

National Security Strategy: Games and Information

Professor Branislav L. Slantchev

January 1, 2014

Overview Because we want to study strategic coercion, we need a way to analyze strategic interaction. We learn how to use game trees to represent stylized descriptions of various strategic situations. We learn to distinguish between complete and perfect information in order to specify what players know when they get to act. We further learn how to transform situations of incomplete information (which we cannot represent directly) into games of imperfect information (which we can). I also provide a brief history of game theory for fun.

Last time we learned that we shall be studying *strategic coercion*. That is, how nations can use force or threats to use force in pursuit of their political objectives. In other words, how nations can coerce others to comply with their wishes. National security concerns itself with not only gaining some worthy objectives, but denying opponents the opportunity to gain theirs. Because the outcomes of these interactions are jointly determined by the actions taken by everyone involved, we need to study *strategy*. That is, how does one choose an optimal action (an action that would lead to a most preferred outcome) given that his opponent is also trying to optimize his action (an action that would lead to his most preferred outcome).

We now focus on **strategic interaction**, which is the study of such interdependent decision-making. Generally, we shall find that getting the most preferred outcome is surprisingly difficult and in many cases an optimal strategy does the best possible with the worst available.

Studying strategic interaction is difficult because one's optimal course of action depends on what the opponent is doing and vice versa. That is, we cannot analyze our strategy in isolation from the strategy of our opponent. But the strategy of our opponent in turn depends on what he thinks our strategy is going to be. Thus, we have to analyze *both* strategies simultaneously, much like you would solve a system of equations.

1 A Brief History of Game Theory

The tool for this purpose is called **game theory**, which, despite its frivolous name, is a fairly serious branch of applied mathematics. Game theory allows us to analyze complex interdependent situations. The development of this theory began in 1944 when the genius mathematician John von Neumann wrote a book with the economist Oscar Morgenstern. The book was called *Theory of Games and Economic Behavior*, and it laid the cornerstone of this discipline, which now pervades economics, is widespread in political science and theoretical biology, and is making steady inroads into sociology and anthropology.

What von Neumann and Morgenstern showed is especially useful. They showed that given several reasonable assumptions about a player's rationality, it is possible to represent his preferences over risky choices with numbers (payoffs or *utilities*), and that choosing the best according to these preferences is equivalent to choosing the one that yields the highest expected utility (payoff). Expected utility is the number one obtains when one takes all possible outcomes of a choice, multiplies the payoff associated with each outcome by the probability of that outcome occurring, and then summing over these products. (We shall see numerous examples of this later, so don't worry if it's too abstract now.) In other words, we can use standard mathematical techniques to find the choice that "maximizes expected utility." Rational players should make choices that maximize their expected utilities.

It is important to avoid a very common confusion about what this means. The numbers attached to the various choices are fictitious, they do not represent monetary payoffs or some level of internal happiness (utility in the sense of the Utilitarians). Choosing an action because it maximizes the player's utility does *not* mean that the player prefers that action to everything else because it yields the highest expected utility. Instead, it yields the highest expected utility because the player prefers it to everything else. Preferences are fundamental, expected utilities are just convenient representations of these preferences.

Given this representation, one can then attempt to *solve* a game. Solving a game means finding optimal (best) strategies for each of the players. Von Neumann and Morgenstern devised a method for solving a class of games called "strictly competitive" or *zero sum*. These are games in which a gain for one of the players is an automatic loss for the other player. Although interesting, this class of games is of limited usefulness because most interactions are not zero-sum. Game theory would have remained a tiny branch of mathematics if it were not for another mathematician.

Game theory, however, got an incredible boost in 1950 and 1951 when John Nash published two short papers, in which he proved the existence of optimal solutions to a large class of games, both zero sum and non-zero sum. This solution is what we now call *Nash equilibrium*. The idea is elegant. Suppose we are given some strategic situation to analyze. The description includes things like: (a) who the players are, (b) what their possible strategies are, (c) what the possible outcomes that can be reached through these strategies are, and (d) what each player's utility for each outcome is. Given this description, we want to know what actions the players are going to take.

Nash equilibrium is a set of strategies, one for each player, such that no player can obtain a better payoff by unilaterally changing his strategy. In other words, each player's strategy is a best response to the strategies of the other players. This means that your optimal strategy is one such that given what your opponent is doing, you have no incentive to switch to another strategy. The one you are using is giving you the best possible outcome. If every player is using an optimal strategy, then the set of strategies is called Nash equilibrium. We shall use this concept, along with some new modifications, to analyze various strategic situations. To solve a game means to find its equilibrium (or *equilibria* if there are multiple solutions).

You may have noticed that the description above assumed that every aspect of the game was common knowledge to all the players. But in reality we are often uncertain about things. We may not know, for example, what utility another player might be getting from a particular outcome. If we don't know that, we cannot form expectations about what this player's optimal strategy would look like, and therefore we won't be able to determine our own optimal strategy.

So game theory ground to a halt until the U.S. government financed a series of studies that culminated with John Harsanyi's publication of three papers in 1967-

68, in which he described how one could analyze games of incomplete information. The application of game theory then exploded.

It is worth repeating that the impetus from this development came from researchers working for the U.S. government during the Cold War when they were commissioned to find ways of analyzing the behavior of the Soviets. We shall be following in the footsteps of these people.

2 The Modeling Enterprise

As I mentioned before, strategic thinking can be quite complex, and the situations that we try to analyze can involve a lot of factors that additionally complicate our efforts to understand what our best options are. In other words, we, as all the students of international relations before us, have to find a way to reduce this complexity and make it manageable. This means that as analysts, we shall **simplify reality** by ignoring aspects of it that we believe are unimportant or too difficult to deal with from the outset. Our **model** will be a theoretical construct designed to capture the essence of the situation we want to analyze. Obviously, each model is only as good as the person constructing it: If that person misses something especially relevant, then the model will fail to capture an important aspect of reality, and its analysis will be misleading, at best.

But how do we decide what an “important aspect of reality” is? What makes one aspect more important than another? The notion of importance is tied to the notion of usefulness in the following sense. We construct models with specific purposes in mind. For instance, we may be interested in how rational actors can end up fighting a war with each other. To analyze a crisis between two states, we need to simplify reality somehow. But what can we discard and what must we put in our model? What assumptions can we make without the danger of distorting our results?

One common simplification of reality is to treat states as *unitary* actors: that is, treat them as if they are individuals instead of the complex organizations that they are. Surely this is a drastic simplification but is it distorting, does it produce answers that are so far from reality that they are meaningless and useless for our task at hand?

This depends on the task: If our goal is to identify general conditions that make war permissible, then this assumption may not be too bad. For instance, if we assume that the leadership of any state has an overriding concern to ensure that the state survives and that the only way to ensure survival is by procuring enough military power (either by producing it internally or by obtaining the protection of powerful allies), then we can say that the international system compels state leaders to behave very similarly in general regardless of how their states are organized. The constraints on behavior imposed by the anarchic structure of the international system provide motivations that must be taken into account by any leader.

If, on the other hand, our task is to explain particular foreign policy choices; that is, how leaders respond to these constraints in any given circumstances, then clearly the unitary actor assumption might be severely distorting. To answer questions like: how should the U.S. get Iran not to produce nuclear weapons must take into account, at the very least, the organization of Iran's state apparatus and decision-making structures.

Simplification for purposes of explaining is akin to making a map of a place. What features you choose to retain and what features you choose to discard depends on the purpose of the map. Suppose you need to know how to get from Grand Central Station to the Metropolitan Museum in Manhattan by means of public transport. A physical map with a detail representation of terrain elevations will not be very useful for that purpose. A map listing tourist attractions will not be useful either unless it includes bus and metro routes. What you want is a map of the public transit system, with the stations clearly marked, and with some of the most important landmarks identified (in case the station name closest to a landmark is different). This map is an extremely simplified view of Manhattan, yet for your purpose it is quite sufficient. Discarding topographical information is inconsequential.

In analyzing international politics, unfortunately, things are a bit more difficult. There is always a lot of controversy about what a useful map is supposed to look like. Hence, there is always a lot of disagreement about the most appropriate ways of simplifying reality. To deal with this problem, we will engage in the **modeling enterprise**: that is, the construction of a sequence of models, each a better approximation of reality than its predecessor. The idea is that we shall begin with very simple models (that we know to ignore many important things), and analyze them to obtain some insights about what we should expect rational actors to do. We shall then engage in a dialogue with empirics (historical cases, statistical analyses) to figure out which of the missing elements seems most crucial. We then go back and revise the model to account for that, and analyze it. We can then see whether the original assumption was distorting.

It may turn out that we are getting essentially analogous results from the more complex model. In that case, the original assumption was not misleading and we can keep it to retain simplicity and parsimony. We also know that the conclusion is more robust because it is not very sensitive to changing this particular assumption. This would be a very good thing to find out and it should increase our confidence about the usefulness of the insights produced by the model.

Alternatively, it may turn out that our conclusions change drastically: that is, they are very sensitive to that assumption. We can then turn back to empirics to judge which of the two conclusions provides a better match empirically; that is, which model has a better fit with the data available. This empirical criterion may serve to select one of the models as the more promising venue for future research. Or we may keep both and then relax another assumption to see how robust the predictions

are. (The problem is that sometimes assumptions interact in unforeseen ways and it may be the case that relaxing one of them produces a model worse than the original but relaxing two produces an improvement. If we discarded the second model on empirical grounds after relaxing one of the assumptions, we would not get to the third version which outperforms empirically the original.)

With this interaction between theory and empirical/historical data, we eventually produce models that are adequate representations of real strategic situations for the purposes at hand. The upshot of all this is that one should not judge one model in isolation from this enterprise. Each model will have some important shortcomings, but each is just a step toward our ultimate goal.

We shall have to make two important assumptions in order to proceed with the modeling enterprise. First, we shall assume that actors are rational. Second, we shall assume that they are intelligent.

3 Representing Preferences: Rationality and Expected Utility

The rationality assumption involves less than the everyday usage of that word implies. All we want here is that actors have preferences over various possible outcomes, and they pursue actions that are consistent with these preferences. That is, they choose actions that are most likely to yield outcomes they like better. This does not mean that actors will not make mistakes, and the rationality assumption does not exclude the possibility that they do. What it does exclude are irrational (crazy) actors whose behavior has no relation to their goals. This is not to say that there are no crazy people. But it is to say that if we assume our actors do not make choices that have something to do with their preferences, then we have to abandon all hope of comprehending such behavior. It will appear to be completely arbitrary to us, and therefore unintelligible.

By rational preferences, we mean that the preferences are not logically contradictory. For example, suppose an actor expressed a preference for peace over war and war over unconditional surrender. Then it must be the case that this actor prefers peace over unconditional surrender as well (that is, the preference ordering is transitive). If he instead expressed a preference for surrender over peace, we would deem such an actor irrational. The second requirement is that the actors cannot refuse to rank some outcome—that is, their preference ordering must be complete. When preferences satisfy these two requirements, we say that they are *rational*.

It is important to emphasize that this definition of rationality has nothing whatsoever to do with the actual content of these preferences. In particular, it embodies absolutely no normative judgment about their morality or ethics. Thus, the rationality assumption does not exclude people like Hitler or Saddam Hussein, who many believe to be “irrational.” Disagreeing with an actor’s preferences is not a basis to declare that actor crazy. An actor may have preferences we find irredeemably odi-

ous, for example, Hitler held the preference for exterminating all the Jews to letting them live. This may be pathological, but since he did pursue a policy designed to accomplish precisely that, he has to be judged rational in our sense: his actions reflected his preferences. By the same token, a suicide bomber has preferences that appear “irrational” to most of us: after all, we’re not taken to blowing ourselves up in order to kill some civilians. However, from our formal perspective, a preference for killing oneself in such a way can still be rational as long as it does not violate the usual logical ordering.

These preferences are *ordinal* because they only require actors to rank the outcomes. They do not require actors to specify intensity. With an ordinal ranking a question like “Just how much worse is war compared to peace for this actor?” is meaningless. Furthermore, these preferences do not take into account the fact that actors may often be uncertain about the consequences of making particular choices, and that their behavior will then depend on the subjective evaluation of risks.

How do we represent preferences for choices under uncertainty? As I mentioned before, von Neumann and Morgenstern’s contribution is to show that we can assign numbers to choices such that the choices that the actor prefers to others also yield higher expected utilities. To represent preferences in this way, we introduce the concept of a lottery, which is defined as follows. For each possible outcome, the lottery specifies the probability which which this outcome will occur if this lottery is “played.” The probabilities must sum up to one; that is, if the lottery is played, some outcome must occur. A risky choice then is essentially the choice which lottery to “play.”

For instance, suppose there are two possible actions in a crisis: attack or stand down. Actor *A* believes that if he attacks, the probability that the opponent *B* will fight is $1/3$, and if he does attack, the probability that the opponent will attack anyway is $1/4$. There are four possible outcomes:

1. If neither player attacks, peace prevails; the outcome is the status quo (SQ).
2. If *A* attacks but *B* chooses not to fight; the outcome is capitulation by *B* (Cap_B).
3. If *A* does not attack but *B* does; the outcome is capitulation by *A* (Cap_A).
4. If *A* attacks and *B* chooses to fight; the outcome is war (War).

From *A*’s perspective, the consequences of his choices are uncertain. If he attacks, the result could be either war (with probability $1/3$) or capitulation by *B* (with probability $1 - 1/3 = 2/3$). If he does not attack, the result could be either capitulation by himself (with probability $1/4$) or the status quo (with probability $1 - 1/4 = 3/4$). The risky choice to attack is a lottery, L_1 , which assigns the probability $1/3$ to War, $2/3$ to Cap_B , and zero to the other two outcomes. Analogously, the risky choice not to attack is another lottery, L_2 , which assigns the probability $1/4$ to Cap_A , $3/4$ to SQ,

and zero to the other two outcomes. Actor A 's choice depends on his preferences over these two lotteries, not just one the preferences over the outcomes, because he must take into account the uncertainty (risk) involved in the different choices. The fundamental preferences here are over lotteries rather than outcomes because the attitude toward risk as an intrinsic property of the individual which cannot be derived from other primitives.

We will maintain our basic rationality assumptions: preferences over lotteries are transitive and complete. In addition, we have to make two new assumptions: *independence* and *continuity*. Independence means that an actor's preference over two lotteries is not going to change if they are combined in the same way with a third lottery. For instance, suppose that B is currently conducting exercises for his armed forces and there is a 1% chance that they will result in an accidental triggering of an unprovoked attack on A . For simplicity, let's assume that this will certainly cause A to retaliate, so the outcome would be war. The maneuvers comprise a third lottery, L_3 , which assign probability $1/100$ to war, and $99/100$ to SQ.

Observe now that this third lottery modifies the existing two as follows. Not attacking could result in (accidental) war with 1% chance and if it does not (which happens with 99% chance), it could result in the outcomes as per L_2 . The new lottery is L'_2 , which assigns probabilities 0.01 to War, 0.7425 to SQ, and 0.2475 to Cap_A .¹ Analogously, attacking produces a new lottery L'_1 , which assigns probabilities 0.34 to War, and 0.66 to Cap_B .² Independence requires that *if* A preferred L_1 to L_2 , then he should also prefer L'_1 to L'_2 . Conversely, *if* A preferred L_2 to L_1 , then he should also prefer L'_2 to L'_1 . Intuitively, since the likelihood of accidental war is the same whether or not A escalates, the preference over the risky choices involving accidental war is entirely determined by his preference over the choices without accidental war.

The second assumption is that of *continuity*. Intuitively, it means that if we take outcomes the player judges to be best and worst, then for any intermediate outcome there exists a probability p such that the player will be indifferent between getting the intermediate outcome with certainty and a lottery that yields the best outcome with probability p and the worst outcome with probability $1 - p$. Let me give a contrived example. Suppose there are three possible outcomes: I get \$500, I get \$0, and I die. The claim is that there exists a probability p such that I would be indifferent between getting \$0 with certainty and a lottery in which I get \$500

¹To obtain the probabilities of the latter two outcomes, note that for SQ to occur, it must be the case that accidental war does not occur (99% chance) *and* B does not attack deliberately (75% chance), or $0.99 \times 0.75 = 0.7425$. For Cap_A to occur, accidental war must not occur *and* B must attack deliberately: $0.99 \times 0.25 = 0.2475$. You should verify that the probabilities of L'_2 add up to 1 so that *some* outcome is certain to occur: $0.01 + 0.7425 + 0.2475 = 1$, as required.

²War can occur either by accident, with 1% chance, or when accidental war does not happen and B fights: $0.99 \times 1/3 = 0.33$. Hence, the overall probability of war is $0.01 + 0.33 = 0.34$. Also, for Cap_B to occur, accidental war must not occur, and B must capitulate: $0.99 \times 2/3 = 0.66$. To verify all of this, observe that the probabilities sum up to one, as required.

million with probability p and get killed with probability $1 - p$. This magic number p is clearly subjective and depends on the degree to which I am willing to run risks of dying. For instance, suppose my salary is \$500 per day, and that I know that the probability of dying in a car accident while driving to work is 1 in 50,000 on any particular day. On the other hand, I could stay home, fail to die in a car crash, and earn nothing. Obviously, I am willing to run some risk to get to work on any given day in order to earn \$500 for that day. The lottery from going to work kills me with probability $1/50000$ and gives me \$500 with probability $49999/50000$. I know (by revealed preference) that I prefer this lottery to one that gives me \$0 with certainty.

Suppose now that we increased the risk of dying to 2 in 50,000. In the new lottery, I get \$500 with probability $49998/50000$ and get killed with complementary probability. I am now offered this new lottery and the certain \$0. It so happens that I prefer the lottery. We now continue gradually making the lottery riskier and riskier until we reach some probability p at which I declare that I am indifferent between the lottery and the certain \$0. This probability must exist because I know that if the risk of dying is too high, I will prefer the certain \$0. For instance, it might be the case that I would not prefer the lottery if the probability of dying exceeds 1 in 5,000. In this case, $p = 4999/5000$ makes me indifferent between the lottery involving the best and worst outcomes, and the certain intermediate outcome.

Given these assumptions, we can represent the best outcome with 1, the worst outcome with 0, and the intermediate outcome with the number p we just found. If we have two intermediate choices x and y , we can find two numbers p_x and p_y such that the actor is indifferent between outcome x with certainty and a lottery that yields the best outcome with probability p_x and the worst outcome with probability $1 - p_x$, and the actor is also indifferent between outcome y with certainty and a lottery that yields the best outcome with probability p_y and the worst outcome with probability $1 - p_y$. Observe now that if the actor prefers x to y , it must be the case that $p_x > p_y$. For instance, if I am offered \$200 with certainty, I will not be as willing to risk a car accident to get \$500 as I am when the alternative is \$0. Hence, $p_{200} > p_0$: to get me out on the freeway, the lure of ending up with \$500 must be stronger when my safe alternative is \$200 than when my safe alternative is merely \$0. To put it another way, if my safe alternative is only \$0, I would be willing to run higher risks to get \$500 than if I am guaranteed \$200 already. Intuitively, among lotteries that involve only the best and the worst outcome I always prefer those that yield the best outcome with higher probability. When I prefer alternative x to alternative y , the probability p_x must be higher than p_y as well. This correspondence allows us to use these numbers to represent the preferences: the number p_x is the utility of outcome x , and the number p_y is the utility of outcome y . In my case, the utility of getting \$500 for sure is 1, the utility of dying is 0, and the utility of \$0 for sure is $4999/5000$.³

³In fact, we can find equivalent representations by transforming these utilities as long as we keep

Von Neumann and Morgenstern then proved that if we assume rationality, independence, and continuity, then we can represent preferences over the various risky alternatives in the *expected utility* form. That is, for each lottery (risky alternative), we can compute the expected utility of that lottery as follows: take each possible outcome and multiply the utility we just assigned to this outcome by the probability with which this outcome occurs in the lottery, then sum over all these products. For instance, the expected utility of the lottery “driving to work” is $EU(\text{drive}) = (1) \times (49999/50000) + (0) \times (1/50000) = 49999/50000$. The expected utility of staying home is $EU(\text{stay}) = (4999/5000) \times (1) = 4999/5000 = 49990/50000$ because I get this outcome with certainty. Recall that I claimed to prefer driving to staying and note that $EU(\text{drive}) > EU(\text{stay})$, just as required. In other words, the expected utilities from the two lotteries maintain the ordering of my preferences over these lotteries. On the other hand, if driving gets too dangerous, I would stay home. I said before that I would prefer to stay home if the risk of dying is above $1/5000$, so suppose it is $3/5000$. If expected utilities represent the lotteries, it must be the case that the expected utility from driving under these conditions is worse than the expected utility of staying. With the new risk, $EU(\text{drive}) = (4997/5000) \times (1) + (3/5000) \times (0) = 4997/5000$, which is clearly smaller than $EU(\text{stay}) = 4999/5000$ as found before. Again, the expected utility form represents my preferences.

It is crucial to observe two things. First, the utilities are thoroughly subjective and depend on my innate propensity to run risks. Second, as the example above shows, when I choose to drive under the original conditions, it is *not* because the risky option gives me a higher expected utility. Instead, it gives me a higher expected utility because I prefer it to the safe option. Choosing the option that yields the highest expected utility is equivalent to saying that I am choosing the option that I prefer to every other option, i.e., what a rational actor should be doing. Hence, when actors maximize utility, they are not actually striving to achieve the maximum amount of internal happiness or money. It is just a convenient short-hand for saying that they are pursuing the best possible course of action under uncertainty.

The second assumption is that actors are intelligent, which is to say, they understand the situation at least as good as the analyst. In particular, this will mean that they see the situation in the same way we see it, and hence our analysis of it should make their actions intelligible to us. This is not as heroic an assumption as it may seem. Again, there is (regrettably) no shortage of stupid people, of people who regularly fail to understand even the simplest things, people who are very likely to misperceive the situation. Still, we should not assume that actors are systematically dumb for otherwise (again) we have no basis on which to form our expectations about their behavior, and no hope for explaining any social situation. I think it’s a pretty safe assumption to think that most people are quite intelligent most of the time.

the transformations linear.

With all these things in mind, we want to learn how to analyze situations involving rational and intelligent actors in the sense just described. Remember that we will engage in a modeling enterprise, and so some of our first models will seem like toys that hardly capture any essence about any situation. However, we shall also move very quickly to rectify that by incorporating progressively more and more interesting aspects of the situation we want to analyze. Before we move on to the analysis, we have to learn how to simplify reality.

4 Representing Strategic Situations: The Game Tree

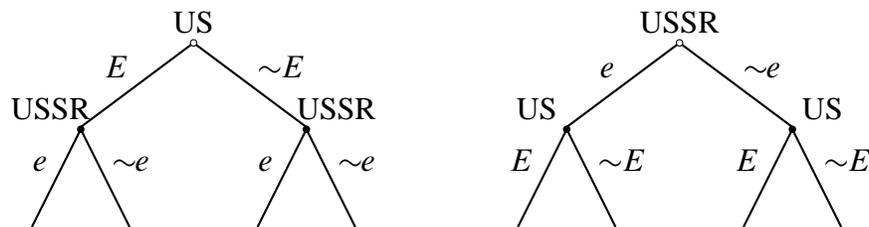
The first step in analyzing a strategic situation is to describe it. That is, we want to know the following things about it:

1. Who are the players involved in this game?
2. What are the actions these players can take?
3. What is the sequence of play? That is, when do players get to act?
4. What are the outcomes? How do the actions jointly determine them?
5. What are the payoffs players receive from each outcome?
6. What do players know about each other?

One easy way to describe a simple situation is with a *game in extensive form*, or a **game tree**. Suppose we want to describe an international crisis between two states, say the US and the USSR. Thus, we have two players. To keep it simple, suppose that in this crisis, each player can escalate the situation or back down. That is, each player has two possible strategies: “escalate” (denoted by E for the US and e for the USSR) or “back down” (denoted by $\sim E$ for the US and $\sim e$ for the USSR). We shall use upper and lowercase letters to help keep track of the identity of the player when we specify the action.

For the order of play, we have five possibilities: (1) US acts first (choosing either E or $\sim E$), and the USSR, after observing this action, chooses next; (2) the USSR acts first and the US acts second after observing its action; (3) the US acts first and the USSR acts second *without* knowing what the US did; (4) the USSR acts first and the US acts second *without* knowing what the USSR did; and (5) the US and the USSR act simultaneously.

After a little bit of thought, it should be obvious that possibilities (3), (4), and (5) represent the same situation. Namely, they describe cases where each player implements his strategy *without knowing what his opponent's action is*. It does not matter who goes first if players do not observe the opponent's action. Therefore,



(a) US Moves First, USSR Moves Second (b) USSR Moves First, US Moves Second

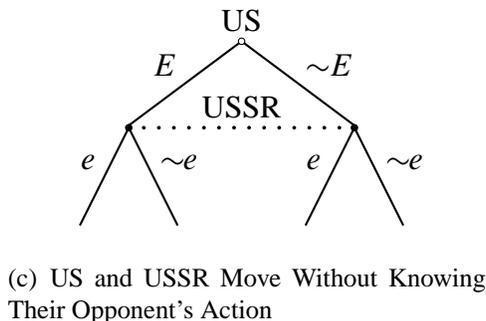


Figure 1: The Three Possible Sequences of the Crisis Game.

we really have three cases to consider for sequencing. So let's use game trees to represent them, as in Figure 1.

Note that in the last of three game trees we could easily exchange the player labels without changing the description of the situation. There is no first or second mover in both cases. The dotted line is called an **information set** and it represents what the player knows when it is his turn to move. In this case, the USSR knows that it has to move but does not know at which node in the information set it is because it does not know whether the US has played E or $\sim E$. (Note that when the US moves it also does not know what the USSR has done.)

The information sets in the other games are all *singletons* because they all contain exactly one node. This means that when a player is about to move, it knows what the other player has done. For example, in (a), the USSR has two information sets: one follows action E by the US, and the other after action $\sim E$. Similarly, in (b), the US has two information sets. In (c), the USSR has only one information set, which contains both nodes.

When the game has information sets that are not singletons, we say that it is a game of **imperfect information**. That is, players do not know perfectly what other players have done. If all information sets are singletons, then it is a game of perfect information. In our example, (a) and (b) are games of perfect information, and (c) is a game of imperfect information.

The game trees describe the players, their actions, the outcomes, and what they know about the sequence of play. All three games have four possible outcomes: (1) both players escalate, (2) US escalates and USSR backs down, (3) USSR esca-

lates and US backs down, and (4) both back down. The see each outcome, simply trace the sequence of actions starting from the top of the game tree (*initial node*) and going to the bottom. The four nodes at the bottom are called **terminal nodes** because they denote the termination of the game. Once the game ends (that is, once a terminal node has been reached), we say that the corresponding *outcome has been realized*.

We now have to determine how players rank the possible **outcomes** of the game (which correspond to the terminal nodes). Think about the situation we are trying to represent.

It is an international crisis. If both countries escalate, war would erupt. This outcome, labeled “war,” is the worst possible for both players. If one country escalates and the other backs down, then the country that stands firm gains in prestige and from the resolution of the crisis in its favor, while the country that submits loses in both. From the perspective of the country that stood firm, the outcome is “victory” and from the perspective of the country that submitted, the outcome is “capitulation”. Victory is better than capitulation and capitulation is better than war (because it avoids the enormous costs of fighting). If neither player escalates, then the outcome is the “status quo” for each. The status quo is better than capitulation and war but worse than victory.

Using familiar mathematical notation, we can write the ranking of these outcomes as follows:

$$\text{Victory} > \text{Status Quo} > \text{Capitulation} > \text{War}$$

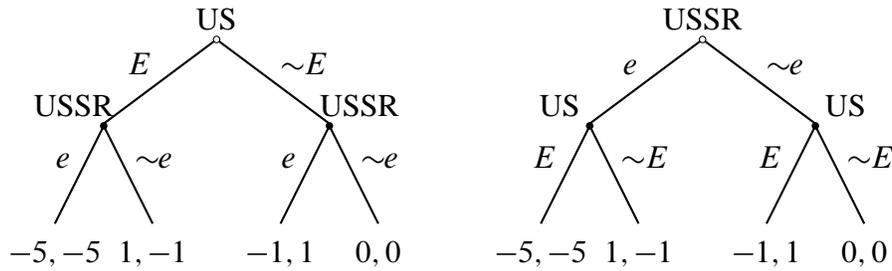
As we know from before, we can now assign numbers to these outcomes. Any numbers will do, as long as their ordinal relationship preserves the ranking of the outcomes. Here are several examples: Note that we don’t care how far apart each

Victory	>	Status Quo	>	Capitulation	>	War
3	>	2	>	1	>	0
12.4	>	11	>	0	>	$-\frac{1}{3}$
-1	>	-56	>	-57	>	$-1000\frac{5}{37}$
1	>	0	>	-1	>	-5

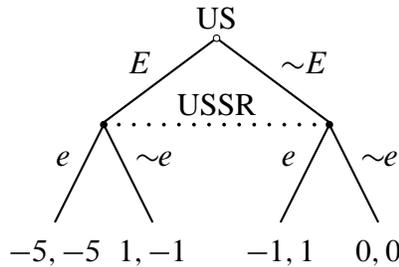
number is from the others (that is, their *cardinal* values), we only care about their ranking (that is, their *ordinal* values).

Obviously, some numbers are easier to work with than others. One common way of coming up with these numbers (first row) is to start with the worst outcome, assign it a value of zero, and then continue working upwards, adding one to each previous outcome. Another common way (last row) is to start with the outcome where “nothing has changed” and assign it a value of zero, and then assign simple numbers that would preserve the rest of the ordering. This is the method we shall use here.

We take the numbers we have assigned to the outcomes and plug them in our game description (the game trees). Each outcome is described by the set of payoffs, one for each player. In our example, each outcome is a pair of payoffs, one for the US and another for the USSR. We use the following convention: for each outcome, the payoffs are listed in the order the players appear in the game. For example, if the US is the first player (listed on top of the game tree), then the first number in the pair specifies the US's payoff and the second number specifies the USSR payoff. Figure 2 shows the revised game trees, with the payoffs attached to the terminal nodes (outcomes).



(a) US Moves First, USSR Moves Second (b) USSR Moves First, US Moves Second



(c) US and USSR Move Without Knowing Their Opponent's Action

Figure 2: The Three Possible Sequences of the Crisis Game, With Payoffs.

One easy way of labeling outcomes is by the sequence of actions that produce them. For example, the outcome *war* is produced if both players escalate; that is, if both players choose their escalatory action. The outcome can therefore be represented by the pair of strategies (E, e) . This pair is called a *strategy profile*. The order in which the strategies are listed in the profile is the same order in which payoffs are listed: first comes the strategy of the player specified first at the top of the tree, and then come the other players, in the order encountered as we work our way downwards through the tree.

5 Representing Incomplete Information

If, in addition to their own payoffs, players know their opponent's payoffs, we call the situation a game of **complete information**. As we have seen already, a game of complete information may be a game of perfect or imperfect information. If a player does not know the payoffs of his opponent, then the game is one of **incomplete information**. Saying that a player does not know the payoffs of his opponent is equivalent to saying that the player does not know that opponent's preferences.

5.1 Types and Beliefs

For instance, in our example, the US may not know whether the USSR prefers capitulation to war or vice versa. Suppose that the USSR can be either a "tough" type (which prefers war to capitulation) or "weak" (prefers capitulation to war):

USSR weak : Victory \succ Status Quo \succ Capitulation \succ War
 USSR tough : Victory \succ Status Quo \succ War \succ Capitulation

The optimal behavior of a weak opponent will probably be different from the optimal behavior of a tough opponent. So, how is the US to analyze this situation? How do we even represent this with a game tree? This is the big question that held back game theory for over two decades until Harsanyi came along and showed how one could model such situations and analyze them.

Using our way of assigning utilities, we can get two sets of numbers to represent these preferences.

USSR weak	Victory	\succ	Status Quo	\succ	Capitulation	\succ	War
	1	\succ	0	\succ	-1	\succ	-5
USSR tough	Victory	\succ	Status Quo	\succ	War	\succ	Capitulation
	1	\succ	0	\succ	-1	\succ	-5

We now have two distinct situations: (1) the US is facing a tough USSR, and (2) the US is facing a weak USSR. Let's suppose that the US gets to move first. With our method, we can easily write the game trees for these, as in Figure 3.

The only difference between the two games are the USSR payoffs in the branches of the trees following escalation by the US. This makes sense because the difference of preferences between weak and tough types is for the two outcomes that involve escalation by US. The preferences for the outcomes that involve US backing down are the same regardless of USSR's type.

Well, we have two games now, but really only one situation. How do we combine them? The US is **uncertain** about which **type** it might be facing: the USSR may be weak or may be tough. Even though the US does not know for sure which type it is playing against, it will form a **belief** about it. That is, the US believes that the

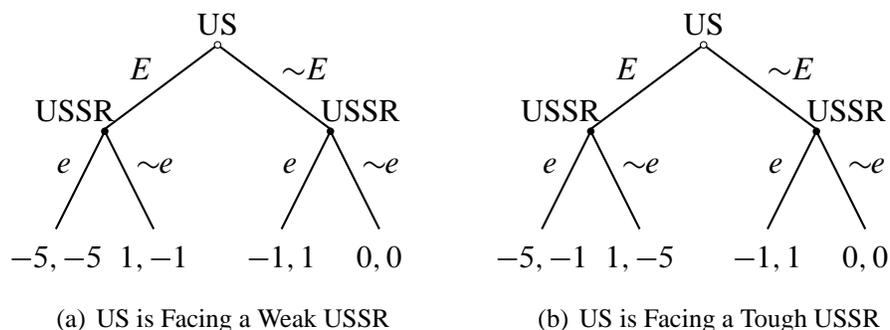


Figure 3: The Two Possible Situations.

USSR is tough with some probability p and is weak with probability $1 - p$, where p is a number between 0 and 1. For example, suppose the US believes that both types are equally likely, then we would use $p = 1/2$. If the US believes that the likelihood of the USSR being weak is 75%, then we would use $p = 0.25$, which is another way of saying that the USSR is tough with probability 25% and weak with probability 75%. Note that $p = 0$ represents the limiting case where the US knows for sure that the USSR is weak, and $p = 1$ represents the other limiting case where the US knows for sure that the USSR is tough. By varying p between 0 and 1, we can represent the entire spectrum of possible beliefs of the US.

Where do these beliefs come from? Mostly from past experience. That is, actors have observed each other in similar situations, have carefully noted verbal statements, behavior of political audiences, alliances, and so forth; all this knowledge is then summarized in an estimate of the opponent that constitutes the prior belief. For example, suppose the US faces down the USSR in a crisis in Iran. Is the USSR weak or tough? Who knows? However, past experience (e.g. with the Greeks, in Berlin, toward the Iranians, etc.) suggests that it is more likely that the USSR is weak and so would prefer to capitulate rather than fight. Hence, the US would begin the game with an estimate that the USSR is very likely to be weak, say $p = .15$, which means there is an 85% chance the USSR is weak. Obviously, these estimates are pretty rough, and we can actually represent various degrees of uncertainty mathematically (although we won't do it in this class to keep things simple).

The situation is quite complicated when there is very little basis on which to form these beliefs. For example, the Palestinians are set to elect a replacement for Yasser Arafat. Although most analysts judge Mahmoud Abbas to be more moderate than Arafat, the simple fact is that we do not yet know what his preferences really are. This will only be revealed by his policies once he is in office. If we want to design our policy before this happens, we have to form some beliefs. One plausible thing to do in the face of such uncertainty is to assume that he is equally likely to be moderate and extreme. Or, we could follow the journalists, and assume he is much more likely to be moderate. In other words, there is an art to figuring out what these beliefs are. This is why we shall analyze our games for all possible beliefs: that

is to say, we shall not assume a particular value for them, but will instead analyze everything in terms of the parameter p that represents the beliefs. When we move back to the empirical realm, we shall “plug in” values that seem appropriate, and our analysis would then yield a particular result for these beliefs.

5.2 Representing Beliefs

It is worth noting that the USSR knows its type but the US only has this belief about the possible distribution of types. Harsanyi showed that we can transform this game of incomplete information into a game of imperfect information, which we already know how to describe. To do this, we introduce a *fictitious player* called “Nature” (or sometimes called “chance”). This player moves first and determines the types of players. Some players learn the outcome of that move but others do not. For example, in our case, Nature determines the type of the USSR and the USSR learns its type while the US does not. However, the US believes that the USSR is rough with probability p , which is another way of saying that it believes Nature chose a tough USSR with this probability, as shown in Figure 4.

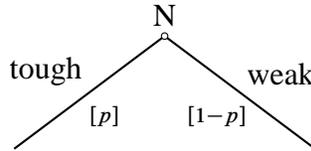


Figure 4: Nature Chooses Tough USSR with Probability p .

We now have everything in place to represent the situation because the move by Nature determines which of the two games in Figure 3 is actually being played. If Nature selects a tough USSR, then the game in Figure 3 (a) is played, and if Nature chooses a weak USSR, then the game in Figure 3 (b) is being played. Putting everything together, yields the tree of our incomplete information game transformed into a game of imperfect information, as shown in Figure 5.

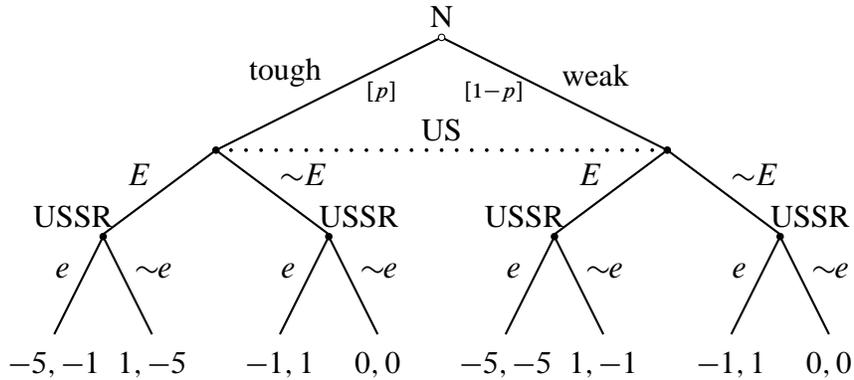


Figure 5: Incomplete Information and Observable Move by Uninformed Player.

Note that there is no payoff for Nature because this fictitious player is only used to model uncertainty. Also, the information set for the US represents the idea that when the US gets to move, it does not know which type of USSR was selected by Nature. All it knows is that it is at the left node with probability p and at the right node with probability $1 - p$. Further, because the US moves first, it also does not know the action of the USSR either.

The USSR, on the other hand, knows everything about the situation because all its information sets (four of them) are singletons. This means that it knows both its type and the action taken by the US. Whenever the USSR gets to move, it has both perfect and complete information, sometimes referred to as *full information*.

5.3 Example: Neither Player Knows the Other's Move

Suppose we wanted to represent a situation where the US has incomplete information about the USSR, and both players have imperfect information. That is, when taking their actions, they do not know what the other player is doing. The only difference from the previous example is that now the USSR does not know what the US has done although it still knows its type selected by Nature. Figure 6 shows the corresponding game tree.

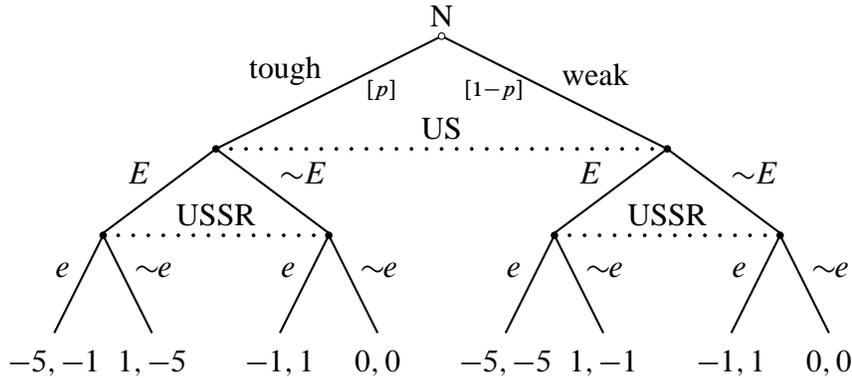


Figure 6: Incomplete Information and No Observable Moves.

USSR now has two information sets, and each one contains two nodes. Let's see what these represent. Consider the information set on the left. When the USSR gets to move in this set, it only knows that Nature has played "tough" but does not know whether the US has played E or $\sim E$. Consider now the information set on the right. When the USSR gets to move at this set, it only knows that Nature has played "weak" but does not know whether the US has played E or $\sim E$. In other words, whenever the USSR gets to move in this game, it knows its type (the move by Nature) but not the move by the US, which is what we wanted to describe in the first place.

5.4 Example: US Knows Move by USSR

Suppose we wanted to represent a situation where the US has incomplete information about the USSR but the USSR moves first and the US observes its move before acting. Caution: this is a bit tricky!

As before, Nature moves first and determines the type of USSR. Then, the USSR learns its type and selects an action. The US sees the action taken by the USSR but does not know which type of USSR actually took it. The question is simply where to draw the information sets that properly represent this situation.

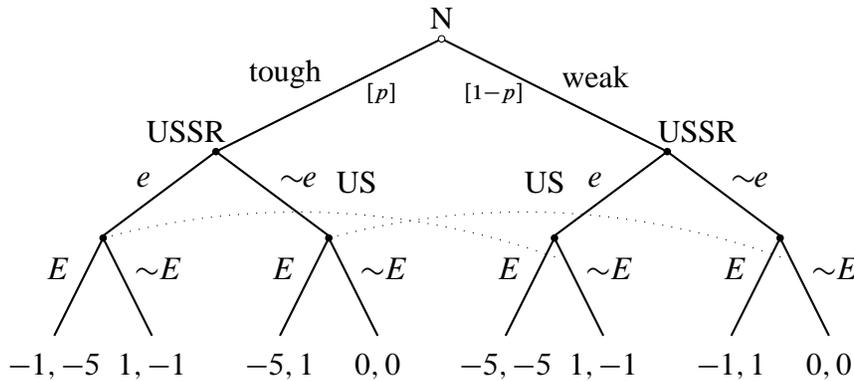


Figure 7: Incomplete Information and Observable Move by Informed Player.

The USSR has two information sets, each of which are singletons: the USSR knows its type. However, because it moves first, it does not know the action by the US. The US also has two information sets, one following each action by the USSR: the US knows the action taken by the USSR. Each such set contains two nodes: the US does not know whether the tough or the weak type has taken this action.

For example, consider the US information set on the left. It contains the node following “tough” by Nature and e by the USSR, and the node following “weak” by Nature and e by the USSR. In other words, when the US gets to move at this information set, it knows that the USSR has chosen e but does not know whether it is weak or tough.

It is this type of game that we shall find of great interest because it involves *signaling*. To see this, note that when the USSR moves, it knows its type and it knows that the US does not know its type. Therefore, it will expect that the US will try to infer something about the USSR’s type from its actions. In other words, the US will carefully observe what the USSR does, and will then revise its prior beliefs before choosing how to respond. Presumably, the US action will depend on how likely it believes the USSR to be tough. If that likelihood is great, then perhaps the US will be more accommodative. So, the USSR will probably attempt to signal that it is tough by taking an appropriate action. The problem, of course, is that the US knows this, so it will not take just any action at face value because it may be the case that the USSR is just pretending to be tough in order to extract

concessions. This means that the US will probably take actions designed to screen out its opponent's type: that is, actions that would cause a tough USSR to respond in one way, and a weak USSR to respond in another. But the USSR knows this, so its initial action will have to take into account these risks... As you can see, it gets pretty complicated pretty fast. But that's the essence of an international crisis: there is incomplete information and actors trying to uncover what they do not know about each other. As we shall see, this makes for a very risky situation.

6 Summary

We have to simplify reality in order to deal with it. The modeling enterprise provides for a disciplined way of doing so through repeated interaction between abstract models and empirical data that is used to refine them. We must assume actors are rational and intelligent although we do not have to assume they never make mistakes. That is, we assume they know what they like best, and their actions are designed to achieve their goals.

A model is an abstract stylization of real strategic situation and it is designed in a way to capture what the analyst believes to be its essence. In order to analyze a model, we have to find a way to represent this situation formally. We learned how to do this with the help of game trees. To describe a situation with a game tree, we need to know:

1. the players,
2. their options,
3. the outcomes,
4. the players' payoffs,
5. what players know,
6. in what sequence they move.

The players may not know two aspects of the game when they get to move: (a) they may not know what actions the other players have taken, or (b) they may not know the payoffs of the other players. When all players observe all actions leading to their information sets, the game is one of perfect information. When all players know all players' payoffs, it is a game of complete information. A game of incomplete information is transformed into a game of imperfect information by the introduction of a fictitious player called Nature.

Now that we know how to describe various situations, we must learn to analyze them. That is, how can we expect players to behave in a given strategic situation. We shall make heavy use of game trees in our analyses.