple-minded – of an election in which voters must not only choose between different types of public spending, but also on the overall size of the public sector. People with ideal point near the budget simplex's vertex on the Z dimension prefer a small, if not insignificant, state wherein all consumption decisions are left to the private sector; people with ideal points near or at the simplex's vertex on the X dimension prefer massive government spending, provided it is spent on national defense; and people with ideal point at the third vertex prefer that most of society's wealth be devoted to social welfare programs.

Presumably, the majority of us prefer something closer to the middle or at least away from such extremes. For that reason, when making use of such *spatial* representations of preference, we forgo drawing triangles and as we have done in Figure 1.4, simply denote the axes of the coordinate system along with the indifference curves and ideal points within it. The important thing, though, is to understand how we can move from the economist's usual representation of consumer preferences to those of voters who must make decisions using a more collective (political) institutional arrangement.

Before we sign off on this subject, it is useful to consider some of the forms spatial preferences can take. Figure 1.4 represents those preferences with some nondescript concentric circles, which suggests that the voter in question weights the two dimensions

or issues equally. Circular indifference curves or contours are especially useful for illustrating basic ideas, and are useful when we take advantage of our natural intuitions about geometry and distance to explore a new idea so that our intuition can lead our reasoning. However, consider Figure 1.5a, which represents preferences as concentric ellipses. In this instance we can say that whoever holds such preferences is more sensitive to changes on the first (horizontal) dimension than the second. And then there's Figure 1.5b where the elliptical indifference curves are tilted relative to the axes. First, notice that in both Figures 1.4 and 1.5a, a person's preference on one dimension does not depend on what choice is made on the other dimension. So if we arbitrarily fix the value of one dimension, the most preferred value on the second is unchanged (i.e., if, for instance, we draw a horizontal line in either Figure 1.4 or 1.5a, the value of X that corresponds to the tangency of that line to an indifference curve, x*, is invariant with the height of the line). In this case a person's preferences are said to be *separable* and their utility can be expressed as $U(x, y) = f(x) + g(y)$. For the case of Figure 1.5b, in contrast, preference on one dimension depends on the value assumed on the other. For this example, the higher we draw a horizontal line across the figure, the higher is the value of X that corresponds to the tangency of that line to an indifference curve (x** versus x*). Thus, to represent preferences for overall combinations of X and Y we must now write something like $U(x, y) = f(x) + g(y) + h(x, y)$.

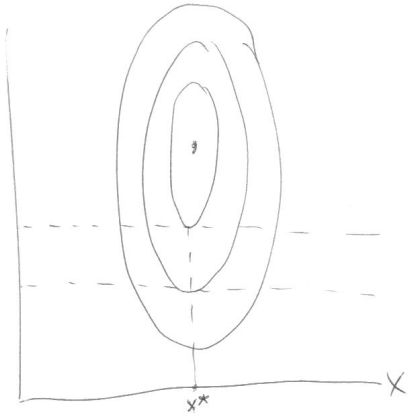

Figure 1.5a

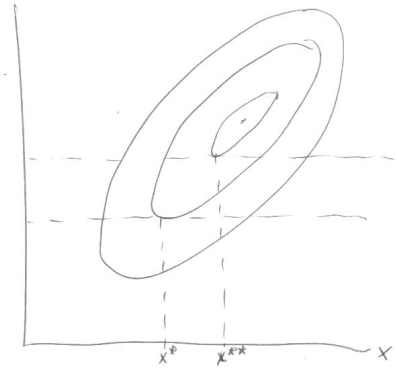

Figure 1.5b

Re-voting at the Philadelphia Convention of 1787: Absent an appreciation of the possibility of non-separable preferences, a naïve reader of James Madison's notes on American's Constitutional convention in 1787 might occasion considerable confusion, or at least leave one with the impression that the delegates there were indeed a confused lot. Specifically, consider the following recorded votes on the character of the presidency:

June 1: agrees to a 7 year term, by a vote of 5-4-1

June 2: agrees to selection of the chief executive by the national legislature, 8 -2 and reaffirms a seven year term, 7-2-1

June 9: defeats selection of the president by state chief executives (governors), 10-1

June 17: agrees once again to selection of president by the national legislature, 10-0, but postpones decision on seven year term

July 19: votes 10-0 to reconsider the executive branch. Passes by a vote of 6-3-1 selection by electors; defeats 8-2 a 1-term term limit, defeats 5-3-2 a seven year term, and passes 9-1 a six year term

July 23: agrees by a vote of 7-3 to again reconsider the executive branch. Passes 7-4 selection of president by national legislature and passes 7-3 a seven year term with a 1-term term limit.

Aug. 24: defeats 9-2 direct election of the president and defeats 6-5 election by electors

Sept. 6: defeats 10-1 a 1-term term limit for the president, and agrees to election of the president by electors via a series of votes refining the electoral college.

It might be true that the delegates were indeed at times indecisive and uncertain as to the best arrangement when dealing with details. The preceding vote history reveals, though, that the delegates were in fact considering three inter-related issues: The method of selecting a president, the president's term of office, and whether there would be a limit to the number of terms. In addition, decisions were being made on a great many other matters between June 1 and September 6, including the design of the national legislature and the powers of the president. There is little reason to suppose that the delegates, which hardly could be said to not have included some of the greatest political thinkers and engineers at this or any other time, deemed these decisions wholly separable. Thus, a decision on one dimension (issue) might reasonably be expect to impact their preferences on others and, thereby, cause them to reconsider prior decisions.

Institutionally induced non-separability: We might be tempted to think that separable versus non-separable preferences are the product of individual taste, but they can also be institutionally induced. Consider the following clause from

Article II, Section 1 of the United States Constitution: "The Electors shall meet in their respective States, and vote by Ballot for two Persons, of whom one at least shall not be an inhabitant of the same State with themselves." The Twelfth Amendment, of course, modified this clause by deleting "for two Persons" and inserted instead "for President and Vice-President" so as to avoid the issues that arose in the election of 1800 when both Aaron Burr and Thomas Jefferson received the same Electoral vote total and it fell to Congress to decide which would be President and which Vice President. For our purposes, though, notice that one's preference for Vice Presidential candidate can be (and generally is) a function of who is nominated for President since the political parties that nominate candidates will quite naturally seek some geographic spread to the two nominees so as to appeal to the electorates' varied sectional interests. The U.S. Constitution sets that preference in stone and dictates that if X is our choice for President, then we cannot choose Y if Y resides in the same state as X. Or, for another example of non-separability induced in part by institutional arrangements, we note that in presidential (as opposed to parliamentary) regimes, while some voters might prefer a unified government in which the same party controls both the executive and legislative branches, there are also those who prefer divided government wherein one party can act as a brake on the actions of the other. One's preference for president, then, might readily depend on who we think will control the legislature and, in the United States at least, whether the same party will control both the Senate and the House.

There are surely other examples of non-separable preferences that are either psychologically or institutionally induced. For example, suppose you must staff a 2-member committee by choosing from a set of 4 candidates, A, B, C and D. It might be that the person you most want to see on the committee is A, because A's preferences most closely match yours. But suppose A as a function of personalities cannot work with either B or C so that any committee that combines A with either of these two people is likely to function poorly or not at all. Thus, if C is chosen first, you most likely would

prefer that either B or D be the second serving member. However, if D is chosen to fill the first seat, your preference is that D be joined by A.

Separable versus non-separable preferences do not, though, exhaust the possibilities we might need to consider when describing forms of individual preference. Consider the following example of preferences we call lexicographic:

> **Diamond Rings and the FDA**: We're not sure how many male readers have had the opportunity to shop for an engagement ring, but those of you who have should immediately understand the following preferences among women (hope we're not being too chauvinistic here). There are several dimensions with which to evaluate such a ring – the size (weight) of the stone, the stone's clarity, its cut and the quality of the setting. But if your experience matches ours, you will quickly learn that preferences here can be *lexicographic* wherein the second, third and other dimensions do not come into play in making choice unless the two top alternatives are equivalent on the first dimension. Specifically, cut, clarity and setting are of little note unless the main stone is "big enough". Indeed, if given a choice between a 2 karat stone of average clarity versus a 1 karat stone of superb clarity, not a few women would choose the first stone. After all, how much can clarity count if people don't first at a distance say "wow"?
>
> For another example, it is often argued that America's Food and Drug Administration is too conservative in its approval of new drugs – that effective drugs are available elsewhere in the world long before they are approved for distribution and sale in the United States. Now consider that there are two basic dimensions with which to evaluate any new drug: Its potential effectiveness in treating some disease versus the risks of its side effects. Ideally, these two dimensions should be balanced against each other with a willingness to assume risk a function of a drug's ostensible effectiveness. But consider the incentives of bureaucrats within the FDA. If they disapprove of a drug that later proves to have few side effects, there are unlikely to be any personal consequences – arguments

can always be made that further study was necessary before a definitive risk assessment could be conclusively offered. Moreover, if they approve a drug that is effective with no risk, they are unlikely to receive any credit since, after all, they have merely "done what's right". On the other hand, if they certify a drug that proves to have negative or even deadly consequences, there's a good chance that those responsible for the approval will have "hell to pay". Thus, FDA's bureaucrats are likely to be risk averse in the extreme with respect to a drug's side effects to the point that only drugs with no apparent side effects whatsoever are approved. If given the opportunity to certify two competing drugs X and Y from two competing pharmaceutical firms, bureaucrats with lexicographic preferences will consider the matter of relative effectiveness ONLY if both offer equally low risk, otherwise they will certify neither or the one with no apparent side effects regardless of its relative effectiveness.

We draw attention to lexicographic preferences not because there are advantages to playing analytically with them. Indeed, the opposite is true, but our examples show that not only can such preferences arise "naturally" for psychological reasons, they can also be institutionally induced and, therefore, are preferences with which we must sometimes deal. Indeed, inducing lexicographic preferences is not the only role institutions can play in determining how to best model preferences in specific circumstances.

A Lesson from Tinseltown: For an example of how the choice of an institution – in this case a voting method – can impact which dimension of preference is most relevant to an individual decision maker's calculus, we note that if an idea is apparent even to those who populate the movie studios of Hollywood – producers, directors, actors and actresses – then the idea must indeed have an element of truth to it. We are reminded then of the ending scene of the movie *1776*, Hollywood's not-altogether historically accurate version of events in Philadelphia at the drafting and signing of the Declaration of Independence. In voting on the Declaration, the delegates abided by the rule of unanimity whereby votes are taken by state and where

a single Nay would send the document down to defeat. In the movie version of events the decision comes down to the Pennsylvania delegation, which, with considerable liberties taken with historical fact, consists of Benjamin Franklin, John Dickerson and Judge James Wilson. Throughout the movie Wilson is portrayed as a weak personality willing to do Dickerson's bidding, who is strongly opposed to declaring independence and prefers instead that further efforts be made at seeking reconciliation with England. Thus, with two votes against one for Pennsylvania and a rule of unanimity in effect for the Congress as a whole, the Declaration seems doomed to defeat. Franklin, however, makes the parliamentary maneuver of calling for a roll call vote of his delegation. With Franklin voting Yea and Dickerson Ney, Wilson becomes pivotal "for or against", in Franklin's words, "American independence." With Wilson wavering, Franklin drives home the point of Wilson's pivotal role by noting that "the map makers of the world are awaiting your decision." If preferences over choices are invariant with context, Franklin's parliamentary maneuver should be of no consequence. But by being made pivotal, the basis of Wilson's decision changes. As Wilson himself states the matter, if the delegates are able to vote anonymously within each state, it would be Pennsylvania that would be credited or blamed for having defeated the Declaration; but under a roll call vote it would be Wilson specifically who did so. As Wilson goes on to explain, if he votes Yea, he will merely be one among many whereas if he votes Ney, he will be remembered as the man who sank American independence. Since his strong preference for anonymity trumps his preference for seeking accommodation with England, Franklin's maneuver changes the basis of Wilson's decision – changes the value (utility) Wilson associates with the alternatives he confronts -- and, thus, the final outcome.

This example is not intended to illustrate a situation in which Wilson's core values changed – that somehow Franklin's strategy changed the judge's preferences over some multidimensional issue space, where those dimensions included perhaps one

that represented America's alternative relationships with England and another his public visibility. But Franklin's parliamentary maneuver -- his switch in voting schemes for Pennsylvania's delegation – did impact the dimensions Wilson deemed relevant to his decision. Only under a voting scheme in which individual ballots are recorded does Wilson's preference for anonymity play a role since only under such a rule are the outcomes "Declaration ratified" and "Declaration not ratified" elaborated to include a specification of how individuals voted. We see here, in fact, yet another door opening to the relevance of game theory – to that of the strategic choice of institutional forms. Hollywood's portrayal of Franklin's genius might not have been historically accurate, but the scene resonates because we know that if that circumstance had in fact arisen, the real Franklin would have understood the strategic possibilities as they were portrayed.

1.4 Collective versus Individual Choice

To this point we have focused exclusively on how to represent the preferences of individual decision makers while admitting that our true concerns are collective or political decisions. What then of collective or group preferences? After all, everyday discourse about politics is laced with statements or assertions that begin with "Society's interests are __," "The electorate prefers __," "The legislature wants __," "The bureaucracy acted __," "The interests of [country X] lie in __" and so on, as if collective preferences are no less real or tangible than individual ones. We are reminded here of Charles de Gaulle's famous comment that France has no friends, only interests. Here, though, we want to end this chapter on a supremely important cautionary note about attributing preferences to collectivities.

The Condorcet Paradox: Suppose three people hold the following preferences:

Person 1: A preferred to B preferred to C

Person 2: C preferred to A preferred to B

Person 3: B preferred to C preferred to A

The question, now, is how to define the *social preference* of these three people as a group. There are, of course, innumerable ways to do so. We could for instance

simply choose one person at random and define his or her preference as the social preference. Absent any bias in our random selection, such a method seems fair since no person is more likely than any other to represent the group. Alternatively, we could assign 2 points to a first place ranking, 1 point for a second place ranking and 0 points for a last place ranking and construct the social preference by adding up the scores of A, B and C. In this case, though, such a method is indeterminate, or at least undiscriminating since each alternative would be awarded a sum of 5 points. Another and seemingly more "democratic" method is to take a majority vote between the alternatives and if X beats Y in a majority vote, then we would say that X is socially preferred to Y or, equivalently, that the group prefers X to Y. The preceding three preferences, though, point to a general problem with this method. Specifically, note that while C beats A in a majority vote, and B beats C, A beats B. Thus, the social preference is *intransitive*!

The grandfather, granddaughter and the horse: Walking through the village accompanying his granddaughter leading the family's horse, the grandfather senses the villager's disapproval of not affording his granddaughter the pleasure of riding on the horse. So up she goes. But soon thereafter there emerges another sense of disquiet among the village: Why is it that such a young girl requires that her elderly grandfather walk while she rides? Not wanting to appear a spoiled ungrateful child, she insists that her grandfather take her place. But nearly immediately the grandfather senses the villager's disapproval of having him alone riding while his sweet granddaughter walks alongside. So up she goes to join him, whereupon the murmurs of disapproval from the village now focus on the horse's burden of having to bear the weight of two people.

The preceding example is but a folky illustration of the more abstract 3-alternative example that precedes it, wherein both illustrate a thing called *the Condorcet Paradox,* named after the 18th century French mathematician who concerned himself with voting systems and finding a fair method for electing members to the French Academy of

Sciences. That our folky example illustrates the same thing as our abstract one can be seen of we suppose that the villagers are of three types:

Type 1: O1 > O2 > O3 > O4

Type 2: O4 > O1 > O2 > O3

Type 3: O3 > O4 > O1 > O2

Where O1 = both ride the horse; O2 = grandfather alone rides the horse; O3 = granddaughter alone rides the horse and O4 = no one rides the horse. In this case, if all three types are represented in the village in approximately equal proportion, the social preference order under majority rule is O1 > O2 > O3 > O4 > O1. The particular paradox here, of course, is that although the individual preferences used to define the social preference in our examples are transitive (and complete), the resulting social preference, at least under simple pair-wise majority rule, is intransitive, in which case we cannot impute a utility function to the group.

Condorcet's Paradox gives rise to any number of important theoretical issues. What, for instance, are the circumstances under which simple majority rule might yield an unambiguous social preference? Are there other ways of applying the idea of majority rule that might avoid the paradox? Do rules other than majority rule also share the property of manufacturing intransitive social preferences out of transitive individual ones? Are there any rules or procedures that guarantee transitive social preferences and if so, what do they look like?

A spatial Example of the Paradox: We cannot answer all such questions in this chapter. Presently, then, the Paradox should be taken simply as a cautionary note – a warning against becoming overly anthropomorphic in our approach to politics by inferring or assigning motives, preferences and the like to collectivities regardless of their identity. The reader, though, should not suppose that the Paradox is a mere curiosity and the product of some artfully created individual preference orders. Rather, it is a feature of group preferences with which we must deal in nearly all applications of game theory to social processes. To illustrate this fact lets us return once again to the spatial preferences and the two-

dimensional form illustrated in Figure 1.3. This time, though, in Figure 1.6 we draw the indifference contours for three people with ideal points at x1, x2 and x3. Now consider the arbitrarily chosen alternative z1, through which we draw the indifference curves of all three persons. Notice that the shaded areas bounded by these indifference contours are all points that are closer to a pair of ideal points than is z1. Alternative z2, for instance, is closer to the ideals of x1 and x2 than is z1. Thus, under majority rule z2 is preferred to z1. On the other hand, now consider alternative z3. As placed, z3 is closer to the ideal points x2 and x3 than is z2. Hence, z3 defeats z2 in a majority vote. But finally, notice that z3 is not in any of the shaded areas corresponding to points that defeat z1. In fact, z1 is closer to the ideals x1 and x3 than is z3. Hence, under majority rule we have the intransitive social order z1 > z3 > z2 > z1.

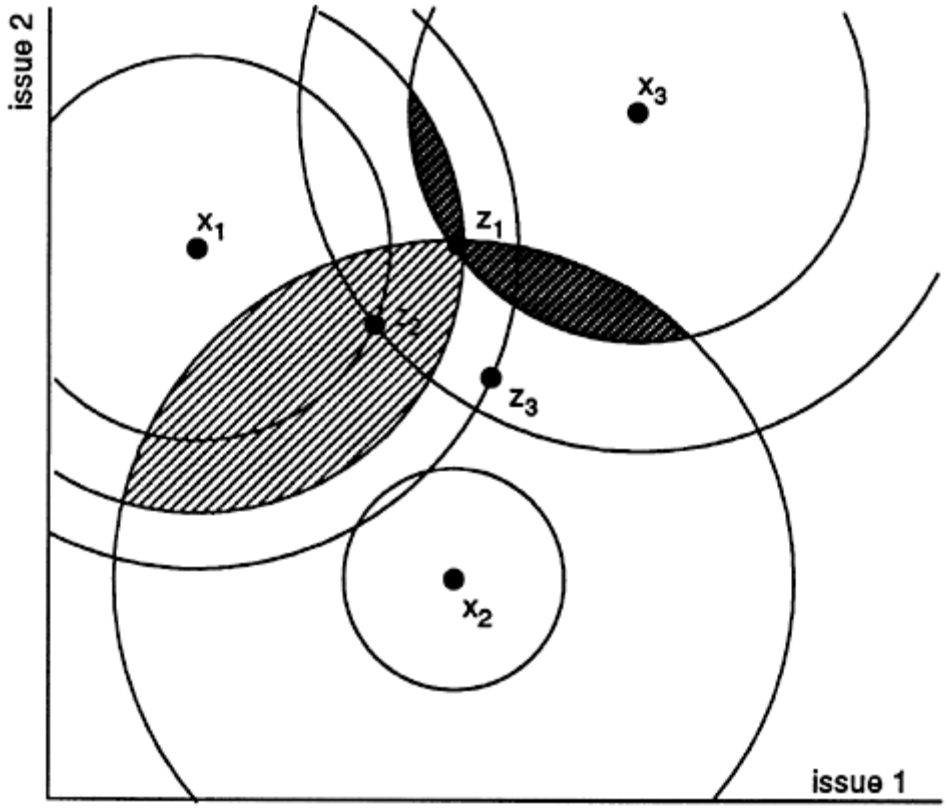

Figure 1.6: Condorcet's Paradox with Spatial Preferences

This simple example – another manifestation of the Condorcet Paradox -- illustrates again the inadvisability of being anthropomorphic about things and, without further analysis, attributing goals, motives or preferences to groups. Barring further developments, we could, of course, assume that the paradox is but an anomalous, albeit unanticipated, characteristic of majority rule. This paradox might merely cause us to question the reverence sometimes associated with outcomes chosen "democratically" by majority rule principles. However, a profoundly important theorem, **Arrow's Impossibility Theorem,** tells us that this paradox is not anomalous or confined to majority rule procedures—it is possible to observe "irrational" social preferences under almost any social process. Arrow's method of demonstrating this fact is to eschew examining each and every rule we might imagine and instead to posit a set of properties or axioms that we think any 'reasonable' rule should follow. Without delving into the finer details of things, the axioms Arrow sets forth are, roughly stated, these:

1. The social ordering is complete and transitive,
2. (Unrestricted domain) No individual preference order over the feasible outcomes is a priori excluded as a possibility,
3. (Pairwise independence of irrelevant alternatives) The social preference between any two alternatives never depends on individual preferences regarding other alternatives
4. (The Pareto principle) alternative x is socially preferred to y whenever everyone prefers x to y
5. (Nondictatorship) No individual should be decisive for every pair of alternatives.

With respect to axiom #2, for instance, there commonly are preferences we prefer to exclude from consideration when making social decisions, such as prohibitions against anti-Semitic or pro-Nazi ideas. Society's current over-indulgence with political correctness is yet another attempt to exclude various preferences from consideration in public discourse. But if we are to fully understand and fully model social processes, then

the method whereby certain preferences are excluded ought to be a part of the general rule we are considering. Axiom 3 requires that we can infer the standing of some alternative relative to another by merely looking at individual preferences over those two alternatives, while axiom 4 in effect requires, among other things, that whatever stands highest in the social preference order be Pareto optimal for the collectivity in question. Despite the reasonableness of those axioms, Arrow's theorem establishes that they are inconsistent. Stated differently,

> *For decisions involving three or more alternatives with three or more individuals, at least one of the axioms 1-5 must be violated. Any procedure consistent with Axioms 2-5 will allow for intransitive social orderings. Equivalently, the only procedure consistent with Axioms 1 – 4 must violate Axiom 5.*

Later we will grapple with the full consequences of this theorem, which is one of the most important in political theory. However, here we want to emphasize the special role played by the transitivity and completeness assumptions in modeling people and the fact that these assumptions cannot play an equivalent role in our discussions of groups. Although it is often convenient to be anthropomorphic and to attribute motives to groups in the same way that we attribute motives to individuals, as our earlier brief discussion of Britain's policies prior to the outbreak of WWI reveals, and as Arrow's theorem precisely formalizes, such linguistic shortcuts are simply that -- shortcuts – and not scientifically valid. Thus, although we may choose to use such shortcuts to convey general meaning and although they may be approximately valid when individual preferences are unanimous or at least nearly so, we must keep in mind that any theoretically valid explanation for social processes and outcomes must rest ultimately on an assessment of the preferences and actions of individuals in combination with the institutions (broadly interpreted) within which those individuals and preferences operate.

Arrow's theorem and the associated Condorcet Paradox, though, are not the only problems associated with attributing goals to collectivities. Another way to bring that fact home graphically is to turn to an example wherein a group appears to act utterly

irrationally (i.e., in its own worst interest) but where it seems reasonable to assume that the individuals within the group are acting in pursuit of perhaps the most intense of all preferences, that of survival.

> **The Curious Behavior of Herring:** Some time ago the Discovery Channel released a video of a large school of herring swimming casually in beautiful clear water when suddenly it is attacked by a number of blue fin tuna. It seems that herring are deemed quite a delicacy by tuna, at least judging by their enthusiasm for catching and eating as many as possible. The reaction of the herring, though, is surprising. Instead of scattering in every conceivable direction (and indeed there were far more potential directions than there were tuna), the entire school began to swim in a tight swirling ball. The ball, of course, provided a far more inviting target than some widely dispersed cloud of fish, so slowly (or actually, not slowly enough from the perspective of the herring), the ball began to shrink. The ball itself, moreover, began to appear wholly disoriented and slowly moved toward the surface, which only made it vulnerable to the swooping pelicans above. The video ends when the ball barely exists and the tuna are fully sated.

We are admittedly not in a position to fully account for this odd and seemingly self-destructive behavior on the part of the herring, aside from noting that it appears to contradict Darwin's rules about species survival. In fact, we will later refer to this example as a way to illustrate problems of collective action and social coordination. Here, though, we can use it to illustrate the distinction between individual and collective preference. Were we, for example, to try to explain the behavior of herring with reference to collective preferences and actions, the conclusion that schools of herring prefer suicide and extinction seems inescapable. Surely, since every herring scattering to the wind (or, more properly, ocean current) seems the best choice for all collectively, the school's behavior is consistent with the hypothesis of a group preference for suicide. It seems safe to assume, on the other hand, that while any one herring could give a twit about the concepts of extinction and group survival, each individual fish does value its own skin (or scales) and if it could, would act accordingly. A careful look at the swirling

ball confirms this supposition – specifically, the ball swirls because each fish is doing its damndest to get into its interior. Each herring has but two choices: To swim unilaterally away from the ball and, most likely, into the waiting mouth of one of the surrounding tuna, or to try to disappear inside the ball in the hope the tuna will satisfy themselves by eating only those on the perimeter. Alas, their choice is a Hobbesian one – there is no good choice for each herring, though getting into the interior of the swirling ball would seem the best of two distinctly poor alternatives (since there are a few herring left when the video ends). There is, of course, no issue of collective intransitivity here as in Condorcet's Paradox. After all, there are only two choices for the school – form the swirling ball or scatter and run (swim) like hell. But now we have an example where the presumed unanimous preference of individuals – to survive – is transformed by circumstances into the seemingly irrational preference (if we are to judge by the "school's choice") of maximizing the ease with which it can be eaten.

More will be said of this example later, but note we have just used a word that itself warrants comment; namely, *irrational*. Much has been made of this word and its opposite, *rational,* and much has been said about whether these words have any proper definition in the context of contemporary social science theory. Some will assert that any behavior is rational if it can be conceptualized to follow from some well-defined set of preferences over ultimate outcomes. Rational action or rational behavior is simply any behavior that can be said to follow from our theories and postulates about preference. This definition, though, would seem to render the concept a worthless tautology since we can always ascribe goals to whatever it is we observe. Even the school of herring, for example, can be deemed collectively "rational" if we are willing to postulate for it the goal of suicide. Others then ascribe rationality only to those actions that can be justified by some "reasonable" set of goals or utility functions. But this merely pushes the pebble of contention to a different place in the mud puddle since we are then left with having to define "reasonable." There is no bar of metal sitting alongside the one in Paris that standardizes the measurement of 'meter' with which we can measure or distinguish between reasonable and unreasonable. Our preference here, then, is to banish the words rational and irrational altogether from our lexicon and to instead simply proceed to the

task of seeing if we can explain (and predict) social and collective actions with concepts that do not require such words. That is the task to which we now turn.

1.5: Key ideas and concepts

decision theoretic
game theoretic
common knowledge
preference
complete preferences
transitive vs intransitive preferences
risk
utility
indifference curves
expected utility
nondictatorship
unrestricted domain
time discounting
rational
tautological
subjective probability
budget constraint
budget simplex
spatial preference
single peaked preference
separable vs non-separable preferences
social preference
collective action
Condorcet paradox
Condorcet winner