U.S. Foreign Policy: Deterrence and Compellence

Branislav L. Slantchev
Department of Political Science, University of California, San Diego

Last updated: July 9, 2014

1 Strategic Coercion

1.1 Brute Force and Coercion

What does it mean to use force? One use is to take possession, or deny possession of an object forcibly. For example, a country can occupy land, exterminate population, or repel an invasion—all through direct use of force at its disposal. A high school bully can simply beat up a smaller kid and take his lunch money. This kind of use of force is direct, and we shall call it **brute force**. The other type is less direct and involves threatening the opponent with pain without actually hurting him, at least in the beginning. Force can be simply used to hurt and, if we manage to uncover the points where it would hurt most, a threat to do so can motivate our opponent to avoid it. We shall call this **coercive use of force**. It is strategic in the sense that it seeks to persuade an opponent to do our bidding without destroying him.

Notice how in the "brute" case force settles everything — there's no room for bargaining. In the second case, our determination to gain our objectives and the opponent's desire to avoid being hurt — opens up room for bargaining. The coercive power is thus aimed at influencing the other side's behavior, primarily through his expectations. For example, our bully does not have to beat up the smaller kid. If his reputation is good (or bad) enough, he can demand the kid's lunch money and get it by just threatening to beat him up. Important to note that while no actual force is used in this case, force is used nevertheless. It is the *latent use of force* here that gets the result. Whereas the power to hurt is destructive, and seemingly aimless (because it does not immediately advance our objectives), it is useful because it can cause others to change behavior in accordance with our wishes.

Thus, **strategic coercion** is a type of bargaining where the opponent's expectations are influenced by the threat to hurt him. The threat must be understood and compliance rewarded. In other words, the opponent must be persuaded through the manipulation of threats. With force one may kill an enemy but with a threat to use force one may get an enemy to comply.

In order for coercion to work, the opponent must receive the threat of force—latent, not actual, use of force — whose success will depend on its credibility. We must then be able to relate it to a proposed course of action; and finally decide whether to proceed. This means that it is the expectation of more violence that will get us desired behavior (if at all), not actual use of force. This is the "coercion" in strategic coercion. "Strategic" refers to the process being a two-way street. Our actions engender reactions, we are influenced by our expectations of his expectations. This interdependent decision-making is called strategic interaction. Hence "strategic" in strategic coercion.

One might say that brute force is what happens when strategic coercion fails. If the threat to use force succeeds, then no (or very little) violence should follow. However, often it is necessary to use brute force to make the threat of further vi-

olence more credible. Consider the depredations of the horse peoples from the steppes who terrorized the Roman Empire (Atilla, the Hun), the Chinese and the Persians (the Mongol Genghis Khan), and almost the entire Middle East and India (Tamerlane). Let's take the greatest of them all, Genghis Khan, whose bloodthirstiness has become legendary. The Mongols were vicious by contemporary standards, no doubt about it. Yet, they were not wanton destructors. It is curious to see how often the great Khan used strategic coercion to compel the unconditional surrender of his opponents. He would lay waste to a city that resisted him but would spare one that would surrender. He would even try a diplomatic approach first before attacking (e.g. when he sent a caravan, which was sacked, and then an embassy, which was murdered, before resolving on the conquest of Khwarazm). Sometimes, the Mongols used tactical coercion: they marched captives from previous raids in front of their army to forestall further resistance.

When all is said and done, it was better for the Mongols if they could enjoy the booty and what we would call today "preferential trade treatment" without risking their skins. Yes, they glorified violence, but skin is skin, and a man has but one. So slaughter they did but they reaped the benefits of fear when remaining rulers voluntarily disgorged tribute to keep the hordes away. It is true that sometimes the Mongols massacred city populations following surrender, but generally they did not. In other words, by indulging in atrocities, they cowed their "audiences" into submission.

And intimidation they truly needed for the so-called hordes were not that numerous. In fact, the Mongols regularly fought out-numbered, and their army probably did not have more than 20,000 warriors at its core, and rarely numbered more than 80,000. In China, they faced over half a million men of war, and emerged victorious. In Russia, they obliterated armies twice their size. Perhaps a less over-awed numerous enemy would have been able to make a successful stand, especially if they realized how to counter the tactics (as people eventually did learn). In the end, however, the Mongols built the largest empire the world has ever known (the reason it's not nearly as famous is that it collapsed very quickly), a lot of it by conquest, and quite a sizeable chunk by intimidation.[1]

The Mongols used terror on a large scale much like the Romans did during their expansion. The Roman armies also regularly massacred entire populations of cities that dared resist them, sometimes going so far as to kill all the animals as well. The idea was much the same: by showing the consequences of defeat, they would discourage further resistance. And, given their nearly unbroken string of victories, the probability that any such resistance would end in defeat was too large for many

[1]Tamerlane, famous for piling towers of skulls, used a similar strategy, although it is arguable whether his particular taste for violence did not exceed the coercive needs (e.g. the massacre of 20,000 residents of Damascus, or 100,000 captured Indian soldiers after the battle of Panipat, among numerous others). Atilla, the Scourge of God, was actually an extremely skillful diplomat who used the fear his Huns inspired with regular success against the Romans.

polities. Whereas the conquest of a particular city or territory was an exercise in brute force, the manner in which it was done had a larger strategic purpose, making the threat of force a credible commitment for everyone watching. Terror, as we have come to relearn painfully, is a strategy of coercion.

1.2 Deterrence and Compellence

Brute force takes two basic forms, offense and defense. Strategic coercion similarly takes two basic forms: deterrence and compellence, which are roughly related to offense and defense in terms of their goals (change or maintain the status quo), and timing (actively pursued or waiting for opponent to engage).

Deterrence aims to persuade the opponent not to initiate action. We make the demand, explain the consequences of acting, and then wait (success is measured by whether something happens); if the opponent "crosses the line" we've drawn we take punitive action. One role for jails (punishment) is to deter potential criminals. The success of prisons is thus measured by how empty they are. It is hard to judge whether an event fails to occur because of successful deterrence or for other reasons. Deterrence is conservative: it seeks to protect the status quo. It is also, like defense, essentially a waiting game: the opponent has to move before a reaction is triggered.

Compellence aims to persuade the opponent to change his behavior. We make a demand of action, then initiate our own, and continue doing it until the opponent ceases. We can distinguish three categories of compellence. We persuade opponent (i) to stop short of goal; (ii) to undo the action (i.e. withdraw from land); or (iii) change his policy by changing government. Success of compellence is easy to see because it entails the reversal or halt of ongoing behavior. Again, this may happen for other reasons but it is hard to avoid the impression of doing it under duress. Compellence is active: it seeks to change the status quo. Also, like offense, it takes the initiative and engages the opponent until the latter relents.

Threats and promises are *conditional strategic moves* that can be used either for deterrence or compellence, depending on what they are supposed to achieve. A **threat** is a pledge to impose costs if the opponent acts contrary to one's wishes. A **promise** is a pledge to provide benefits to the opponent if he acts in accordance with one's wishes. Both threats and promises are intended to influence the expectations of the opponent and cause him to change his behavior. Both threats and promises are costly to the one making them although threats are costly if the player fails to influence the opponent, and promises are costly if the player succeeds.

In principle, both threats and promises can be used for either deterrence or compellence. Suppose we wish to compel the North Koreans to abandon their nuclear

program: we could threaten a punishment (cut off economic aid, limited strikes on the power plants) if they fail to comply, or promise a reward (invest in the country, build other plants) if they dismantle the program. Similarly, if we wish to deter them from pursuing such a program, we could try either a punishment or a reward. Although both could be used, in practice deterrence is best achieved with a threat, and compellence with a promise.

The difference is in the timing, initiative, and monitoring. A deterrent threat can be passive and static. One sets up the **trip wire** and then leaves things up to the opponent without any time limit. Throughout the Cold War, the U.S. constantly worried about the possibility of the USSR attacking Western Europe. The problem was that in conventional armaments, the Red Army was much, much stronger than what NATO could muster against it. A general war over Western Europe almost invariably meant that the U.S. would have to resort to nuclear weapons. The Americans could say "If you ever attack Western Europe, we shall fight back with all we've got, including nukes." Then they could sit back, wait, and watch. Only if the Soviets ever invaded would the Americans have to do anything.

The deterrent threat can be eroded by **salami tactics**, a strategy that takes steps that are small enough not to activate the threatened action, yet that bring the player closer to his goal. For example, the Soviets could send "military advisors" to Eastern Germany. Is this an invasion? Of course not, they are helping an allied communist nation organize its defenses against the imperialist Western aggressors. Before you know it, they bring several tank brigades to Berlin. Is this an invasion? Of course not, they are using the equipment to train said defense forces. Then they instigate a couple of incidents along the perimeter with West Berlin. Is this an invasion? No, these are provocations by the imperialists which demonstrate the need for defenses, which is why we are sending a Red Army division there to make sure things stay calm. They cut off the corridor to West Berlin. Is this an invasion? No, they are exercising their right to sovereignty, which was threatened by the West in those border clashes. West Berlin suffocates and the East Germans offer to begin supplying it (while Soviet tanks are making sure nobody else can get through). Is this an invasion? Before you know it, the Soviets are in possession of Berlin, with a sizeable contingent of the Red Army ready to strike. By the time you think of an answer, you find yourself hoping they would spare Britain.

Thus, the deterrent threat had to be invulnerable to salami tactics, and it would have to ensure that the Americans would actually want to respond to an invasion by defending Europe. As we shall see, stationing American troops in Europe provided a trip-wire (or **plate glass**) that performed these functions. The presence even of a significant U.S. force there was not enough to win a land war against the Red Army. However, it did ensure that if the Soviets ever decided to attack, they would have to do so in strength that would be sufficient to overcome these forces. This meant that the Soviets would have to use such a large number of troops that there would remain no doubt about their intentions. An attack on the U.S. contingent in Europe

would be nothing less than the opening salvo in a general war. It would shatter the plate glass, so to speak.

This should therefore tend to discourage the Russians from adventurous policies that would probe American resolve to defend Europe (it did). Whether it would work like that elsewhere in the world was an open question (it did not). Further, apart from making the Soviets reveal the scope of their aggressive intentions, stationing Americans in Europe would enhance the credibility of the threat to fight the Red Army if it did invade. As we shall see, many Europeans (and Americans) doubted whether the U.S. was prepared to go to general, possibly nuclear, war with the Soviet Union over Western Europe. If the Russians did invade, they would inevitably have to overcome the resistance of the American forces by destroying them. It is highly unlikely that the U.S. would calmly accept the deaths of tens of thousands of its citizens: the U.S. would be compelled to react and fight even if it cared little for Europe itself. As Schelling put it, the purpose of these troops there was to die gloriously.

Thus, stationing troops in Europe could serve as plate glass by forcing the Soviets to come in strength, and as a trip-wire by forcing the Americans to respond in kind. Attack would be unequivocal, and defense nearly automatic.

Trying to achieve such deterrence with a promise is possible but harder. The U.S. could say something like "Every year that you do not attack Western Europe, we will provide you with economic aid." This requires continuous action which could actually strengthen the enemy and perhaps encourage him to do the very thing that the promise is supposed to help avoid. However, this is not to say that deterrence cannot be achieved through promises. A powerful argument can be made for improving the status quo for dissatisfied powers to such an extent that destroying it would not be in their interest. (You should carefully read John Mueller's chapter on this topic.)

Unlike deterrence, compellence must have a deadline. We cannot follow U.S. ambassador to the United Nations Adlai Stevenson who, when told by the Russians that they would inform the U.S. about the movement of nuclear weapons toward Cuba in "due course," responded by saying that he was prepared to wait until hell froze over. Quite a dramatic statement, but exceedingly bad strategy. Why? Because the Soviets could procrastinate, if not until hell froze over, then until they had their missiles in place and operational. Without a deadline (e.g. "tell us in 24 hours or we shall assume you are installing them and strike to remove them"), the compellent threat can be seriously undermined by delay.

A compellent promise can induce the other party to bring to your attention its good behavior. For example, we could tell the North Koreans that if they dismantle their nuclear program, we shall provide them with economic aid. This should encourage them to come to us with evidence of such dismantling because they will be eager to persuade us to fulfill our promise. (Of course, this does not guarantee that they would not cheat. As we see below, any evidence that they produce must be a

costly signal or we would not believe them.)

Generally, if deterrence is the goal, you would do best by choosing a status quo such that if your opponent acts contrary to your wishes, what you do is punishment. This usually involves making the status quo sufficiently pleasant and threatening to make it much worse if he disrupts it. You can also promise to make it progressively better as long as he persists in compliance.

If compellence is the goal, you would do best by choosing a status quo such that what you do if the opponent complies with your demand becomes a reward. This usually requires that you make the status quo sufficiently unpleasant and promise to improve it if he complies. You can also threaten to make the status quo progressively worse if he persists in non-compliance.

2 Typology of Deterrence

We can distinguish between two types of deterrence with respect to the relationship between the defending actor and the challenger, and the perceived timing of the action. The idea is that the defender issues a deterrence threat that is supposed to prevent the potential challenger from attempting to overturn the status quo.

First, the question is the identity of the actor the threat is designed to protect. **Direct deterrence** refers to threats that are designed to prevent direct attacks on the defender itself. Examples include any posturing that attempts to persuade the potential challenger not to initiate an action against the state that issues the deterrent threat. During the Cold War, both the U.S. and the USSR engaged in direct deterrence with respect to each other, each seeking to prevent the other from trying to attack the two mainlands. By its very nature, direct deterrence is usually quite credible: after all, an army would defend its homeland almost always.

Less clearly credible is **extended deterrence**, which refers to those occasions on which the defender extends his protection to a third party, usually called a *protégé,* and warns that he would resist an attack upon the protégé by the challenger. For example, these days Taiwan is an American protégé, with the U.S. engaged in extended deterrence to prevent China from absorbing the island which it regards officially as a renegade province. Because, by its very nature, extended deterrence involves expanding the "national interest" on a larger sphere than protection of the homeland, it is inherently more amorphous and less well-defined.

A second way to differentiate among types of deterrence is with respect to their timing: is the deterrent commitment intended to prevent some vague potential threat posed by a would-be attacker, or is it intended to prevent an immediately pending action? **General deterrence** refers to situations where there is no clear and present danger of attack and yet an underlying antagonism persists. An example of such a commitment is the American treaty with Japan that secures the island nation against any potential aggressor even though no such threat is apparent at present. The treaty

was designed at a time when the USSR could be counted on to press for concessions from the country recently battered into submission by the Americans (and the Red Army in Manchuria), and totally demilitarized. General deterrence is also an apt characterization of U.S. protection of Western Europe from the potential menace of the Red Army during the Cold War.

Immediate deterrence, on the other hand, refers to situations where the challenger can mount an attack at any moment. For example, in 1950 the Chinese attempted to deter the U.S. from pursuing a war of conquest into North Korea but their warnings were ignored, and the Chinese swarmed across the Yalu River to push back the American forces. A successful example would be the 1970 warning by Israel against potential invasion of Jordan by Syria. Every crisis that ends in the outbreak of war is a case of failed immediate deterrence.

Combining these dimensions of deterrence at hand, we can distinguish four generic categories in Table 1.

		Threat Posed by Attacker	
		Actual	*Potential*
Target of Attack	*Defender*	**Direct-Immediate** (Outbreak of Winter War, 1940)	**Direct-General** (Sino-Soviet border dispute since 1970)
	Protégé	**Extended-Immediate** (U.S.-Chinese crisis over North Korea, 1950)	**Extended-General** (U.S. forces in South Korea since 1953)

Table 1: A Typology of Deterrence. Source: Paul Huth, 1988. *Extended Deterrence and the Prevention of War,* p. 17.

The Arab-Israeli conflicts would usually fall into the direct deterrence categories: with Israel attempting general deterrence to ward off attack upon its territory, with the periodic failure of its policies and an eruption of yet another war. In the critical days preceding the Six Days War of 1967, for example, Israeli policy-makers were crucially concerned with the credibility of their deterrent posture against Egypt. Once they convinced themselves that immediate deterrence (which they tried to achieve by mobilization) would fail, the road to war lay open. Conversely, the stunning success of 1967, persuaded Israel that its posture would not fail to deter in the future, and this belief goes a long way in explaining their unpreparedness in 1973 when the Arab forces struck back exposing the weakness of the general deterrence policy.

The Great Powers are the states that can afford to indulge in extended deterrence, and many wars have occurred when the protégé drags its protector into conflict by its intransigence, which itself is a result of the promise of security. This was the case with both Serbia and Austria-Hungary in 1914. The Russians had guaranteed the security of Serbia and encouraged the government to resist the ultimatum delivered

by the Austrians. The Austrians themselves were goaded by the Germans who issued the so-called "blank check" promising to come to the aid of the empire come what may. In the end, the two defenders found themselves at war with each other over a conflict between their protégés.

This problem of entrapment is what usually causes commitments of extended deterrence to be somewhat less than firm and absolute, which, of course, in turn contributes to them being less credible, and therefore open to more frequent challenges. Hence, such commitments are inherently riskier for the defenders. Take the example of American commitment to Taiwan. The problem is well-known: should the U.S. promise unconditional defense of the island, it may well choose to defy China and declare full independence, something that the Chinese have repeatedly insisted would be a *casus belli* (cause of war). Such a commitment may encourage Taiwan to pursue a reckless policy that would endanger the peace between U.S. and China. On the other hand, should the U.S. appear neglectful in its promise to defend the island, China may well find the courage to attempt to take it over by force, an outcome that (for now) is not in American interests for it would alter the security balance in the region and throw into doubt American commitments in South Korea and Japan, perhaps triggering an arms race when these countries seek to defend themselves from possible future Chinese aggression. This is why the U.S. has pursued a rather vague policy of *strategic ambiguity*, which means it sometimes supports Taiwan and sometimes does not, and it is never clear exactly how committed the U.S. is and to what. All that both sides know is that the guarantee is not absolute, and yet it is perhaps strong enough to ensure defense against unprovoked Chinese attack.

The biggest problem with using threats and promises is that one may have no incentive to follow through on them because they are always costly to the player making them.[2] That is, they may not be credible. But as we have seen, if they are not credible, they will have no effect on the expectations of the opponent, who will ignore and refuse to believe them. If they fail to influence his expectations, he will not change his behavior, and we shall be stuck with having to deal with the consequences. Thus, the art of credible commitments constitutes an enormously important part of achieving the goals of national security.

We now investigate several strategies for making commitments credible. We divide the discussion into three broad categories: (i) reducing freedom of choice, (ii) manipulating future payoffs, and (iii) manipulating risk. We want to know how one could act strategically to acquire credibility, and avoid capitulating because of the credibility of its opponent. Generally, we shall see that the strategies involve choosing how to sequence one's actions (that is when to act), and deciding how

[2]It is worth repeating that a threat is costly if it fails, and a promise is costly if it succeeds. If the threat fails, one must carry out the costly action that was threatened. If the threat succeeds, one need not do anything. If the promise succeeds, one would have to deliver the benefits, which is costly. If the promise fails, one need not do anything.

costly these actions should be, or what risks to run. Finally, we investigate whether the credibility of a threat depends on hurting your opponent more than you hurt yourself by executing it.

3 Reducing Freedom of Action

The first method of acquiring credibility is to structure the situation in such a way that you would have no choice but to carry out the action you have threatened or promised. Conversely, you may attempt to maneuver the opponent into a position where it will be up to him to make the painful decision.

3.1 Constraining Choice

Limiting one's choices in an *observable* and *irreversible* way may help establish a credible commitment by eliminating an embarrassing richness of choices that tempt one to escape the commitment. When you think about it, the credibility problem arises from the temptation not to carry out the action you are supposed to. If you remove these tempting alternatives, then you would have no way of choosing them. That is, you will have no choice but execute the threat or promise you have made.

3.1.1 Automatic Fulfillment

An extreme way of constraining your choices is by ensuring **automatic fulfillment**. The idea is to remove the element of human decision from the course of action altogether. If you set up a system that automatically retaliates and that cannot be stopped once activated, and if you can demonstrate to your opponent that such a system is in place and you do not have the freedom to change that, then your commitment will be credible. There is no sense in risking an action against a system that makes automatic decisions.

If you ever see Stanley Kubrick's famous film *Dr. Strangelove* (you should it is very funny), you will note the so-called *doomsday device* designed by the Russians. This device is triggered by an atomic explosion on Soviet territory. When it explodes, it contaminates the entire atmosphere. The only problem is that the Soviets did not tell the Americans about it. You should watch the film to see what happens.

Obviously, even though such a commitment is perfectly credible, it can be incredibly dangerous if there is even a tiny chance that things could go wrong. During the heated years of the Cold War, the United States had a strategy that kept a significant portion of Strategic Air Command (SAC) bombers in the air at all times. In the event of a crisis, they automatically proceeded to their destinations, mostly targets in the Soviet Union. The danger, of course, is that if they did not receive the cancelation command (failure of communications), they would actually cause war even if a crisis was resolved. Hence, the fail-safe protocol according to which, planes

were to proceed first to pre-designated points around the globe (outside Soviet territory) and hold there until they receive an explicit command to attack. If no such command arrived, they were to abandon the mission and return to base. The idea was that if communications failed, the potential error (they might fail because the Russians jammed them or destroyed the command centers) would be one the safe side. The film *Fail-Safe* is an excellent take on how things might go terribly wrong anyway.

For example, a warning system that activates the automatic defenses has to be sensitive enough to detect an attack early and not be fooled into ignoring a scattered attack that does not rely on obvious concentration of missiles and bombers. Such a warning system can never be perfect. In particular, if it is sensitive enough to react when necessary, it will also sometimes get triggered by innocent events (e.g. a stray satellite falling into the atmosphere). Even with a minuscule danger of such an error, the fully automated solution ensures that disaster will occur with certainty. Generally, human intervention will be required for sound judgment, which, of course, would mean that the system is not fully automated.[3]

Automating the response was actually a tactic that the Russians claimed to be pursuing for a while. Chairman Khrushchev told the Americans that it did not matter whether Berlin was worth more to them or to the Americans; if a military confrontation ensued, Khrushchev claimed that the Soviet rockets would *fly automatically*. The interview was published in the premier policy journal "Foreign Affairs" and caused quite a stir at the time.

3.1.2 Delegation

A somewhat more plausible way of constraining your choice is to delegate it to someone else. It is not mechanical, but it is not in your hands either. It may help your credibility if the agent responsible for implementing the action is less tempted to avoid it than you are. For example, if Congress is more hawkish on foreign policy issues than the President, the President can benefit from delegating all responsibility for agreements to Congress. He can then tell the Soviets that even though he would love to sign an agreement very favorable to the Soviets, he cannot do it because it is the responsibility of Congress to ratify it, and they (being hawkish) would never accept it in this form: the Soviets must concede more.

More interestingly, a leader may constrain his choices by simply making it impossible for him to make decisions. For example, a civilian government may delegate control of nuclear weapons to the military, which has a clear mission to defend the country, may not be subject to the pressures and debates of a civilian government,

[3] Although perhaps infeasible for national security, automatic fulfilment systems are quite common in other areas, such as trade policy. Many countries have procedures that automatically retaliate with import tariffs if another country tries to subsidize its exports to that country. These usually go under the name of "countervailing duties."

and so may be prompt with their use. The French, for example, toyed with this idea for a long time. Similarly, there were serious proposals to let the Germans have direct control over NATO nuclear weapons in Europe because they could commit much more credibly to using them against invading Russians than the Americans. Or, one can let computers play out the warfare scenario and relinquish choice completely.

Of course, delegation is not fool-proof because it may backfire (and it does reduce your flexibility), and it may not be believed. For example, a leader used to totalitarian mode of government may simply refuse to believe that the President is constrained in any meaningful way by Congress. If the constraint is real and is not believed, it may end up producing the exact insurmountable obstacles it was designed to solve.

3.1.3 Burning Bridges

An even more plausible strategy is to eliminate the possibility of taking the tempting action altogether. This is called **burning bridges** and comes from the ancient practice of armies burning the bridges behind them to ensure that they have no choice but proceed forward.

To illustrate this idea, consider our original crisis game with imperfect information, and recall that it has three Nash equilibria. Suppose that player 2 could move first and eliminate the possibility of backing down, as shown in Figure 1.

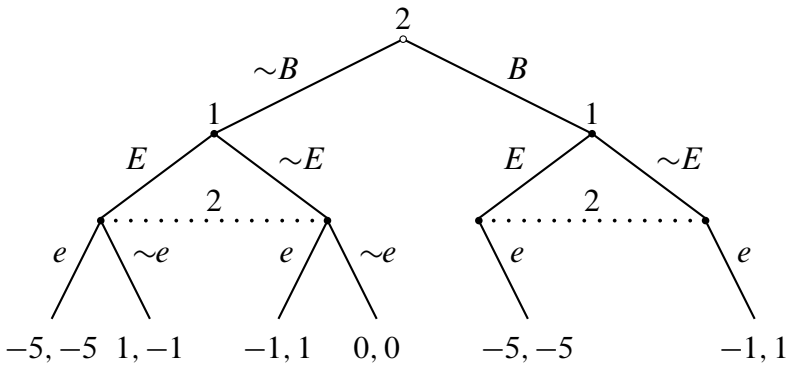

Figure 1: The Crisis Game with Burning Bridge Commitment.

The initial action is B (burn the bridge) or $\sim B$ (do not burn it). If player 2 chooses not to burn the bridge, then the original crisis game is played. If player 2 burns the bridge, he cannot choose not to escalate in response to player 1's escalation. Consider now the subgame that begins with player 1's move at his second information set (following B by player 2). Player 2 will always escalate because he has no other choice, and so player 1's best response is to choose $\sim E$ because doing so yields -1, while escalation yields -5. Thus, playing B gives player 2 an expected payoff of 1 (because player 1 will not escalate while player 2 will).

Consider now the subgame that begins with player 1's move at his first information set (following $\sim B$ by player 2). We know that this subgame has three Nash equilibria: two in pure strategies and one in mixed strategies. We have argued that for a meaningful crisis, the mixed-strategy equilibrium is the reasonable prediction.[4] Recall that this equilibrium is $\left\langle \frac{1}{5}, \frac{1}{5} \right\rangle$, that is, each player escalates with probability 20%. Disaster occurs with probability 4%, submission by player 1 and submission by player 2 each occur with probability 16%, and the status quo prevails with probability 64%. Let's compute player 2's expected payoff from this game:

$$\mathrm{U}_2\left(\left\langle \frac{1}{5}, \frac{1}{5} \right\rangle\right) = (0.04)(-5) + (0.16)(1) + (0.16)(-1) + (0.64)(0) = -0.2.$$

Thus, player 2 could expect to get -0.2 if he chooses $\sim B$, which is strictly worse than 1, which is what he would get by choosing B. Therefore, in the subgame perfect equilibrium of this crisis game, player 2 would choose to burn the bridge, which would lead to the capitulation by player 1.

The core idea is to make the tempting option unavailable to you. Thus, when Hernan Cortez landed in Mexico, he beached his ships to ensure that the soldiers would have no way of retreating, which would cause them to fight as hard as possible. During the last months of the Second World War, the Japanese resorted to *kamikaze* attacks: the planes only took enough fuel to reach the American ships, in which the pilots were supposed to ram them. In the less violent arena, the common European currency (the Euro) is a similar commitment device: by making abandonment of the Union exceedingly costly, it ensures that the participating countries would work hard to make it work and would comply even with painful decisions. In fact, it was precisely because of this high level of commitment the Euro created that Great Britain chose to stay out of the monetary union.

Alternatively, one could try to make tempting options available to one's opponent in the hope that he will make use of them. That is, while you may want to burn the bridges behind you, you definitely do not want to burn the bridges behind your opponent. As Xenophon observed during his march with Greek troops across Persia, in battle you want to leave your opponent a way out: when things get tough, he will take it. In other words, we are applying the logic to the opponent. The same thing that would cause us to renege on our commitment would cause him to renege on his. Hence, giving him a graceful way out eases our task: if we know that he can back down because we have given him a loophole, and if he knows that we know, our threat to press him becomes credible.

[4]The analysis that follows can be done for the pure strategy Nash equilibria as well. Suppose players expect to play the $\langle E, \sim e \rangle$ Nash equilibrium: that is, player 1 escalates and player 2 does not. The expected outcome for player 2 will be -1. This is strictly worse than 1, which is what he would get by playing B, and so burning the bridge is optimal. You can see that if players expect the equilibrium $\langle \sim E, e \rangle$, then burning the bridge is just as good as not burning it, so it is still optimal to burn it.

Although this makes straightforward sense and seems obvious, people often get it completely wrong. Just look at the famous *Illiad* by Homer (now a major motion picture directed by Woflgang Petersen). Much of the book concerns repeated attempts by the defending Trojans to burn the ships of the invading Greeks! Instead of encouraging the Greeks to leave, accomplishing this mission would have caused exactly the opposite. You want to burn your bridges, but you often want to build many for your opponent.

3.2 Relinquishing Initiative

Relinquishing initiative saddles the opponent with the painful choice of making the last step that results in disaster for both. If he has a chance to back down, he will take it. Therefore, it is crucial not to maneuver the opponent into a position from which he cannot retreat. In particular, if the opponent has managed to preempt you and constrain his choices, relinquishing initiative automatically leads to disaster.

Consider a highly stylized example of the Cuban Missile Crisis. After finding out about the Russians secretly placing nuclear missiles in Cuba, the U.S. considered several options, from the mildest (quarantine, which is what got implemented), to progressively more dangerous and escalatory ones, like a limited air strike designed to take out the missile sites, a massive air strike, and even a land invasion.

The quarantine stood apart from the more military responses in terms of who had to take the next escalatory step. Suppose the U.S. can choose between a military action, (M), and a blockade (B). If it chooses the military option, then the USSR can respond by fighting or not. If it fights, a war results where both suffer greatly. If it does not, the U.S. wins and the USSR loses a lot. In fact, because of failing to respond to a direct military challenge of the rival superpower, it loses more than by fighting a limited engagement over Cuba.

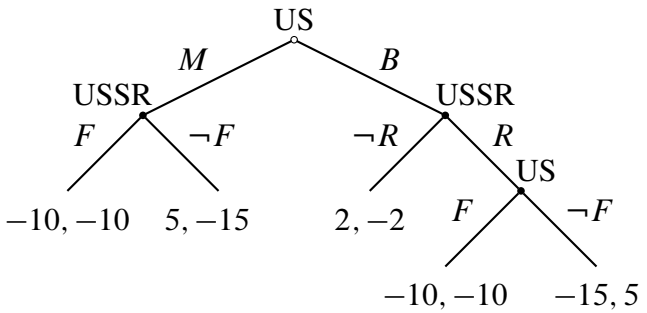

Figure 2: A Stylized View of a Missile Crisis.

If the U.S. picks the blockade, the USSR can choose whether to run it or not. If it does choose to run it, the U.S. can decide whether to initiate the military option or not. Again, if the U.S. fails to respond militarily to direct Soviet challenge, the

Soviets gain and the Americans lose badly. If it does respond, war results. If the USSR does not run the blockade, the Americans win concessions from them.

We solve by backward induction. Given blockade and the Soviets running it, the U.S. prefers to fight. Given that the U.S. would fight should they run the blockade, the Soviets prefer not to run it. On the other hand, given a military action by the U.S. the Soviets prefer to fight. Given that the Soviets would fight a military action but would not run a blockade, the U.S. strictly prefers to impose a blockade instead of risking war.

Of course, this is a very simple setup that does not do justice to many other considerations that went into the frenzied weeks of October 1962. However, the basic feature is clear: Imposing the blockade shifted to the Soviet Union the responsibility of making the escalatory step that would have resulted in war. Note that we have *not* assumed that the Russians would not fight if challenged. On the contrary, we assumed that both the Russians and the Americans would fight if they had to! However, saddling the Russians with the choice to initiate the war conferred a great advantage on the U.S., causing the Russians to back down.

The U.S. relinquished initiative. Instead of initiating the military strikes (and thereby ensuring an automatic reprisal by the Soviets), the U.S. put up the blockade and let the Russians take the initiative in running it. Having been maneuvered in this position the Russians had no choice but back down or start a war.

3.3 The Dynamics of Mutual Alarm

The most important limitation of using these tactics (aside from making actions truly irrevocable and observable) comes from the very mechanism that generates their credibility: Your inability to do something else and avoid incurring the costs. Decisions in international crises are made under intense pressure, and without knowledge of the exact actions (or intentions) of the opponent. This means that irrevocable commitment always carries the real danger that either the opponent will not see it in time or will see it only after having himself made a similar irrevocable commitment. Because there is a race to pre-empt the opponent with your own irreversible commitment, there is a huge incentive to do it as quickly as possible. This holds both for you and your opponent, and so in the rush you may both become committed to a course of action you both want to avoid.

Here's an example from the July Crisis of 1914 that led to the First World War.[5]

[5]This is a highly simplified version of events that focuses exclusively on the pressures of mobilization moves. The crisis was much more complicated, and these moves were not the primary reason it escalated to war. An excellent recent account is Christopher Clark. 2013. *The Sleepwalkers: How Europe Went to War in 1914.* New York: HarperCollins. Despite the title, which seems to suggest that policy-makers drifted into the war without realizing the impact their seemingly rational actions had, the book goes a long way toward the conclusion that Russia and France were the two powers responsible for triggering the war on the continent instead of allowing Austria-Hungary to coerce Serbia without fighting (that is, for ending up with a regional war instead of either no war at

Mobilization is the process through which a country gears up for war. It involves calling the reservists, arming them, and transporting them to the front lines along with piles of equipment, food, fuel, and support personnel. Mobilization is enormously complicated and every country has carefully prepared plans on how to execute its own. It is also terribly expensive because it involves not only removing men from their jobs but also disrupting commercial schedules of railways and, in more modern times, aviation.

Once mobilization is under way, it is hard to stop, and nearly impossible to restart if stopped. Once completed, it cannot be maintained indefinitely. Once its resources and armies are mobilized, a country must use them or lose them. That is, nobody can afford to field armies without action for a long time. The forces either get used or the soldiers must be sent home.

This momentum implies two things. First, a country is vulnerable if it stops its mobilization midway before it is completed because the resulting chaos makes it next to impossible to restart the process quickly. If it stops then, an adversary could use this opportunity to strike. Second, once mobilized a country becomes a great menace to its potential adversary because it must either strike or demobilize. This brief window of opportunity makes it hard to negotiate at leisure a way out of the crisis.

Now think about the combination of these two effects. A country that begins mobilization will be extremely dangerous to its adversary once mobilization is completed. However, it is also extremely vulnerable during mobilization and in the event it stops the process. Knowing that it will eventually have to face the fully mobilized resources of this country, an adversary might be tempted to strike sooner, making the crisis even more unstable. (Crisis stability refers to the likelihood that the crisis would end up in war.)

Let's look again at that fateful summer of 1914. Austria-Hungary had issued its ultimatum to Serbia and it looked like it would go to war with the little Balkan state. The Russians faced a dilemma. They had to mobilize to threaten the Austrians sufficiently to prevent them from finishing off the Serbs. A full mobilization, however, would also threaten Germany and perhaps provoke it into mobilizing itself.

---

all or a very limited local war of Austria-Hungary against Serbia). This should be complemented by reading Niall Ferguson. 2000. *The Pity of War: Explaining World War I.* New York: Basic Books. He presents a strong case that Great Britain should have stayed out of the continental war, and it was its involvement that turned a regional conflict into a global total war. These recent works contradict the long tradition of blaming Germany for the war, a tradition that goes back to the Versailles Treaty itself, but which also received a boost in 1961 when Franz Fischer published *Germany's Aims in the First World War*, in which he argued that Germany had deliberately instigated the war in a bid for world power status. Few today would accept this version without serious modifications and many would not accept it at all. See, for instance, Annika Mombauer. 2002. *The Origins of the First World War: Controversies and Consensus.* London: Longman. In case you are curious about just how complex an explanation of the outbreak of this war can be, read the superb survey by James Joll and Gordon Martel. 2006. *The Origins of the First World War,* 3rd Ed. London: Routledge.

The Russians did have plans for partial mobilization in the south, which is exactly what they needed to threaten the Austrians only. However, once started, this partial mobilization could not be converted into full mobilization because of the way the railroads were scheduled. This was a problem because initiating partial mobilization, while not threatening to Germany, would expose the Russians to a German attack. The Russians had to trust the Germans not to exploit this opportunity.

Or they could hedge against it and order full mobilization just in case. But full mobilization is preparation for total war and Germany's reaction was, of course, to mobilize itself. Germany also faced a dilemma. The Russians were allied with the French and if Germany attacked Russia, it would find itself fighting on two fronts when the French, in accordance with their agreements with the Russians, attacked from the West while Germany was engaged in the East. Or, even without the alliance, Germany had reasons to fear that France might use the opportunity and try to regain Alsace and Lorraine which she had lost after the Franco-Prussian War of 1871.

At any rate, there was a real danger that if Germany mobilized and threw all its forces in the east, the French would attack across its exposed western borders. The German high command believed that finishing off the French would be quicker and easier than defeating the Russians, and so in an event of a war with Russia, the German war plans called for a surprise attack on France first. The mobilization plans, just like the ones of the Russians, were also impossible to reverse once put into motion, and so the Russians ordered full mobilization out of fear that Germany might exploit a partial mobilization, the Germans mobilized for war against France out of fear that the French might exploit their potential vulnerability. To make matters worse, Germany's plans for France required the capture of the Belgian city of Liege with its major railroad junction. The Belgians had declared neutrality but were expected to mobilize when Germany did, just for security purposes. This would make the capture of Liege very difficult and would, at the very least, delay the thrust into France putting the German operation in jeopardy. As a result, the German plan was to attack Belgium by surprise within two days of starting to mobilize. For Germany, more so than for any other country, mobilization meant war and there was no time to backtrack without incurring serious tactical disadvantages. Britain was the guarantor of Belgium's neutrality, and such an attack would certainly help the British government bring the country into the war against Germany. The war was destined to become at least European in scope.

The military doctrine at the time emphasized speed of mobilization and surprise attack. It was believed that the country that could finish its mobilization first and attack its opponent before the latter was ready could gain a significant advantage and perhaps even win the war. This creates an awfully dangerous situation. A statesman who has the military instrument at the ready and knows that he must use it or lose and who further knows that his opponent is in the same position, faces a fateful decision where hesitation to strike first may mean national defeat.

Notice how this provides a motivation for war quite apart from its other causes. This one is mechanical, it is produced by the military technology of coercion and planning. A vulnerable military force provides a temptation to the enemy to strike until this window of vulnerability exists. Therefore, a vulnerable military force cannot afford to wait and must attack first.

If striking first carries such an advantage, the other side may think that you want to do it even if you really do not. But if it thinks you might do it, then it is tempted to do it first even though it may not want to do it. But if you know that it might be tempted in this way, you now think that it might strike, and so you might prefer to strike first because you think that it would do so anyway. Both of you provide each other with justification to strike first. These interacting expectations produce a chain of the now familiar logic: he thinks that I think that he thinks that I think... he thinks that I think he will attack, so he will, so I must.

The end result is war that neither may have wanted, an accidental war that is not due to some mechanical failure but to the expectations that shift in such a way due to the constraints of technology that both sides become convinced that war is inevitable, making it truly inevitable in the process. In a way, because technology commits the players to following certain strategies, they may become victims of circumstance and make the fateful decision to start fighting even though they would rather not.

It is the fear of surprise attack that influences expectations in this way, and this fear is generated by one's own vulnerability and that of its opponent. Especially that of his opponent because what generates the escalating reciprocity of fear is the expectation that because the opponent is vulnerable, he might strike first.

We reach the somewhat paradoxical conclusion that to increase crisis stability one must work to *decrease* the vulnerability of its opponent's military forces. But compelling one's opponent requires destroying a significant portion of these forces, which makes it desirable to *increase* their vulnerability. Herein lies the problem: An action that is designed to reduce the likelihood of war makes it more difficult to win the war should the war occur. Conversely, an action that increases the likelihood of war also makes it easier to win the war. You can see how a prudent state would probably hedge against losing a war and will choose a strategy of the second type, making crises less stable and far more dangerous.

Still, during the Cold War, the two superpowers pursued strategies that decreased the vulnerability of the military forces and increased the vulnerability of the civilian population, thereby providing powerful incentives not to jump the gun in a crisis. Once each side acquired second-strike capability, the era of mutually assured destruction (MAD) began. Each country could absorb a first strike by the enemy and then return a devastating counter-blow.

Acquiring this capability involved (a) building a lot more missiles—what some people mistakenly called "overkill" in the belief that once the U.S. had enough nuclears to blow up the Russians it was unnecessary to build more, completely

missing the point that the relevant quantity was not the total number of nuclears but the number that could survive a surprise attack by the Russians; and (b) rendering the existing forces invulnerable to enemy bombs. The second strategy involved dispersing of missile sites and bombers, hardening missile silos, and, once it became technologically possible, placing nuclear weapons on hard to detect submarines.

In addition to making their military forces less vulnerable, the two superpowers made their civilian populations more vulnerable when they agreed not to build anti-ballistic missile systems (ABMs). This venerable treaty persisted until George W. Bush unilaterally withdrew the U.S. from it. The purpose, however gruesome, was to supplement the stability-inducing invulnerability of the military. If you have second strike capability and your enemy's cities are vulnerable, then your enemy is unlikely to attack you first by jumping the gun in a crisis. But if your enemy is unlikely to launch a surprise attack, then you have no reason to launch one either, and so crises become much more stable.

3.4 Severing Communication

Also note the requirement that these commitments be observable by the opponent. One tactic to undermine such commitments is therefore by cutting off communications and making yourself unavailable to receive the threat. We have all used this strategy when screening calls from people we do not want to talk to. We know that if we pick up the phone, common courtesy would compel us to waste several minutes, which we really want to avoid. It would be rude to answer only to cut them off in mid-sentence with "Ah, it's you!" followed by a click as you disconnect. Most of us simply screen our calls and pretend we are not available (an acceptable excuse not to answer).

This works at the international level as well, although in this day and age it is becoming more and more difficult to make yourself scarce. Consider, however, the following example from the height of the Second World War. Bulgaria was ruled by King Boris III, and was allied with Germany. Bulgaria was also home to 50,000 Jews, whom the Germans wanted deported and exterminated like the others throughout the conquered or allied territories. The Bulgarians did not like the idea a bit, and this included the Christian Church and the King. Thus, once the deportation orders arrived from Berlin, the Church organized clandestine evacuations of the Jews from the cities and dispersed them among other friendly Bulgarians throughout the country. When the government forces, delayed on purpose, finally began scouring the cities for the Jews, they did not find any. Bulgarians innocently claimed no knowledge of any Jews living among them. The Germans became outraged and tried to strong-arm the King into pursuing deportation more vigorously, like a real ally. The King, however, was nowhere to be found. He had disappeared in the woods, "hunting," for two weeks until every Jew was safely hidden. "Unfortunately," he was not available to receive the German threats in time, and when

he emerged, he could pursue the policies fully with absolutely no consequences for the Jews. Bulgaria ended up as the only belligerent with a significant Jewish population that saved it from extermination during the Second World War even though Germany exercised serious control over the country's affairs.[6]

4 Manipulating Future Payoffs

Another general way of acquiring credibility is to change your own future payoffs such that what was not in your interest to do, becomes optimal (and therefore credible).

4.1 Reputation

Reputation is a concept often bandied about by policy-makers. As we shall see, much of the American (and Soviet) behavior during the Cold War was driven by reputational concerns: each superpower felt compelled to demonstrate its resolve and superiority to the other and to the audience of uncommitted other states. The fall of one country under communism was interpreted by U.S. policy-makers as a dangerous sign that the Soviets were on the move, but, and perhaps more importantly, that it would seduce others to follow in the wake of the apparently triumphant communism. The idea was to react in a way that would demonstrate to the rest of the world that the Americans were taking things seriously, and that they were prepared to incur significant costs in the defense of their allies or friendly regimes. In other words, the U.S. wanted a reputation for toughness and trustworthiness.

Acquiring reputation is a strategy that allows one to restructure the future payoffs in a way conducive to making commitments credible. For example, it may not be worth the expense for the U.S. to defend Kuwait from Iraq for the sake of the Kuwaitis or West Berlin from the East Germans for the sake of the other Germans. A threat to use costly force for such a purpose can be dismissed as incredible. However, if the U.S. manages to convince Iraq or the USSR that it considers such defense a matter of reputation, it just might work.

It might work because the U.S. would be telling its opponents that it expects grave consequences from the failure to act: not only the (admittedly negligent) loss of the current prize at stake, but future losses resulting from losing the reputation for being a trustworthy ally. Thus, the relevant calculation is not between this loss and the costs of avoiding it, but between these costs and a stream of future losses in addition to the present one. This may well tip over the cost-benefit balance and make it rational to bear large costs today to avoid even larger losses in the future.

For such a tactic to work, the players must care sufficiently about the future, the interaction must be expected to continue for a long period of time, and reputation

[6] Denmark also managed to preserve its 1,000 Jews through slightly different tactics.

must carry over into related areas. These are all pretty difficult to achieve.

4.2 Salami Tactics

Sometimes it may be possible to divide a single large game into a series of smaller steps, none of which carries excessive risk by itself. The idea is to proceed slowly and allow for the reputational mechanism to kick in. As opponents demonstrate with each successive step that they can be trusted not to renege on their promises, their mutual confidence in the successful resolution of each following step increases.

That is one reason you often pay in installments for ongoing projects. This is also why the IMF distributes its huge loans in tranches, and not all at once. The loans have conditionality provisions attached to them that make successive disbursements contingent upon satisfactory implementation of desired macroeconomic policies. A country that receives the entire loan in one lump sum is much less likely to follow painful IMF demands as faithfully as a country whose additional funding depends on meeting such conditions.

Of course, this momentum becomes increasingly difficult to sustain as the end of the game approaches. Here is a very famous example that demonstrates what happens when we carry this to its logical extreme. The game in Figure 3 illustrates the problem. It is a hypothetical description of the Middle East problem: Israel is relinquishing territory in exchange for security from Palestinians.

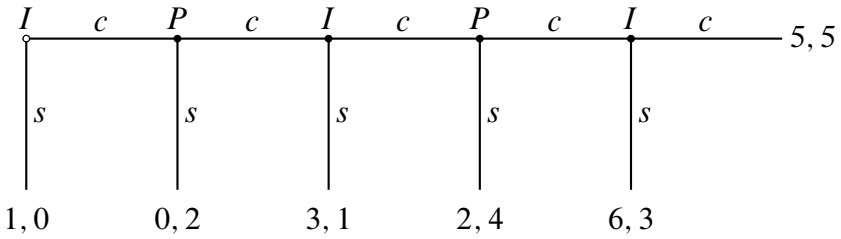

Figure 3: The Land for Security Trade-off Game.

The game begins with Israel in possession of the land. It can choose to stop the peace process (s) or continue it (c). If it continues, it gives up some land and the Palestinians decide whether to stop the process with the land in their possession (in which case Israel is worse off because it gets neither land nor security) or continue it and abandon some of their terrorist activities. If they continue, Israel benefits from reduction in terrorism, and gets to choose again whether to continue or stop. This continues until only one piece of land and very few terrorists remain. This is called the "endgame." At this point, Israel can benefit more from stopping the process and simply capturing the remaining terrorists than conceding the last piece of territory.

Solving this game with backward induction tells us that in the endgame, Israel would prefer to retain the territory and go after the terrorists, so it will choose s.

Given this outcome, the Palestinians would strictly prefer to stop too because they would avoid giving up additional bargaining leverage for which Israel is not going to reciprocate. Stopping at their second node yields 4 which is better than continuing and getting 3 after Israel plays *s* in the endgame.

But since the Palestinians are expected to stop the process at their second information set, Israel will not continue past its own second information set, which in turn makes the Palestinians unwilling to reciprocate even the first concession, which in turn renders the Israelis unwilling to even offer it. The unique perfect equilibrium of this game involves all actors playing *s* at each of their information sets. The equilibrium outcome is that the process does not even get started!

The endgame effect can be very strong and persistent. The above example just demonstrates the extreme case, of course. In reality both sides will be eager to see some progress made because they are unsure about the exact incentives of the opponent. Under these conditions, one would expect them to take a couple of steps forward. But as the endgame approaches, it will become increasingly tempting to preempt the opponent by stopping first. Although it is difficult to say which side will be the first to terminate the process, we can be fairly certain that the process will end before it gets a chance to go to its last part.

While giving up the territory in one fell swoop may be utterly unreasonable from Israel's standpoint, proceeding in smaller steps, while better and more likely to yield some results, will still fall short of ensuring that the process will go through to its conclusion. Generally, the closer the endgame, the more tempted are opponents to preempt each other.

4.3 Irrationality

If I can convince you that I am irrational or stupid and therefore cannot understand your commitment, I render myself immune to your threats and win because you (being the rational and smart one) would have no choice but back down. Children often understand this much better than adults. A kid pretending to be dumb or not hear is simply implementing a pretty good tactic of making himself unavailable to receive information about your very credible commitment that is not in its interest.

This idea of **rational (strategic) irrationality** is not limited to children. President Nixon, for example, once remarked to his National Security Advisor and later Secretary of State Henry Kissinger that it would be good for the Russians and the North Vietnamese to think that he was "out of control" and so could use the nukes if an agreement on peace is not achieved soon. This was an attempt to escape the rational logic that precluded the use of nuclears in such a peripheral theater. It did not work (not that Nixon was entirely sane).

Motives for irrationality that get used frequently with variable success abound. Appealing to honor is a way to claim that you will deliver the action threatened or promised even if you are tempted not to do so. Naturally, one way to undermine

such a strategy is to allow your opponents graceful ways to bow out of commitments. You are, in effect, destroying the grounds for appealing to honor. If no honor was tarnished by the exchange, there is no need to defend it.

5 Manipulating Risk: Brinkmanship

Sometimes, a threat is simply too big to be credible. Two strategies share an underlying logic between themselves. One is the **threat that leaves something to chance** and the other is the strategy of **limited retaliation**. These strategies depend on the willingness of the players to run a **risk of undesired and unintended consequences**.

Imagine a chess game. You are playing the Whites and I am playing the Reds. The game, as usual, can end in win, loss, or a draw. However, we now modify the game by adding a fourth outcome called *disaster*, which is strictly worse *for both players* than simply losing the game. For example, if disaster occurs, we both pay hefty fines to a third party.

The new rules specify very clearly what causes disaster. Specifically, if either player has moved his knight across the middle of the board and the other player moves his queen across the middle, then disaster strikes immediately. It does not matter whether the knight or queen are moved first.

How would two rational players play this game? One thing we can tell for certain is that it will never end in disaster because this outcome is always under control of the players and they both have incentives to avoid it. The disaster outcome can only occur if some player deliberately makes a move that ends the game according to the new rule. Since disaster is the worst possible outcome, no rational player would ever make this move.

This is not to say that the knights and the queens will stay on their side of the board. Indeed, because of this certainty of disaster on the last move, players can use strategic moves that exploit the situation for its inherent credibility. If I, for example, am the first to move his queen across the board and keep it there, you are effectively deterred from moving your knights across. As long as the queen is on that side, I have credibly committed to threatening you with disaster should you move the knights across.

In fact, I am threatening you with something that you would cause should you take the proscribed move. The consequences follow automatically and I am unable to do anything about that. To wit, I am threatening you with a war that you start! As before, disaster is unpalatable to both, and even if it were more costly to me than to you, the threat would still be effective as long as your costs are sufficiently high compared to the other possible outcomes, and so you would still be deterred. I have successfully relinquished the initiative to you, and it is you who gets to be embarrassed by the multitude of choices at your disposal.

The virtue of this modified game is that the rules are completely clear and it is always known with certainty who has committed and who has the last move that avoids disaster or causes it. In real-life, of course, things are not as clear. We don't always know (or can even calculate) who would be the last to move. Certain situations create their own escalatory logic that might blow up in both our faces with neither really intending it.

5.1 The Threat That Leaves Something to Chance

We now modify the modified chess game. We keep disaster outcome and amend the rule to say that should the necessary conditions occur a referee rolls a die and if six comes up disaster occurs. If the die shows any other number, the game continues. If the conditions still exist after a player makes the next move, the die is rolled again, and so on. That is, every time the conditions are met, there is a one-sixth chance of disaster. (In our language, we transform the necessary and sufficient conditions into ones that are only necessary but not sufficient.)

This is now a very different game indeed. In particular we can easily imagine circumstances where knights and queens would move to the "wrong" side of the board, creating a *shared risk of disaster*. If, for example, you move your queen across, I can try to compel you to move it back by deliberately placing both of us in a risky and dangerous situation. I can move my knight across and at every turn while the situation persists we both risk a one-sixth probability that we end up badly. If you lose your nerve before I do, that is, if your willingness to run risks is not as high as mine, I win because you would retreat.

Notice how different this is from before. In the original modification, whoever moved his relevant piece across the board first won. There were no imaginable circumstances where we would both have the queens and the knights one the "wrong" sides of the board. The reason for that, of course, is that the threat is extremely effective: in fact, its fulfilment is completely automated by the rules.

In the modified version of the modified chess game, however, this certainty is gone. What's more interesting, players are able to threaten each other with a disaster that would hurt both. This was not a possibility in the original modification because once someone commits, the other cannot pressure him to retreat by threatening to move his chess piece across too. The certainty of disaster ensures that no such threat can be credible. In this version, on the other hand, such threats can be made and probably will be made.

You can apply the technique of constraining your own choices to this environment as well. For example, suppose you have moved your queen across and I want to compel you to move it back. However, you are much more resolved than I am and we both know it. If I can bring myself to run the risk of disaster at least twice, however, I can win nevertheless: I move my knight across, thereby placing us both in jeopardy. However, since I know that in the war of nerves you will probably win,

I then move another piece such that it blocks the knight's way back. Now I cannot retreat even if I wanted to and it is up to you to do something to relieve the risk. If I can commit myself to continue to run the risks and make clear to you that you are the only one who can diffuse the situation, you would have no choice but back down and retreat.[7]

The strategy of taking your opponent to the brink of shared disaster and compelling him to turn back first. Schelling calls it "manipulating the shared risk of war" and it really involves the deliberate creation of risk that can only be relieved when the opponent takes an action that suits your purposes. Brinkmanship is a war of nerves, it is about risk-acceptance and fear more than it is about cool rational calculations.

Why don't we just threaten with something certain? Why "simply" create a *risk* that something *may* happen? Threatening with too big a stick can be a problem because it may lack credibility. For example, consider the original modification of chess. Suppose you move your queen across and I verbally tell you that unless you retreat I will move my knight and we both end up with the disastrous outcome.

We have already seen that it does not matter whether this outcome hurts you more than it hurts me. As long as it hurts me sufficiently (and it does because according to the rules it is even worse than a loss), my threat will not be credible. You obviously cannot avert the disaster *if* I make the final move. I know it. You know that I know it. And I know that you know that I know it. We also both know that it is up to me to make the fatal last move. You can just sit smugly and smile at me while I rail against the rules being stacked in your favor, the world being cold and heartless, and nobody caring about my predicament. None of that would help, of course. You win and we both know it.

A similar problem occurs with threatening massive retaliation in response to conventional military infractions. The stick is too big and too dangerous to be believable. Even when the United States had first-strike capability many wondered if this nation could use the nukes for a third time with impunity and with total disregard of the extent of the threat they are supposed to diffuse. Say the Soviets invade some

[7] You can also think of a variant with *escalating risks* of disaster. For instance, if the conditions still exist after the move following the first roll, the die is rolled again, and if either six or five comes up, disaster strikes. If the conditions still persist the next time, the die is rolled again, and disaster occurs if six, five, or four comes up. In other words, every next time the conditions for disaster are met, the risk of suffering it increases by one-sixth. Clearly, the sixth time the die is rolled, disaster will strike for sure. This increases pressure on the players to remove themselves from the situation. Of course, the player who has to make the move before the sixth roll essentially faces certain disaster unless he defuses the situation. But, knowing that, he has every incentive to move his piece into a position from which it would be impossible to retreat. If the pre-commitment succeeds, the opponent will be forced to back down even if she would have taken a high risk in the fifth roll, and so on. Again, this becomes a game of preemption: who will maneuver first into a position from which it will not be possible to extricate within the time-frame? As you can guess, if players misjudge the time-frame they think they have or the counter-moves of the opponent, such tactics may make disaster certain.

dinky little third world country with a population of 1 million. Can the United States threaten to blow up Moscow (population of 10 million) in retaliation? Probably not and the Russians knew it. The gun is too powerful and so the threat to use it is not credible.

Recall our hypothetical escalation game. The solution was that for any $p > 0.8$, the defender would never resist. This upper limit is the defender's tolerance for risk: only if $p < 0.8$ would the defender be willing to resist and risk war. Of course, if we put different numbers for the payoffs, we will get different tolerance levels. However, the principle would hold: a threat is "too large to make" if the probability of it going wrong is above this critical limit. In this game, the defender's threat "I will resist if you escalate" is too large, too risky, and too costly to make.

When it is not possible to threaten credibly because the action would hurt you too much, you can threaten with the *risk* or *probability* that the action would be carried out *despite your best intentions to avoid it*. Uncertainty, so the speak, scales down the threat (you will read about this in Schelling's book where he talks about randomized threats).

The risk of carrying out the action in spite of your own attempts to prevent is inherent in many complex situations. First, you may simply make an error in assessing your opponent's freedom of choice and intentions. Maybe the opponent cannot or would not back down. In any case, the risk of misperception is clearly present. Second, and more interestingly, the threat may be carried out even when it should not have been. Maybe your opponent backs down but before you have the chance to stop it, events are set in motion that lead to disaster anyway. Brinkmanship is a slippery slope, maybe at some point it is no longer possible to avert disaster and nobody is quite sure where this point really is. That's the third possibility: we both may become committed to the escalatory steps without even realizing it and may not be able to escape them even if *we both wanted to*.[8]

The threat that leaves something to chance (very aptly named) depends on creating this shared risk of disaster. Once created, the players engage in a competition in risk-taking in the sense that the outcome depends on resolve and nerve.

We now examine two claims often made by analysts and show that their logic has important gaps in it.

1. "A state willing to run the greater risks will prevail."

   Paradoxically, it is not always the side with the most resolve or steely nerves that prevails and succeeds in getting the other one back down. If you think about this a little bit, you will probably remember the signaling game we

[8]If you have not seen the film *Fail-Safe*, I absolutely recommend it. In it, the Americans and the Soviets become committed to escalatory actions that result in disaster with neither side wanting it and both trying to help each other avoid it. What begins as a routine day and a small technical mishap turns into a global disaster. See the original film with Henry Fonda and Walter Mathau not the recent George Clooney remake.

analyzed. The difference in behavior between tough and weak types came from the uncertainty of the defender about which type it was facing. The weak types try to bluff and exploit this uncertainty (and the defender's desire to avoid war). The same can occur with running risks: a challenger may not be as resolved as the defender and *know it for a fact*, but as long as the defender is unsure, he can be exploited by a bluffing strategy, at least up to a point. Thus, contrary to the often asserted conclusion that the state "willing to run the greatest risks will prevail," a state that may be less willing to run risks may still come out victorious in such a confrontation.

2. "An increase in the resolve of the defender should make challengers less likely to escalate."

   The logic seems straightforward: if the defender is more resolute, he is more likely to resist, and thus the risk of disaster is greater. This increased risk means that challengers are less likely to escalate.

   This logic, however, is not quite complete. Again, our signaling game can provide some clues. If the defender is stronger and more likely to resist, then the expected payoff from escalation is lower because the risk of disaster is high. This means that the weak challenger will be less willing to escalate. But this now affects the defender's beliefs. Because the weak challenger is less willing to escalate, upon observing escalation, the defender will believe that it is more likely that its opponent is tough, which reduces the expected payoff from resistance to the defender because it increases the probability of disaster. But this in turn means that the defender is now less likely to resist a challenge, which would increase the expected payoff from escalation to the weak challenger, and the latter would find it more profitable to escalate with higher probability.

   Thus, the usual logic ignores the complicated interactive dynamic when analyzing the consequences of increased resolve for the defender. Interestingly, a player may be *more* instead of *less* likely to escalate the more resolved its opponent is. That's because if it is public knowledge that the opponent is resolved, escalation is a very strong signal about the other player: only resolved types would be willing to do it.

This is how our game theory models can help disentangle the logic of claims that sometimes defies even smart experienced people.

5.2 Coercive Pressure with Limited Retaliation

The other very similar strategy that depends on the generation of risk is the strategy of limited retaliation. Instead of creating a situation where ultimate disaster may strike, one takes a series of small steps (hence the word "limited" in the name

of the strategy) that do two things. First, they increase the probability that the ultimate disastrous event may occur because they generate an additional risk of that happening and further steps presumably escalate that risk. Second, they involve giving the opponent explicit incentives to back down that are unrelated to the risk of disaster.

By destroying methodically but in limited quantities things of value to the opponent, you give him the chance to stop the destruction while he still has something of value left. The problem with the big stick (again) is that if the threat is carried out, the opponent has nothing left to care for. In the strategy of massive retaliation, we destroy the Soviet cities, for example. But if the opponent stands to lose everything, he will fight back as hard as he can, which is not what we want. We only want them to back down.

Suppose that instead of initiating a nuclear war, whether deliberately or by accident, we target Soviet cities but only destroy one. We then tell them that unless they retreat we will destroy another. If they don't retreat, we destroy a second city. And so on and so forth, gradually turning the pressure up, but always letting them back down. The reason such a strategy might work is because despite of the pain, the Soviets are left something they care for: their other cities. It is the threat to destroy these cities, not the pain of having already lost some, that might compel them to back down.

This strategy gradually imposes costs on the opponent but, more importantly, it threatens to impose more costs in the future. A player would be unable to threaten with more costs if it destroys everything his opponent values in one fell swoop. A threat that leaves quite a bit to the adversary is a lot more credible than a massive murderous one. In fact, part of the credibility problem with the massive threat is generated by the consequences of nuclear war. If we threaten with a massive nuclear strike, then the Soviets, with nothing to lose, have incentives to strike back and impose as great costs on us as possible. With a limited strategy, on the other hand, they may be induced not even to retaliate because they are afraid that if they do, they would lose even more.

If you think that this is cold and heartless, you are right. However, Robert McNamara, the U.S. Secretary of Defense during the Kennedy and Johnson administrations made a speech in 1962 in which he proposed this very strategy, the so-called "No-Cities Doctrine". The Russians were very quick to denounce it by claiming that no limited option existed in a nuclear war. Once the missiles start flying all bets are off. The Soviets quite correctly perceived how such a strategy would deny them bargaining power. They had a lot of imprecise missiles with which they can threaten massive strikes but not careful limited retaliation in return. So they did not like it.

The essence of this approach is very similar to the one used by the threat that leaves something to chance. The strategy of limited retaliation also increases the credibility of the threat of future destruction. By exercising the limited option, a

player can demonstrate that its resolve is greater than that of its adversary, just like with the threat that leaves something to chance, where it did so by revealing its willingness to run risks of disaster.

5.3 The Generation of Risk

Obviously, these are very dangerous tactics; *they would not work unless they were dangerous because it is the generation of risk that makes them potentially worthwhile*. How is that risk generated?

Rational opponents would never cross the brink of disaster willingly. However, even rational opponents may do so unwittingly, unintentionally, and by accident or sheer bad luck. The essential idea here is to blur the brink. If you cannot clearly see where it is, you can walk perilously close to it. If you could see it, then you might be tempted to stay away, just to make sure nothing actually tips you over.

So how do we blur the brink? By generating the fear that things may get out of hand. Many have heard of the "fog of war" a situation during tense moments of conflict where communication is uncertain, decision makers are not fully in control of events, accidents happen, and everyone's nerves are so tight that they might snap. Many of the mechanisms that generate risk actually preclude firm control of its escalation or its degree, thereby further enhancing the fear factor. This is sometimes called an **autonomous risk** because it is generated by events beyond one's control.

The crucial point is that you have to arrange things in such a way that neither you nor your opponent knows precisely just where the brink is. If you know, you would definitely never escalate beyond it. If he knows, he can push up to it and you run the risk of giving up because you think it is dangerous while he knows that it is safe. The threat is therefore one of unintended consequences, an inadvertent escalation, not a cool rational one.

6 The Hurt-More Criterion

It is often said that a threat that damages the threatener more than it damages the threatened party cannot be credible. This reflects a rather profound misunderstanding about the considerations that enter the decision to resist the threat or comply with the demands. The credibility of the threat does depend on whether the costs incurred in executing it are prohibitive relative to the pain of not getting what's being demanded. But suppose the threat is credible in that way but still damages the threatener more than it does its opponent. Would the opponent comply? He would if the pain of no compliance (resulting from the threat being executed) exceeds the pain of compliance. Nowhere in this calculation would the pain relative to his opponent appear.

To illustrate this, let's assume that war is two times costlier for the U.S. than it is for the Russians. We modify the crisis game in Figure 2 such that the payoffs to war to reflect this, as shown in Figure 4.

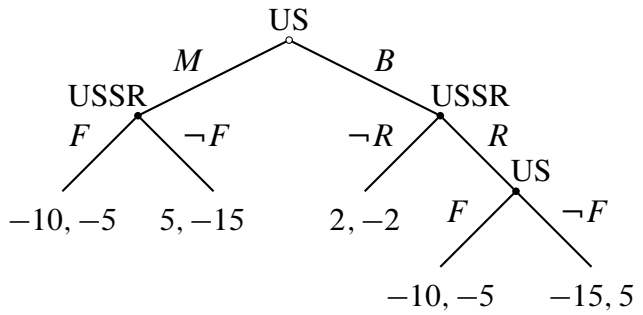

Figure 4: War Hurts the Americans Much More.

We do the backward induction again and we find that our results are completely unchanged. In other words, in this setup, the U.S. still manages to compel the Soviets to back down even though it threatens with a war that would damage it twice as much as it would the Russians. Does this go against your intuition? What's going on here?

It does not matter how much the U.S. hurts itself in war relative to the Soviet Union. What matters is how much the Soviet Union gets hurt *compared to its other choices*. However costly the war is for the U.S., the relevant calculation that the Russians make is the one where they compare *their* costs of backing down versus *their* costs of fighting a war. None of these include the U.S. costs and so it is not surprising that these do not matter in the end. All that matters is that war is sufficiently painful to the Russians given the pain of backing down. If war is more painful, they will back down.

This is not to say that U.S. costs do not matter at all. They do, but only for the calculations of the Americans. The threat to go to war must be credible if the Russians are going to believe it. If war is so costly that even backing down in response to a direct military challenge is preferable, then the U.S. has no viable threat. However, we assumed here that the U.S. would fight if challenged, so this was not a problem.

We conclude that **the threat does not depend on the threatener having to suffer less than the threatened party**. All that matters is that the threatened party would suffer more if it does the action it is being threatened not to do compared to another action. However, we must keep in mind that for the threat to be credible, the threatener must have an incentive to carry out the threat.