National Security Strategy: Revealing and Eliciting Information

Professor Branislav L. Slantchev

January 1, 2014

Overview We now know that asymmetric information can be a serious problem that may lead to war between rational opponents who otherwise like the status quo. We now study some tactics that are designed to deal with these informational problems. We know that asymmetric information can cause conflict when players try to bluff and exploit the uncertainty of the their opponents. We also saw that tough opponents may suffer from inability to separate themselves from weak ones. We now study ways that such credible signaling can be achieved. We also study ways of eliciting information from strategically-savvy opponents.

We have seen the consequences of incomplete information in action. War cannot occur in the subgame perfect equilibria of the escalation game with complete information. If the challenger is weak, he never escalates the crisis because he knows that he will be compelled to capitulate by the resisting defender. If the challenger is tough, then the defender capitulates, and deterrence fails. In any case, war cannot occur.

The situation changed quite dramatically once we introduced the possibility that the defender may be unsure whether the challenger is tough or weak. The pooling sequential equilibrium that exists when the defender believes that the challenger is very likely to be tough is similar to the subgame perfect equilibrium in which the defender always capitulates. The probability of war is (again) zero. However, when the defender is more optimistic, the only semi-separating sequential equilibrium that exists predicts a non-zero chance of war even though all players prefer the status quo to fighting.

The problem is the presence of **asymmetric information**: the challenger knows something that the defender does not. A tough challenger would really like to reveal his type to the defender because this would convince the latter to capitulate, and war would be avoided. However, if the defender believed every statement about toughness, then even weak challengers would make such statements and escalate. Of course, the defender cannot allow just anybody to win such confrontations because she knows that if she resists the weak challenger, she would force him to capitulate without the danger of war.

In the following series of models, we shall take the basic preferences as they were in the games we have already analyzed, and we shall then make appropriate modifications as necessary. To keep things straight, recall that each player (regardless of type) obtains a payoff of 0 from the status quo, 10 from victory (the opponent's capitulation), and -10 from defeat (his own capitulation). The difference is in the payoffs from war, which vary depending on the player's own type and the type of his opponent. We have already seen the cases in which the challenger faces a weak defender. The cases in which a challenger faces a tough defender will be symmetric.

	D_W	D_T
C_W	$-12, -12$	$-15, -1$
C_T	$-1, -15$	$-5, -5$

Figure 1: The Type-Contingent War Payoffs.

The payoff matrix in Figure 1 shows the war payoff for each type of challenger matched against each type of defender (and vice versa). The first number is the challenger's payoff, and the second number is the defender's payoff. For instance, the weak challenger will get a war payoff of -12 if his opponent is weak and -15 if she is tough. Analogously, a tough defender will get a war payoff of -1 if her opponent is weak and -5 if he is tough.

1 Costly Signaling and Signal Jamming

The potential for weak challengers to exploit the uncertainty causes the defender to risk war in order to deter them, producing exactly the outcome a tough challenger would wish to avoid. In such a situation, it is in the interest of both parties that tough challengers identify themselves beyond doubt; that is, that they separate themselves from weak ones. How can one **signal** such information credibly? That is, how can one reveal such knowledge and be believed? Talk is cheap.

It should be clear from the model that the only way to do this is by doing something that a weak challenger would not do.[1] Escalation carries the risk of war which is less painful to the tough type than the weak type. This is precisely why the weak type does not simply mimic the behavior of the tough one: it is too dangerous to do so. If the tough challenger could do something that the weak one would under no circumstances want to do, he could convince the defender of his strength.

As we have seen, escalation by itself is not quite sufficient for clean revelation (recall that the game does not have any separating equilibria) because the weak challenger's strategy is designed precisely to confuse such distinctions. A strategy whose goal is to make inferences imprecise is called **signal jamming**, and typically involves a mixed strategy.

1.1 Sinking Costs

To overcome the consequences of signal jamming, the tough challenger must undertake an action that the weak one would not want to mimic with positive probability even if doing so yields the best outcome. Think about it this way. Suppose the tough type did find a way to convince the defender of his strength by taking an action c. In this case, the defender is guaranteed to capitulate. This, as we know, is the best possible outcome for the challenger. The action c must be such, however, that the weak type would not want to take it *even if doing so would produce capitulation by the defender.* Otherwise, as we have already seen in the escalation example, there will be no separating equilibria.

The only way to convey information credibly in such instances is through an action that is simply too costly or too dangerous for the weak type to take. To see how this works, suppose escalating is costly: the challenger must marshal resources, probably put the forces on full alert, perhaps begin mobilization of troops, all of which are extremely costly activities by themselves and by the disruption they cause

[1] Or could not do. For example, an unambiguous revelation of capability. Israel paint-bombed the Aswan Dam during the Yom Kipur War, which unambiguously revealed capability to destroy it (which would have resulted in flooding the entire Nile valley with its millions of civilians). The signal is unambiguous because it simply cannot be imitated by an air-force that cannot actually destroy the dam. This, however, is quite rare: nations are usually loath to reveal their capabilities because alerting the opponent to them without actually using them in war simply gives the opponent a chance to come up with a counter-strategy.

in civilian life, economic production, transportation, and so on. Let the cost of escalation be 5 units for the tough type, and 11 units for the weak type. Figure 2 shows the modified game, where the costs of escalation have been subtracted from all payoffs of the challenger except the status quo.

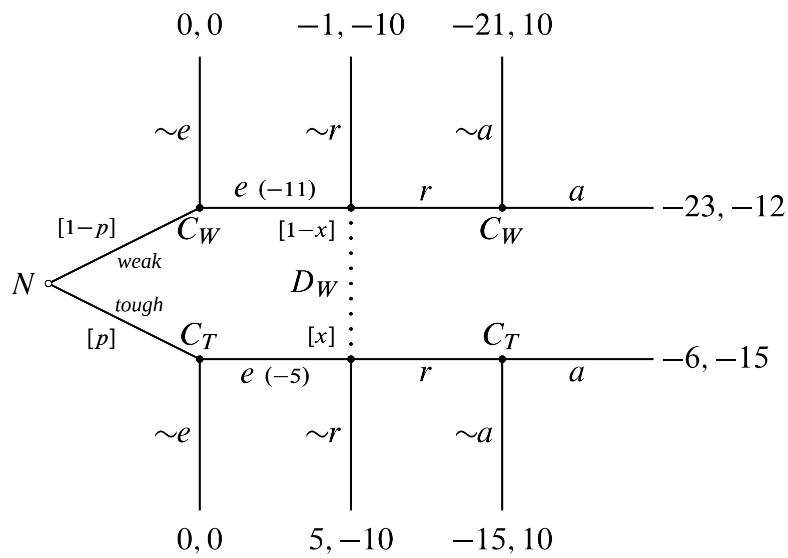

Figure 2: Costly Escalation Game with Incomplete Information.

This game is much easier to solve than the previous version. We begin with the backward induction. As before, the weak challenger never attacks at his last node because the payoff from a is -23, which is worse than the payoff from $\sim a$, which is -21. The tough challenger, on the other hand, always attacks at his last node because the payoff from a is -6 which is strictly better than the payoff from $\sim a$, which is -15. Thus, introducing the costs of escalation has not changed the basic distinction between the two types. Tough challengers still go to war, and weak ones still capitulate.

Figure 3 shows the simplified game tree of this situation. The two last information sets for the challenger have been replaced with the payoffs for the corresponding outcome that would result when each type takes his optimal action at his information set.

The optimal action for the defender depends on whether she believes her opponent is tough or weak. If the defender is sure that the challenger is tough ($x = 1$), then she would not resist because $\sim r$ yields a payoff of -10, while r yields a payoff of -15. If, on the other hand, the defender is sure that the challenger is weak ($x = 0$), then she would choose to resist because playing r yields a payoff of 10, while $\sim r$ yields a payoff of -10. Thus, just like in the previous game, the sequentially rational strategy for the defender critically depends on her belief about the type of opponent she is facing.

However, this is where the similarities between the two situations end. Consider

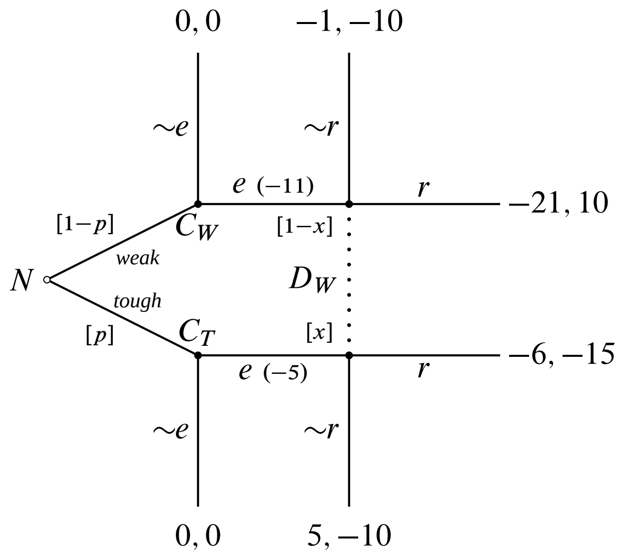

Figure 3: Simplified Costly Escalation Game with Incomplete Information.

the choice to escalate by the weak challenger. If he chooses to escalate, then he would get -21 if the defender resists, and -1 if the defender does not resist. However, if the weak challenger chooses not to escalate, he would get the status quo payoff of 0. But this payoff is strictly better than anything the weak challenger could ever hope to get by escalating regardless of the defender's response. Thus, it can never be rational for the weak challenger to escalate: even if the defender believed that the challenger was tough and did not resist, the weak challenger would still do worse by escalating.

The action $\sim e$ **strictly dominates** the action e for the weak type because it yields an expected payoff that is strictly larger regardless of the actions of the defender. Consequently, e can never be used in equilibrium because it is never a best response to anything the other player might do. Hence, in any sequential equilibrium, the weak type would play $\sim e$ for sure.

Consider now the tough type's decision. If he chooses $\sim e$, he would get zero from the status quo. If he chooses e, then, because the weak type always chooses $\sim e$, the defender will update to $x = 1$; that is, the defender will become convinced that his opponent is tough. The best response to this belief is to capitulate, and so the tough challenger would get a payoff of 5, which is strictly better than the payoff from not escalating. Hence, escalating is optimal for the tough type.

We conclude that the profile $\langle(1, 0), 0\rangle$ is the unique **separating** sequential equilibrium of this game.[2] The tough challenger escalates and fights if resisted, the

[2]We could still obtain the pooling equilibrium in which no type escalates but this would still require the strange off-the-path belief that D infers that her opponent is weak upon seeing escalation. In this instance, this belief is even less plausible because for the weak challenger escalation is strictly dominated, and as such worse than non-escalation regardless of D's beliefs.

weak challenger never escalates and backs down if resisted, and the defender capitulates when challenged. The outcome is that weak challengers are always deterred, and tough ones are never deterred. However, the probability of war is zero in equilibrium.

Why is this different result? As we have seen, the strategy of the weak and tough challenger at the last stage of the game remains the same, and so does the defender's strategy. However, for some reason, the weak challengers do not attempt signal jamming, and consequently the tough ones can reveal themselves fully by escalating. The reason this could happen in this game but not in the previous one has to do with the relative costs of escalation for weak and tough types. Although escalation is costly for both, it is far costlier for the weak type. In fact, it is so costly for him, that he would rather live with the status quo than escalate even if the defender were sure to capitulate.

Clearly then, incomplete information is not a sufficient condition for war. The problems it produces can be overcome by **costly signaling**. Signaling is credible when it is costly precisely because it is only in the interest of the tough type to engage in it. Weak types can never benefit from mimicry of the tough types, and so they reveal themselves by their silence. We have, therefore, found how a player can reveal private information in a credible way: by engaging in an action that is excessively costly for others to imitate.[3]

It is worth asking what would happen if the signal is costlier to the weak challenger than the tough one but not as costly as we assumed above. In particular, what if it does not make non-escalation a strictly dominant strategy. To see that the benefit of costly signaling will persist in the scenario, let's analyze the game assuming that the cost is still 5 for the tough type but is now 7 for the weak type. Figure 4 shows the modified game. It is the same as Figure 2 except that we have now subtracted 7 instead of 11 from all payoffs of the weak challenger following escalation.

Again, the challenger will fight if weak and will capitulate if tough when resisted, which gives us the simplified version of the game in Figure 5. The defender will

[3] You have seen many examples of costly signaling in your life, even if you did not realize it. Consider the quality of new cars. You are likely to have far less information about the quality of a car you are thinking of buying than its manufacturer. It is in the manufacturer's best interest to convince you that his car is good and charge more for it, but he has a credibility problem because every manufacturer would make such claims. The manufacturer could then offer you a warranty as a signal. If the car is really good, it is unlikely to break down in 10 years, but if it is a lemon, it is very likely to break down in half that time. The manufacturer of the good car can offer you a 10-year warranty because he knows that he is highly unlikely to have to make any repairs during that period. The manufacturer of the lemon finds such a warranty too costly because he knows that he will be forced to honor it frequently as your car develops its problems. Therefore, he will only offer you a 2-year warranty. Thus, the length of the warranty is a costly signal of the confidence the manufacturer has in his car, and it is credible.

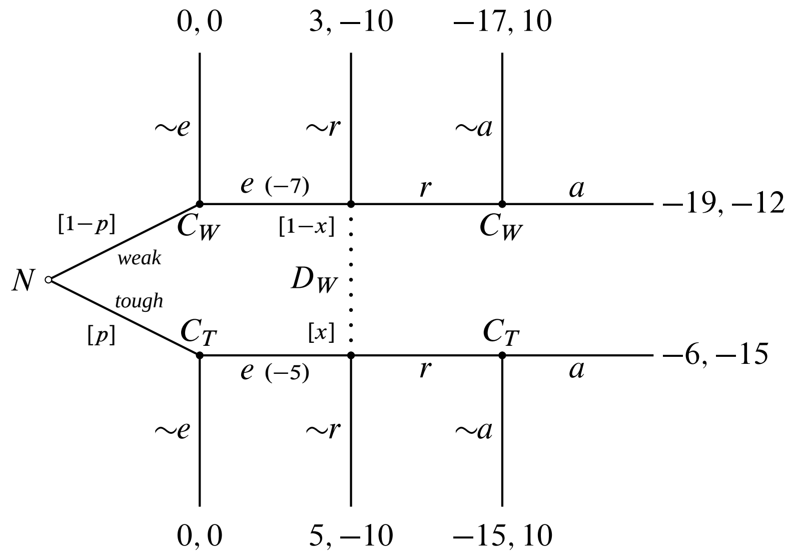

Figure 4: Less Costly Escalation Game with Incomplete Information.

prefer to resist if there's a high enough probability that the challenger is weak:

$$\mathrm{U}_D(r) > \mathrm{U}_D(\sim r) \Leftrightarrow x(-15) + (1-x)(10) > -10 \Leftrightarrow 20 > 25x \Leftrightarrow x < 0.80,$$

that is we have precisely the same best-response for D as before. Unfortunately, not escalating is no longer strictly dominant for the challenger because now it is possible to profit relative to the status quo if the defender should capitulate (this gives the weak type a payoff of 3). Hence, the weak type will escalate if

$$\mathrm{U}_{C_W}(e) > \mathrm{U}_{C_W}(\sim e) \Leftrightarrow q(-17) + (1-q)(3) > 0 \Leftrightarrow 3 > 20q \Leftrightarrow q < {}^3\!/_{20}.$$

where q is the probability that D will resist.

This game will have the pooling equilibrium with escalation and the semi-separating equilibrium with bluffing. So, let's see when pooling can happen. Since both types escalating for sure, the defender learns nothing new from this act, so $x = p$. He will therefore capitulate for sure if $p > 0.80$, which in turn rationalizes the challenger's strategy. The game with sunk costs does not differ from the basic escalation game in that respect. Consider now the equilibrium in which the challenger always escalates when tough and escalates with probability β when weak. Because the weak type is willing to randomize, he must be indifferent between his pure strategies. This implies that $q = {}^3\!/_{20}$ or else he would strictly prefer one of them. This, in turn, means that the defender must be randomizing in equilibrium as well, and because this is only possible when $x = 0.80$, the challenger's mixing must induce this posterior belief. By Bayes rule, the posterior belief is defined as:

$$x = \frac{p(1)}{p(1) + (1-p)\beta} = 0.80 \Rightarrow \beta^* = \frac{p}{4(1-p)},$$

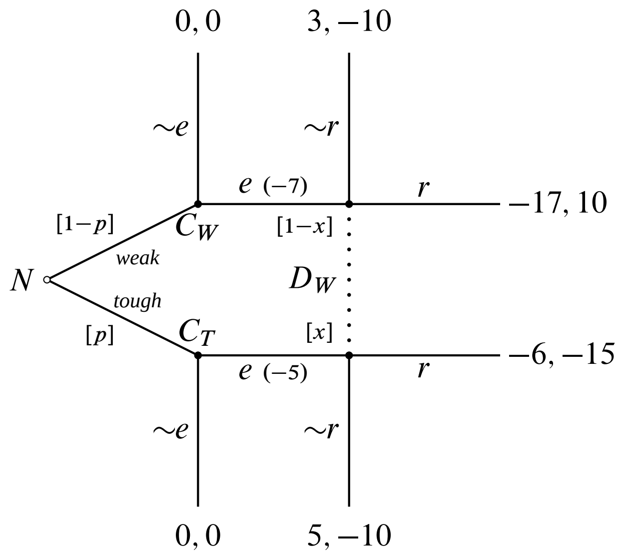

Figure 5: Simplified Less Costly Escalation Game with Incomplete Information.

that is, this is the same as before. Sinking costs has had absolutely no effect on equilibrium behavior of the challenger compared to the original game. However, it has had an effect on the behavior of the defender. Observe now that the probability that he will resist a challenge is $q = {}^3\!/_{20}$, which is strictly *less* than the corresponding probability in the original game, where we found it to be ${}^1\!/_2$. This implies that the probability of war in this game is also *lower*. In fact, it is now:

$$\Pr(\text{War}) = p(1)({}^3\!/_{20}) = ({}^3\!/_{20})p,$$

that is, it is about *three times* lower at any given value of p. Why is the defender less likely to resist in the new game even though the weak challenger is just as likely to bluff as before? This appears a really odd result: the action of the challenger has *not* affected the defender's beliefs at all relative to the original game. Think about what costly escalation does to the weak challenger though: because the expected payoff from escalation is now strictly worse than in the original game (it is that payoff minus the cost of escalation), it follows that the only reason a weak challenger might escalate in equilibrium is when the probability that the defender resists such escalation is correspondingly lower or else the overall payoff will be strictly worse. To see that, observe that if the defender resisted with probability ${}^1\!/_2$, then escalation would yield the weak challenger a payoff of ${}^1\!/_2(-17) + {}^1\!/_2(3) = -7$, which is strictly worse than staying with the status quo. With costly signaling, it must be the case, therefore, that the defender is less likely to resist.

The crucial point here, however, is that *even though the defender is less likely to resist, the weak challenger is not more likely to bluff.* This is now the real effect of the costly signal: it deters bluffs in situations where the weak challenger would have escalated in the original game for sure: any $q \in ({}^3\!/_{20}, {}^1\!/_2)$. Why is the weak type unwilling to bluff more often despite a chance of victory that is nearly twice as high as before? Because obtaining it is still costly. Conversely, because the weak type is not going to react to a lower probability of resistance by exploiting it with increased bluffs, the defender can "relax" and resist less often. The benefit to the defender here is that doing so is less likely to cause war should the opponent turn out to be committed. The upshot of all this is that *the tough type benefits substantially* relative to the original game, also because the defender is less likely to resist. Hence, costly signaling does benefit the tough challenger and it does reduce the probability of war even when it does not fully reveal the information.

Notice that in both cases we analyzed here, costly escalation does not change he actual commitment of the challenger: the weak type will still prefer to capitulate, and the tough type will still prefer to fight when resisted. Intuitively, this is so because the cost of escalation must be paid regardless of the outcome of the crisis, and as such represents a type of *sunk cost:* once paid, these costs should have no bearing on subsequent behavior of a rational player. Whatever he does at this point will not undo the payment of these costs, and as such they will not influence his actions. What matters here, however, is the willingness to pay these costs: it is precisely this willingness that can help the tough type reveal information about its commitment. Yes, it may be imperfect, as we saw in the second example, but it is still better than nothing. Unfortunately, this tactic does nothing to improve the lot of the weak challenger: his lack of resolve is still a critical problem. We now turn to tactics that the challenger can use that can even benefit him if he is weak.

1.2 Tying Hands

Recall now that in the example in Figure 2, the costs of escalation are paid regardless of the outcome of the crisis. We call these *sunk costs* and note that by themselves they do not alter the credibility of the challenger's threat to fight: the tough one still prefers war to capitulation and the weak one still prefers capitulation to war. The crucial point is that the costly action changes the ranking of the victory and status quo payoffs for the weak type and makes victory so costly that it is unattractive relative to the status quo even if assured. Since only a tough challenger can benefit from such a costly victory, escalation becomes a credible signal of his type.

Observe that the crucial problem with the credibility of the threat to fight remains intact. We now study a way in which a challenger can rearrange his own incentives and change that. In other words, we now study how a weak challenger can *establish a credible commitment to go to war*. There are basically two ways he can do this: he can either eliminate his choice of going to war altogether ("burn the bridge") or he can rearrange his incentives by manipulating his payoffs from war and peace. The general rubric for these tactics is "tying hands" in the sense that the challenger can establish a credible commitment to fight by tying his hands; that is, making it impossible or exceedingly difficult to capitulate after being resisted. The tough challenger does not need to tie his hands because his threat to fight is credible already. We shall talk about the burning-bridges tactic in more detail next time. Right now, we focus on manipulating one's own incentives to fight.

The fundamental problem with the weak challenger's threat to fight when resisted is that his payoff from peace exceeds the payoff from war. There are at least two ways he may try to change that. He can reduce the payoff from peace until it becomes worse than the payoff from war or he can increase the payoff from war until it becomes better than the payoff from peace. Either way, the relative ranking of the two outcomes will change, making the threat to fight credible in the process.

1.2.1 Audience Costs

Consider the first tactic: making peace through capitulation worse. One way a leader can do that is by making a public promise not to back down when resisted. This puts the nation's honor and credibility on the line. If he capitulates after having publicly committed not to, then the nation will suffer diplomatic humiliation, the hollowness of its promises and threats would be exposed, her allies would be alienated, and her enemies would be encouraged to probe for further weaknesses. In other words, the nation would have to pay costs *over and above* the costs of relinquishing the good in this particular crisis.

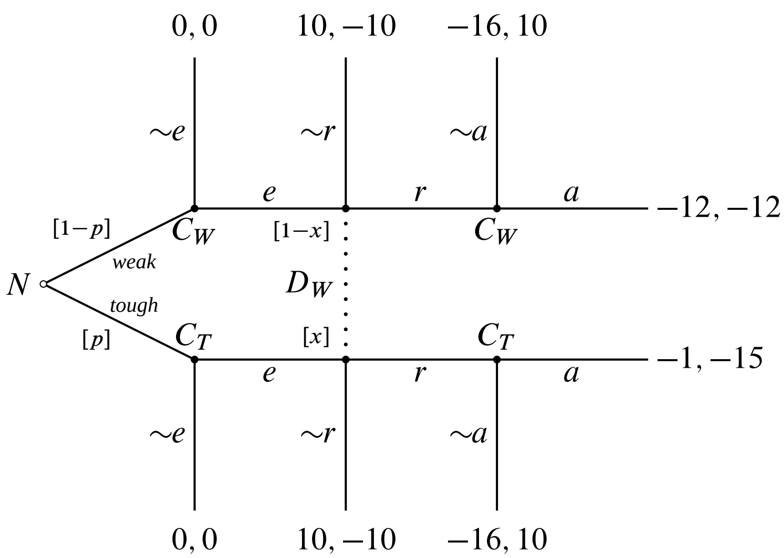

Figure 6: Escalation with Audience Costs.

There are two types of costs here, depending on the audience that imposes them. First, there's the international audience that consists of other states, international organizations, and perhaps non-state actors (e.g., insurgent or terrorist groups). If your allies decide you are weak, they may abandon you. If your enemies decide you are weak, they may bully you. The *loss of reputation* can be a major cost imposed by the international audience.

Second, there's the domestic audience that consists of one's own citizenry or powerful elites. In a democracy, the fate of the leader is decided in elections (the electorate). In an autocracy, the fate of the leader is decided by a much smaller group of important people, e.g., military generals or heads of secret police (the selectorate). In either case, if the leader displeases enough members of the groups on which he depends for retaining power, he will be removed from office and may even suffer additional penalties (autocrats often forfeit their lives with their office). Since capitulation is likely to be perceived as detrimental to the national interest by these domestic audiences, the leader will have to suffer these *domestic political audience costs*.

Whether audience costs are domestic or international, the challenger only pays them if he fails to follow through on his public threat to resist. If these costs are sufficiently high, they may make capitulation so unpalatable that even a weak challenger will prefer to go to war rather than suffer them. Consider our original escalation game and suppose that escalation incurs audience costs of -6. The capitulation payoff is now -16 for the challenger regardless of type, as shown in Figure 6.

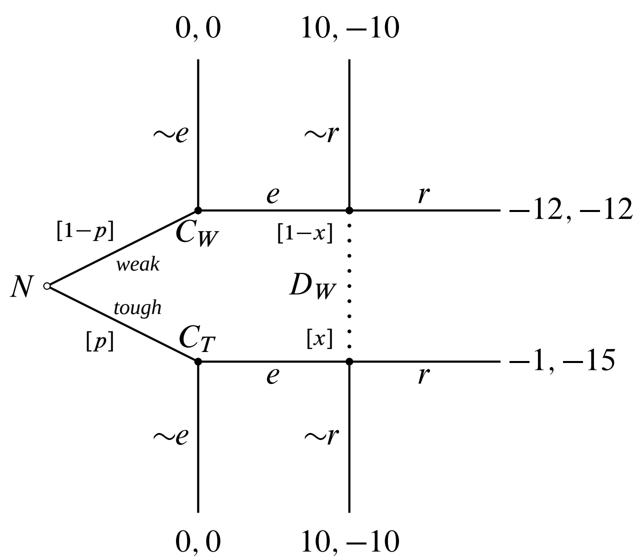

Figure 7: The Pruned Escalation Game with Audience Costs.

Observe that audience costs do not change the credibility of the tough type's commitment: war is still preferable to capitulation ($-1 > -16$). However, they do change the commitment of the weak type for war is now preferable to capitulation for him as well ($-12 > -16$). Subgame perfection requires that the challenger now attack regardless of type, yielding the simplified game tree in Figure 7

The defender's choice now becomes very simple: since resistance will lead to certain war (and a payoff of -12 at best and -15 at worst), she will opt for capitulation (a payoff of -10) regardless of her belief x. Therefore, the game has a unique pooling equilibrium in which both types of challenger escalate and the defender backs down (the crisis is peacefully resolved). Whereas the strategic dynamic is the same as in the pooling equilibrium of the original game, this one does not depend on the prior belief. Recall that in the original setup, the equilibrium only worked if the defender was sufficiently pessimistic ($p > 0.8$). This equilibrium works no matter her prior beliefs are. This is because the public commitment ties the weak type's hands and he becomes certain to go to war. Since the tough type is certain to do that to begin with, from the defender's perspective it really does not matter whether he is weak or tough anymore: resistance leads to certain war. The rational response is to back down even if she is positive the challenger is weak. In this way, incurring domestic audience costs can *create a credible commitment* even where no such commitment existed before.

Many people have argued that in democracies, with their regular elections, the citizens can keep the leaders on a tighter leash than in autocracies. Thus, the logic goes, democratic leaders can incur higher audience costs than autocrats. This makes public commitments by democratic leaders much more credible: when a democratic leader says the country is committed to a war, the statement will presumably carry more weight compared to one made by an autocrat. Since democracies are better able to generate these audience costs, they can signal their intent better than autocracies. Democracies will therefore tend to bluff less and fight more often when resisted. In particular, crises between two democracies should be extremely unlikely to end in war because both sides can credibly reveal their type. Hence, this is one explanation for the *democratic peace*: the interesting, if relatively recent, empirical phenomenon that democracies seldom, if ever, go to war with one another.

Before we rush to the conclusion that domestic audience costs are a good signaling device (or accept the rationale for the democratic peace), I should note four things. First, the mechanism requires audience costs that make peace worse than war. Given the enormous destructiveness of war, this assumption may be very hard to stomach. While it is true that perhaps the leaders (who do not normally go to fight themselves or often don't even send their children to fight in the wars they start) do not suffer the same costs of war as the citizens do, it is still a demanding assumption.

Second, the challenger must tie his hands in a publicly observable way that the defender can see and understand. If this were not the case, then from the defender's perspective the weak challenger would still be expected to capitulate, so she will resist (depending on her beliefs) causing war by mistake. But if that's the case, the weak challenger would not want to tie his hands: it would be useless. While the public statement of the commitment is often deemed sufficient for the defender to hear and understand it, it's not at all certain that an autocrat who is accustomed to one type of political system can readily believe that a democratic leader is subject to a very different set of constraints. In other words, an autocrat may disbelieve a totally sincere and real commitment anyway. This means that a democratic leader who ties his hands may suddenly find himself at war anyway because the opponent simply ignored the public commitment.

Third, if public threats engage one's honor, they may also engage the opponent's honor. In other words, by publicly threatening to fight if the opponent does not capitulate, one may very well back the opponent into a corner from which she cannot retreat without losing face herself. Such a public threat can backfire badly for it can then create a stand-off in which the two sides would rather fight than negotiate. This is an undesirable outcome, especially when they may have both entered the crisis preferring to capitulate (in private). This lock-in can be a dangerous consequence of attempts to establish a credible commitment.

Fourth, incurring audience costs may not always work. Suppose, for example, that the defender is tough (she prefers war to capitulation) and that this is common knowledge. (The war payoffs are listed in Figure 1.) This situation is depicted in Figure 8.

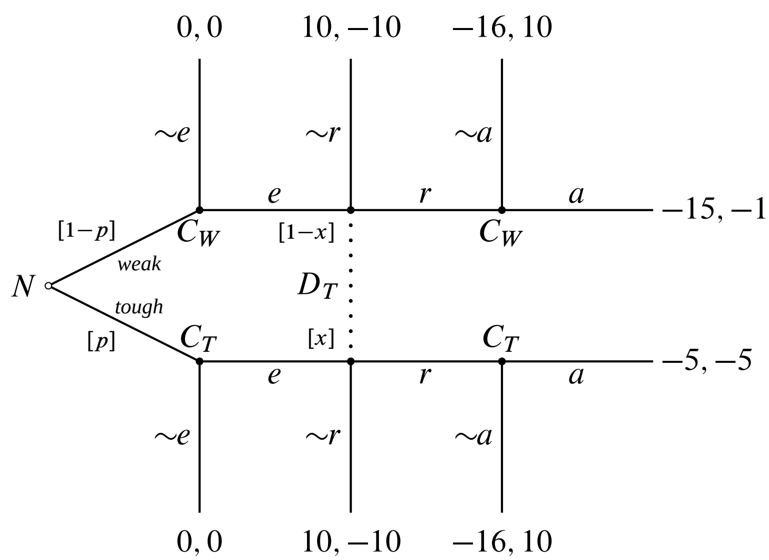

Figure 8: Escalation with Audience Costs and a Tough Defender.

We begin with the challenger's last decision to attack. Since escalation incurs audience costs, capitulating in the face of resistance is unattractive and the challenger will attack regardless of type (audience costs have committed even the weak challenger to a war against a tough defender). Subgame perfection then allows us to simplify the game as shown in Figure 9.

Turning now to the defender's decision, observe that if she resists, her payoff is always greater than what she obtains from capitulating. If the opponent is tough, she gets $-5 > -10$ and if he is weak, she gets $-1 > -10$. This means that resistance is strictly dominant and is therefore sequentially rational regardless of her belief x. This implies that in every equilibrium the only possible response by the defender

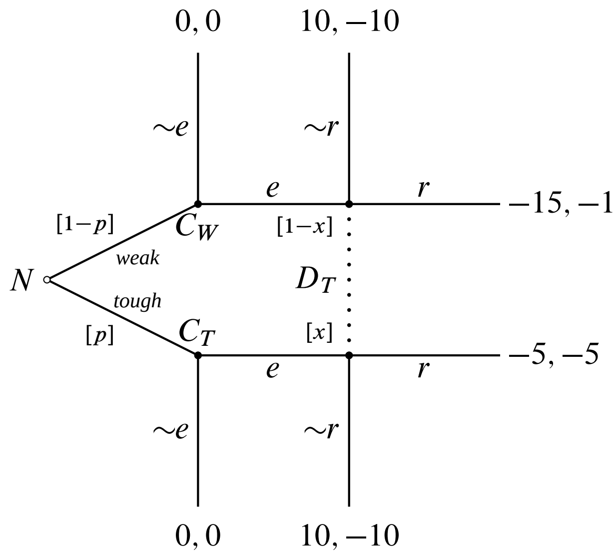

Figure 9: The Pruned Escalation Game with Audience Costs and a Tough Defender.

is to resist with certainty. Knowing this the challenger will never escalate because even the tough type prefers the status quo to war. To see this, note simply that the status quo payoff for the challenger is 0 and the war payoff is -5 for the tough type and -15 for the weak type. Hence, deterrence will succeed even though the challenger could commit credibly to war. This is so despite the fact that the defender herself prefers the status quo to war. We are essentially back to the case where she can threaten the challenger with a war that he must initiate (the audience costs he will incur from escalating ensure that even the weak type will fight if resisted). Tying hands cannot help the challenger here.

We now have an interesting situation: audience costs the challenger incurs may compel the defender to capitulate if she is weak but may simply lead to war (causing the challenger to be deterred from escalating) if she is strong. In other words, we are "back" to our peaceful subgame-perfect equilibrium outcomes under complete information: we either have assured victory for the defender or assured deterrence. Of course, the underlying dynamic of the interaction is slightly different because here we vary the type of the defender as well. Still, you can probably already see what will happen if the challenger was unsure about the defender's type. In that case he would not know whether tying his hands would cause the defender to back down or to fight. Hence, escalation may either lead to war or to victory… meaning that the challenger may sometimes attempt escalation causing war in the process.

To elaborate on this a bit, our simple model assumes the defender's type is common knowledge (that is, the challenger knows it, the defender knows that the challenger knows it, and so on). This means that if the challenger tied his hands, the defender will definitely capitulate if weak. This makes tying hands a safe tactic for it can never lead to war. However, suppose the challenger is uncertain about the defender's preferences. What if the defender is strong? Well, in that case tying hands becomes a risky proposition. It can commit the challenger to war but if the defender prefers war to capitulation, then she will resist anyway, and war will be the outcome. Of course, with complete information about a tough defender, the challenger will simply avoid tying his hands and will not escalate at all. The problem occurs when the challenger does not know for sure if he's facing a tough defender. Then, thinking that she might be weak, he may tie his hands to ensure her capitulation. If the defender turns out to be tough, however, this commitment will lead to war. Showing this requires modeling two-sided incomplete information so there's uncertainty on both sides.

Suppose, for example, that the challenger thinks the defender is weak with probability $^2/_3$ and strong with probability $^1/_3$. This belief is common knowledge. As before, escalation incurs audience costs that the challenger must pay if he capitulates in the face of resistance, so he will attack regardless of type if the defender resists. This means that capitulation is strictly dominant for the weak defender and resistance is strictly dominant for the tough defender. In other words, the sequentially rational strategy for the defender is to capitulate if weak and resist (causing war) if strong regardless of what she believes about the challenger's type. From the challenger's perspective, this implies that escalation will lead to victory with probability $^2/_3$ and to war with probability $^1/_3$. What will he do? We must calculate the expected payoff from escalation. It is:

$$
\begin{aligned}
\mathrm{U}_{C_T}(e) &= (^2/_3)(10) + (^1/_3)(-5) = {}^{15}/_3 = 5 > 0 \\
\mathrm{U}_{C_W}(e) &= (^2/_3)(10) + (^1/_3)(-15) = {}^5/_3 > 0.
\end{aligned}
$$

In other words, the tough challenger expects a payoff of 5 from escalation, and the weak challenger expects a payoff of $^5/_3$. In either case, this is better than the status quo, so the challenger will escalate regardless of type. Hence, the unique equilibrium here is pooling for the challenger (he escalates and then fights if resisted regardless of type) and separating for the defender (she capitulates if weak and resists if tough regardless of her beliefs).

In this equilibrium, war occurs with positive probability because whenever the defender is tough, escalation locks in the challenger into starting a war. Notice that this is *independent of the prior belief the defender has*.[4] The probability of war is, as before,

$$\Pr(\text{War}) = \Pr(e) \times \Pr(r) \times \Pr(a) = 1 \times {}^1/_3 \times 1 = {}^1/_3.$$

[4]For example, suppose the defender thinks the challenger is strong with probability $p = {}^9/_{10}$ and weak with probability $^1/_{10}$. Recall that in this region, when there is one-sided incomplete information and the defender is weak, the unique equilibrium also involves certain escalation but it also involves certain capitulation by the defender. This means the crisis will certainly be peacefully resolved. In the two-sided incomplete information scenario, however, this is not the case.

That is, the probability that the crisis will end in war equals the probability that the challenger escalates (since both types do that, it is certain) multiplied by the probability that the defender resists (since only the tough type does it, it is $1/3$) multiplied by the probability that the challenger attacks (since both types do that, it is certain as well). The overall result is that the chance of war is $1/3$. Note that it did not matter what the defender thought about the challenger's type. (In more complex situations, it will matter.)

One very unpleasant implication of this analysis is that the weak challenger's own behavior in the crisis may commit him to war against a tough defender. In retrospect, this behavior appears irrational: why would a player who would rather capitulate than fight (and loves the status quo) would initiate a crisis and then find himself at war with the toughest opponent possible by the time the crisis ends? The logic here shows why this puzzle should not lead one to conclude that the player was irrational. Even an unresolved player could compel a weak opponent to give up the disputed benefit short of war if he could just commit himself to fighting somehow. This commitment carries certain risks because it will not induce the tough opponent to concede. However, if the player thinks there's a good chance that his opponent is weak, this risk of war will not be enough to dissuade him from attempting to coerce her. Hence, even a weak player can deliberately court the risk of war. In this instance this is "war with regret": when the fighting begins, this player would dearly wish he had not escalated and committed himself to fighting. But now it's too late to back down. Looking back at instances like this, we (as analysts) cannot blame the actor for being stupid and making a mistake: instead, his behavior should be seen as the result of a deliberate risk-taking. And risk-taking, of course, can sometimes backfire. This does not mean that the original decision was wrong. In fact, quite the contrary—it was the best the actor could do under the circumstances (which is why it happens in equilibrium).

Tying hands with audience costs is, of course, a tactic that is potentially available to both sides in the dispute. You can probably see what will happen if both players attempted to commit themselves to war in an effort to compel the other side to capitulate. When they do this under uncertainty, they may create a situation in which they become locked into positions from which neither wants to retreat, and war becomes certain upon escalation (in this scenario two players, both of whom initially prefer capitulation to war, can find themselves at war with each other). For that reason, tying hands tends to be a very risky tactic and is not something leaders often resort to.

All these problems suggest that perhaps public commitments that depend on domestic audience costs may ether not work at all or be a cure that's worse than the disease: we have seen at least two ways in which attempts to incur these costs may actually cause war. With two-sided incomplete information and audience costs on both sides, it is actually quite possible for two weak opponents to lock themselves into war while trying to maneuver each other into capitulating.

1.2.2 Military Mobilization

The two crucial parts of the audience cost signaling mechanism are (a) escalation decreases the payoff from peace until war becomes preferable even for a weak challenger, and (b) the costs do not change the defender's own commitment. As we have seen, this may not work against a defender known to be strong, or it may cause war against a defender whose strength is not known. Recall now that the fundamental reason audience costs worked as a credible signal of strength is because they commit even the weak challenger to fight when resisted. In other words, they alter his payoff from capitulating under duress relative to the payoff from war. You should immediately see that there is another way the weak challenger can tie his hands: if he could *increase* the expected payoff from war enough to make war better than capitulating, he would have a credible commitment to fight too.

One way a challenger can increase his war payoff is by preparing to fight. If he mobilizes his army and then attacks, he is more likely to prevail than if he started a war without extensive preparations. Notice that if the challenger becomes more likely to win, the defender must be becoming more likely to lose. In other words, the challenger's military preparations *reduce the defender's payoff from war*. This is the first time we have seen a tactic that alters the opponent's payoffs as well. To summarize, the military instrument has three functions that are not available with any of the other signaling/commitment mechanisms we have seen:

1) it is costly regardless of the outcome of the crisis: sunk cost;

2) it improves one's payoff from war: device for establishing commitments;

3) it worsens the opponent's payoff from war: device for undermining commitments.

Let us now represent these three functions by modifying the payoffs in our model appropriately. Since mobilization is a sunk cost, let's suppose, as in the costly escalation game, that when he mobilizes, the challenger pays -5 if he is tough and -7 if he is weak. Clearly, by itself this will not alter his resolve: if he is tough, capitulation would yield -15 and war -10, and if he is weak, capitulation would yield -17 and war -22. In other words, the weak type would still capitulate and the tough type would still fight. (Because this cost is sunk, it will reduce the victory payoff by mobilization costs for both types as well.) Since mobilization improves the challenger's war payoff, suppose that the added benefit from these preparations is 10 if war occurs for each type. This means that the war payoff is now -12 for the weak type, and 0 for the tough type. This does alter the resolve of the weak type given the capitulation payoff of -17: by mobilizing he has committed himself credibly to war. Finally, since war preparations affect adversely the defender's payoff, suppose they reduced it by 10, so a tough defender now expects -11 from war against a weak challenger and −15 against a tough one. Figure 10 shows the game with a military threat.

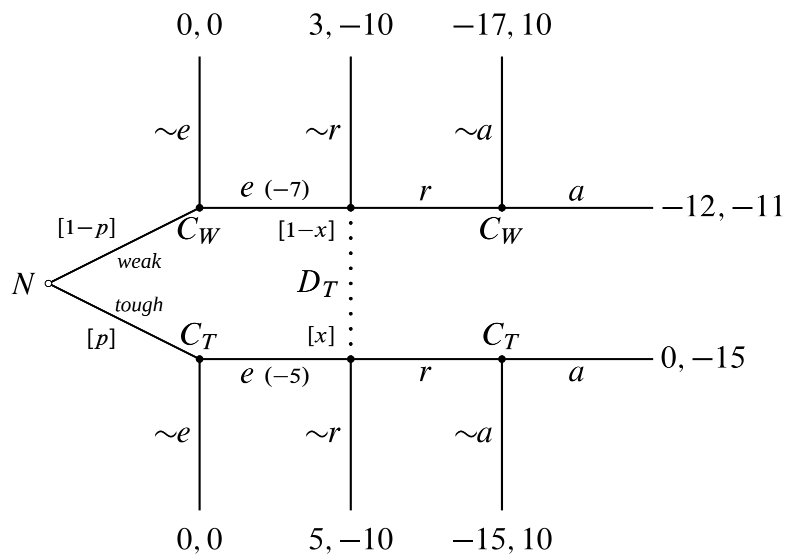

Figure 10: Escalation with Military Mobilization and a Tough Defender.

Observe now that the military mobilization has tied the challenger's hands: even the weak type prefers to attack under these conditions. Subgame perfection therefore requires that the challenger fight regardless of type when resisted. Just like with audience costs, militarized escalation creates a credible commitment to fight. Unlike audience costs, however, militarized escalation has also undermined the defender's resolve by lowering her expected payoff from war. Now she expects −11 if the challenger is weak and −15 if he is tough. Either of these is worse than capitulating and getting −10. Hence, the only sequentially rational strategy for the defender is to capitulate regardless of her beliefs (capitulation strictly dominates resistance). This makes escalation safe for the challenger, so he escalates regardless of type. The unique sequential equilibrium is for the challenger to escalate (and fight) regardless of type, and for the defender to capitulate for sure. The outcome is peaceful revision of the status quo.

Notice now an interesting aspect of this result: the military move tied the challenger's hands committing him to war, but also "untied" the tough defender's hands by undoing her commitment to fight. Observe that this is so irrespective of the defender's beliefs. Military moves represent a combination of sinking costs and tying hands: although they are inherently costly, they can alter the configuration of commitments by improving one's own war payoff and decreasing the opponent's war payoff. They can *undermine* the opponent's commitment making a threat that used to be believable incredible in the process. Hence, a tough defender could not be compelled to back down even with a credible threat to fight unless the threat also untied her hands. This cannot be done with audience costs.

Of course, before we get carried away and conclude that military moves are always good, I should note that it is quite possible for the challenger to miscalculate if he is uncertain about the defender. Observe that since war preparations are costly, the challenger will try to mobilizing the minimum force he thinks is necessary to get the defender to quit. Obviously, if the defender is strong, he would have to mobilize (expensively) a lot more than if she were weak. If the challenger is unsure about the defender's type, he may end up mobilizing less than necessary to get a strong defender to quit (because he erroneously believes the opponent is weak and tries to save on the mobilization costs). In that case, the tough defender would still resist, causing war. Hence, military moves themselves are not a panacea either. However, they can help establish credible commitments or, failing that, at least undermine the opponent's commitment sufficiently to cause her to capitulate.

1.2.3 Conclusion about Costly Signals

In general, the challenger must try to convince the defender that he is tough in order to compel the defender to capitulate. To be credible, the signal that reveals this information must be costly or risky enough: it has to be unattractive to the weak type even when it succeeds in persuading the opponent that he is tough. The challenger may also attempt to rearrange his incentives and alter his prior commitment. If he makes peaceful capitulation sufficiently bad, then even the weak type would prefer to go to war. This makes escalation a credible signal. However, as we have seen, there are some problems with the audience cost mechanism. Alternatively, the challenger may attempt to undermine the opponent's commitment through military moves. If he succeeds, he may compel her capitulation even in the absence of a credible threat to fight. Generally, none of these tactics can eliminate the risk of war: in equilibrium the probability of war will tend to be positive under two-sided incomplete information.

Although the strong types would normally expect better deals which they would get through credible signaling, they would also usually face higher risks of war that the weak types. Unfortunately, it is precisely the willingness to incur these costs and run these risks that makes their behavior credible because the weak would not want to mimic it. Hence, in many crisis situations the attempt to resolve the crisis on attractive terms must carry a positive risk of war. The only way to get the opponent to concede is to make the alternative sufficiently unattractive, and this involves convincing her that war is quite likely if she fails to concede and when war comes, it will be highly unpleasant.

2 Screening

Revealing information through costly signaling is what informed parties do. Eliciting information through screening is what their uninformed counterparts try to do.

If your opponent knows more than you do and this informational asymmetry is a disadvantage for you (it may not always be), then you can use strategies that induce your opponent to reveal what he knows. You are **screening** your opponents if you behave in a way that induces different types to respond differently. These responses, which are conditioned on their privately known information, tell you about what they know.

The most famous example of screening comes from the Bible. Remember the story about King Solomon who has to figure out which of two women claiming to be the mother of a baby is telling the truth (this was in the age before DNA testing)? He ordered the baby cut in two, with each woman getting a half. The fake mother agreed to the decision. The real mother, who would rather the child live than have it, asked the king to spare the baby's life and give it to the other woman. Solomon inferred that the latter was the real mother and ordered the child returned to her. Solomon screened the "type" of women (where "type" refers to her being a real mother or not) by the cunning use of a completely transparent strategy that induced the different types to choose separating strategies.

Whether he was wise or not I can't say, but the fake mother was truly dumb. We, being schooled now in strategic thinking, would immediately see that the correct way to avoid Solomon's crude **screening device** is to pool with the other woman (engage in signal-jamming) by saying whatever she says. As we know, in this case the uninformed party has no way of learning anything he does not already know.

Going back to the escalation game, suppose before it begins, the defender gets to choose whether to implement some sort of defense. There are two types of available defenses that she could deploy. The first is not very expensive but does not help if fighting actually occurs. It only makes escalation costlier to the challenger. Suppose then that if this defense is built, the weak challenger would have to pay a cost of 11 in order to escalate, and the tough challenger would have to pay a cost of 5. We know what happens in this game from our previous analysis of sunk costs.

The second type of defense is quite costly, but it works after fighting breaks out as well. It makes escalation costly but in case the challenger attacks, it imposes additional costs. If this defense is built, the challenger has to pay 11 if weak and 5 if tough just to escalate, but if the game ends in war, each type of challenger pays an additional 10. The new game is shown in Figure 11. The only difference between this an the game in Figure 2 is in the challenger's payoffs in case of war where an additional 10 was subtracted.

In this game with strong defense, neither type of challenger finds it worthwhile to attack if resisted. Even the tough challenger would not fight under these conditions (payoff from attacking is -16, which is strictly worse than payoff from not attacking, which is -15). Given that neither challenger would attack, the defender always resists regardless of her beliefs because doing so, given that all challengers capitulate, would yield 10, while not resisting yields -10. But if the defender always resists, then escalation is equivalent to capitulation for the challenger, and so

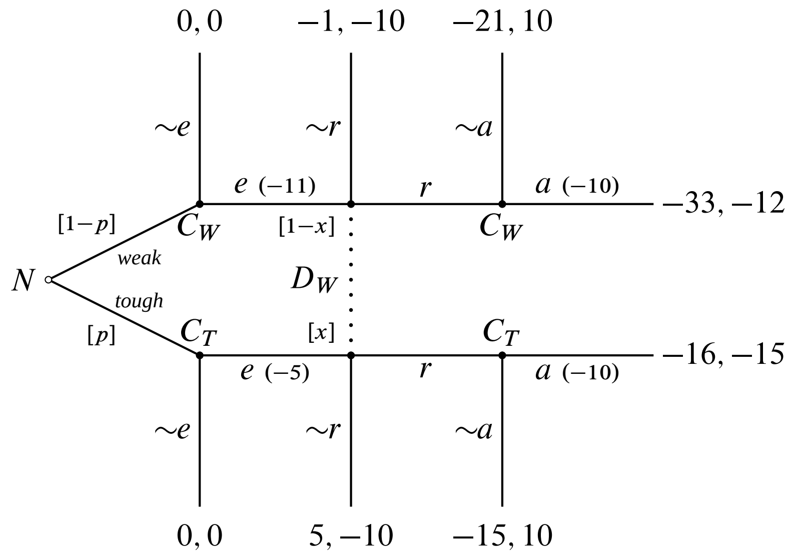

Figure 11: Costly Escalation With Strong Defense of a Weak Defender.

no challenger would ever escalate.

The only sequential equilibrium in this game is for the challenger to never escalate and always capitulate if resisted, and for the defender to always resist. In other words, the strong defense is a perfect defense because it would deter all challengers from creating a crisis.

We now know three things:

1. if she builds no defenses, the defender would end up in the original crisis game where she would capitulate if $p > 0.8$ and risk war if $p < 0.8$;

2. if she builds the cheap defense, the defender would end up in the costly signaling game where only the tough challenger escalates, so from her perspective before the crisis begins, the she would capitulate with probability p (the probability that a challenge will happen, in which case she always submits), and get the status quo with probability $1-p$ (the probability that no challenge happens because the opponent is weak);

3. if she builds the expensive defense, the defender would end up in the strong defense game where she would get the status quo regardless of p.

Which of these defenses should the defender purchase? It really depends on their costs. Suppose the cost of the cheap defense is 1 and consider an expensive defense that costs 8 units to the defender.

Suppose first that $p > 0.8$. Building no defenses results in capitulation and a payoff of $\mathrm{U}_D(\text{none}) = -10$. Building the cheap defense results in capitulation with probability p and the status quo with probability $1 - p$, and so the expected payoff is $\mathrm{U}_D(\text{cheap}) = -10p + (0)(1-p) - 1 = -10p - 1$. Finally, building the expensive defense results in perfect deterrence, so the outcome is the status quo: $\mathrm{U}_D(\text{expensive}) = 0 - 8 = -8$. Note now that $\mathrm{U}_D(\text{expensive}) > \mathrm{U}_D(\text{none})$, so we know that the defender will definitely build some defense rather than remain defenseless. But which one? she will choose the expensive defense whenever $\mathrm{U}_D(\text{expensive}) > \mathrm{U}_D(\text{cheap})$, or $-8 > -10p - 1$, which holds when $p > 0.7$. Since we have assumed that $p > 0.8$, this condition certainly holds. Therefore, if $p > 0.8$, the defender will build the expensive defense.

Suppose now that $p < 0.8$. Building no defenses produces the semi-separating equilibrium of the original escalation game. The expected payoff there is:

$$\begin{aligned}\mathrm{U}_D(\text{none}) &= \Pr(\text{War})(-15) + \Pr(\text{Cap}_C)(10) + \Pr(\text{Cap}_D)(-10) + \Pr(\text{SQ})(0)\\ &= (0.5p)(-15) + (0.125p)(10) + (0.625p)(-10) + (1 - 1.25p)(0)\\ &= -12.5p,\end{aligned}$$

where we obtain the relevant probabilities from our solution to the escalation game in the previous lecture.[5] Nothing has changed for the other two calculations because the equilibria remain the same regardless of the value of p. Note now that:

$$\begin{array}{rll}\mathrm{U}_D(\text{none}) > \mathrm{U}_D(\text{cheap}) & \Leftrightarrow -12.5p > -10p - 1 & \Leftrightarrow p < 0.40\\ \mathrm{U}_D(\text{none}) > \mathrm{U}_D(\text{expensive}) & \Leftrightarrow -12.5p > -8 & \Leftrightarrow p < 0.64\\ \mathrm{U}_D(\text{cheap}) > \mathrm{U}_D(\text{expensive}) & \Leftrightarrow -10p - 1 > -8 & \Leftrightarrow p < 0.70\end{array}$$

We can now conclude that the rank-ordering of the defenses is as follows:

$$\begin{array}{rl} p > 0.70: & \text{expensive} \succ \text{cheap} \succ \text{none} \Rightarrow \text{expensive}\\ 0.64 < p < 0.70: & \text{cheap} \succ \text{expensive} \succ \text{none} \Rightarrow \text{cheap}\\ 0.40 < p < 0.64: & \text{cheap} \succ \text{none} \succ \text{expensive} \Rightarrow \text{cheap}\\ p < 0.40: & \text{none} \succ \text{cheap} \succ \text{expensive} \Rightarrow \text{none}\end{array}$$

[5] We could have calculated this also as follows. Suppose D faces the tough challenger, who always escalates. Recall that D resists with probability $1/2$, in which case C attacks, and capitulates with probability $1/2$. The expected payoff against the tough challenger is then:

$$\mathrm{U}_D(\text{none}|C_T) = 1/2(-15) + 1/2(-10) = -12.5.$$

Suppose now D faces the weak challenger. This one escalates with probability $\beta^* = p/[4(1-p)]$, in which cases D resists with probability $1/2$, which leads to C's capitulation, and capitulates herself with probability $1/2$. Of course, if the challenger does not escalate, the status quo remains. Hence, D's expected payoff against a weak challenger is:

$$\mathrm{U}_D(\text{none}|C_W) = \beta\,[1/2(10) + 1/2(-10)] + (1-\beta)(0) = 0.$$

Since D believes she will face C_T with probability p, her overall expected payoff is:

$$\mathrm{U}_D(\text{none}) = p\mathrm{U}_D(\text{none}|C_T) + (1-p)\mathrm{U}_D(\text{none}|C_W) = -12.5p,$$

which is, of course, precisely what we found using the probability distribution over outcomes.

To recapitulate, the defender would choose defenses as follows, depending on the probability p that the challenger is tough:

- If p is less than 40%, then no defense will be built; tough challengers always escalate, weak ones bluff and sometimes do, and the defender sometimes resists; the probability of war is strictly positive in equilibrium, and the outcome is possible deterrence failure leading to possible compellence failure and war.

- If p is between 40% and 70%, then the cheap (weak) defense will be build; only tough challengers escalate, and the defender capitulates; probability of war is zero in equilibrium, and the outcome is peaceful revision of the status quo whenever deterrence fails (since this happens only if the challenger is tough, we have possible, but not certain, deterrence failure).

- If p exceeds 70%, then the expensive (strong) defense will be built; the defender gets perfect deterrence, and no challengers escalate; probability of war is zero in equilibrium, and the outcome is peaceful maintenance of the status quo (deterrence success).

Two things follow from this analysis. First, the defender will choose a defense to screen out her opponent's type. Depending on her prior beliefs, she will decide whether the expense is worth weeding out the weak types or not. If the probability of the opponent being tough is too high, the defender will not screen at all and will be vulnerable to exploitation by weak challengers. If, on the other hand, the probability is sufficiently high, then the defender will invest in perfect defense that would stop everyone.

The true screening begins when p drops below 70% but is still above 40%, in which case the defender will build a defense that would only stop weak challengers. She will prevent them from bluffing, but will not deter tough ones. In other words, the defender will screen out the challengers such that when escalation occurs she will know for sure that her opponent is tough, and will be able to capitulate, avoiding war.

Finally, if p drops below 40%, screening is no longer optimal either. The defender takes her chances with war. It is very significant to note, however, that with the use of screening techniques, the dangerous region where war is possible shrank from anything below 80% to half of that. The largest probability of war in this equilibrium now does not exceed 20% as opposed to 40% in the original escalation game.

Eliciting information is useful although it is costly. Revealing information is also useful even though it is costly as well. It is worth repeating that perfect deterrence, even if available, may not be chosen by a rational defender. Why?

Because, as the strategist Bernard Brodie put it, **strategy wears a dollar sign**, the second major conclusion from this analysis. The costs of alternative forms of defense will determine which one we will choose to acquire. Perhaps not surprisingly, the defender sometimes chooses to forego the strong defense because it is too expensive under certain conditions (e.g., when she thinks her opponent is not exceedingly likely to be tough). More surprisingly, however, the defender will often forego even the cheaper alternative. Sometimes she will opt for a weaker defense that does not work that well (but at least prevents war) but sometimes she will build no defense at all, taking her chances with war. These considerations are driven by the costs of defense relative to what it is expected to achieve in deterrence.

3 Incentive Schemes

Eliciting information may not be your goal. Instead, you may be far more interested in ensuring that the other party behaves appropriately even if you cannot observe their actions. This is a common problem with arms-control agreements: Is the opponent reneging and secretly stashing deadly weapons, or perhaps even secretly building new ones? You could demand verification, which is costly and difficult. Or you could design the agreement in a way that the opponent would find it worth his while to behave as you want him to. A strategy that attempts to influence such an unobservable action (compliance with the treaty) by punishing or rewarding observable outcomes of that action is called an **incentive scheme**.

Suppose you are the U.S. negotiator and you are trying to get the North Koreans to dismantle their nuclear program. The outcome of this project is uncertain and depends on the quality of North Koreans' efforts. If successful, the benefits from nuclear disarmament are estimated to be worth \$60 billion (no need to maintain large army close to the border, normalization of relations, trade benefits, etc.) You have estimated that the probability of success is 60% if the Koreans make a half-hearted effort to comply with the program, but raises to 80% if they make a more determined effort. (It is never 100% because you never know what their erratic leader is going to decide in the end.)

Expending effort entails costs to the Koreans. Let's say you have to pay them \$10 billion for routine efforts and an additional \$5 billion if they make an extra quality effort. (They may have to bribe important officials along the way to guarantee compliance, and so on.)

Is the extra payment in return for extra effort worth your while? Without the extra effort, your expected payoff is $0.6 \times \$60 - \$10 = \$26$ billion; that is, 60% chance of success with routine effort, for which you have to pay \$10. With the extra effort, your expected payoff is $0.8 \times \$60 - \$15 = \$33$ billion; that is, 80% chance of success with extra effort, for which you have to pay \$15. Clearly, the extra payment is worth your while because you should expect to get \$7 billion more in returns.

The problem is how to implement this agreement. You could sign a contract with the Koreans stipulating a payment of \$10 billion with \$5 for extra effort. But how would you know that they made this effort? They could simply take your additional money and implement the routine effort anyway. Much of this effort will be behind-the-scenes work involving bribes or unobservable actions in locations you don't know about and therefore cannot monitor, and the Koreans can (and will) always claim that they have made every possible attempt to get the results even if they have not. You have no way of knowing. If the nuclear program gets dismantled, they claim success and attribute it to their "spectacular" efforts. And if the program does not get dismantled, they blame failure on bad luck in spite of their "spectacular efforts." Because the chances of failure are 1 in 5 even with high effort, you cannot be sure that they are lying.

This is a general situation in which decisions about compliance have to be made on the basis of an observable and verifiable outcome which, unfortunately, is *probabilistically determined by compliance*. That is, it gives some information about effort, but that information is not perfect. How do you solve such a problem?

Let s be the compensation you offer for the routine effort and b be the bonus in case of observed success. Consider the Koreans' expected payoffs. If they make the routine effort, they would get s for sure, and b with 60% chance, that is: $s + 0.6b$. If they make the extra effort, they would get s for sure and b with 80% chance, that is: $s + 0.8b$. The expected benefit from the extra effort is therefore: $(s + 0.8b) - (s + 0.6b) = 0.2b$. Since you are offering \$5 billion for the extra effort, they will only expend it if:

$$
\begin{aligned}
0.2b &\geq \$5 \\
b &\geq \$25.
\end{aligned}
$$

That is, if the bonus offered exceeds \$25 billion! The intuition is that the bonus multiplied by the increase in probability of getting it through extra effort should be enough to compensate the Koreans for the cost of that extra effort. If you offer a bonus that is sufficiently high, you will be able to get them to expend extra effort for sure. This condition is called the **incentive-compatibility constraint**. Without it, your offer will not get the Koreans to work hard.

Of course, the Koreans would not work at all, let alone hard, if your offer is not sufficient to compensate them for routine work. Suppose you offer the appropriate bonus and so the incentive-compatibility condition is met. Then the Koreans would expected to get $s + 0.8b$, which must be greater than the \$15 that you are going to pay them in this case. That is, $s + 0.8b \geq \$15$. This is their **participation constraint**.

Since you don't want to spend taxpayers' money on North Koreans unless absolutely necessary, you want to keep their compensation as low as possible and still consistent with the two constraints. So, you will choose $s + 0.8b = \$15$, and since $b = \$25$, this means that $s = -\$5$. What? You are going to offer a negative base compensation?

There are two ways to interpret this. First, you may require the Koreans to put up \$5 billion of their own money for the transaction. The bonus in this case would simply be their share of the "partnership." Second, you may want to fine them \$5 if the project fails. Of course, it remains dubious that such a thing is enforceable at all.

So what next? The base compensation s must be non-negative. In this case, we set it at the lowest possible level (0), but since it requires $b = \$25$ to meet the incentive-compatibility constraint, you will be overpaying because $0 + 0.8 \times \$25 = \$20$, which is the expected amount the North Koreans are going to make. This, then, is the cost of **asymmetric information** to you, the less informed player.

We have already seen how one might deal with this situation. There are ways of eliciting appropriate behavior from better informed players, but this always entails costs. Under complete information, you would pay \$10 for base effort and then the additional \$5 only if you observed the extra effort. Koreans making extra effort would therefore cost you \$15. The calculations above show that Koreans making extra effort under asymmetric information would cost you \$20. The difference of \$5 billion is what you pay to overcome your informational disadvantage. Is it worth it?

What do you get by spending the additional \$5? You get extra effort, and so your expected payoff is: $0.8 \times \$60 - \$20 = \$28$. That is, you expect to get \$28 billion from the project. If you did not spend the extra \$5, you only need to pay the basic effort \$10 (you would never offer the additional \$5 bonus for a total of \$15 because it would not make the Koreans work harder anyway). Your expected payoff then is: $0.6 \times \$60 - \$10 = \$26$. That is, you should expect to get \$26 billion from the project. In this case, even with the extra cost resulting from the informational asymmetry, you get more expected profit by using the incentive scheme.

This need not always be the case. If, for example, the benefit was only \$40 billion, then paying the costs to overcome informational asymmetries would not be worth your while, and you would settle for the low-effort, low-probability variant. This is very similar with the situation the defense acquisition where a low enough probability of a tough challenger does not make the expense of the defense worth while and the defender simply takes his chances in crises.

You have seen incentive schemes in action many times. Just think about the insurance company who has insured you against a large loss (e.g. collision and theft coverage on your car). You can affect the probability of them having to compensate you by the effort you spend trying to avoid the loss. For example, you could be a really careful driver who never speeds, never engages in aggressive driving, and never tailgates others on the highway when pissed off at them. But if your car is fully insured, you may be tempted to indulge in riskier behavior: after all, what's another dent if someone else is going to pay for it?

The simple act of insuring you can induce you to take less care to avoid what you are being insured against. This is called **moral hazard**, and is a problem analogous to the one you faced with the North Koreans. This problem is controlled (not solved) in a similar manner: you are required to retain part of the risk through the deductible. This works as an incentive scheme because you will be less willing to run the risk that would cause you to have to pay this deductible. This is why agreeing to a larger deductible (meaning you are carrying a larger portion of the risk) lowers your premium. If you demand the lowest possible deductible you are telling the insurance company that you will most likely be negligent (because you show yourself unwilling to pay to prove you won't be... costly signaling, remember?) and so they compensate for the additional risk by jacking up your premium. Nice, eh?

We shall have an opportunity to think about incentive schemes when we discuss arms control and disarmament agreements in particular. As you can already see, many of these issues are of generic interest that you will see in your everyday life.