National Security Strategy: Credible Moves and Perfect Equilibrium

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

February 6, 2008

Overview. We have now seen how to solve games of complete information (perfect and imperfect) by finding the best responses of the players and then identifying the strategy profiles that contain only strategies that are best responses to each other. We now investigate best responses in games where players move sequentially. We develop the concept of credible threats and promises, which is the fundamental building block of our theories of the use of force. We incorporate the idea of credible moves in our refined solution concept of perfect equilibrium and learn how to find it by backward induction.

Recall that last time we solved the Crisis Game with imperfect information and found three equilibria, two in pure strategies and one in mixed strategies. I argued that the mixed-strategy equilibrium can provide us with numerous insights about general crisis situations: the element of danger that must be present, and how that relates to the probabilities of a crisis eruption in the first place and a crisis ending in disaster once begun.

A critical feature of the general crisis model was the information available to the players when they made their decisions. We assumed that neither observes the other's move when choosing the optimal strategy. That is, we assumed a game of imperfect information. We now turn to a variant of the game with perfect information: one of the players will observe the first move of its opponent before deciding what to do.

The optimal solution to this crisis game will allow us to define precisely what we mean by credible threats and promises. The ideas of credibility and commitment are the fundamental building blocks in the theories of the use of force, and so we must ensure that we understand them thoroughly. The goal is to use a series of simple models to illustrate each concept, and then investigate possible counter-moves for players who will try to cope with threats. This will give us a solid background for the historical analysis of deterrence and compellence as they were used in practice throughout the Cold War and after.

1 Nash Equilibrium With Perfect Information

1.1 The Crisis Game

Let us now consider the variant of the crisis game under perfect information, as shown in Figure 1. In this game, player 1 makes the first move, and player 2 observes it before choosing her response. Perhaps surprisingly, the analysis may appear a bit more involved. However, after we do the best-response exercise, we shall learn of a very simple method for solving these games! We shall solve the game for pure strategy equilibria because this is sufficient to demonstrate the difference between analyzing the perfect and imperfect information games.

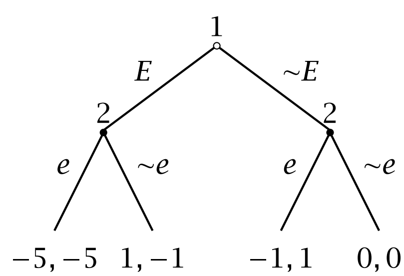


Figure 1: Crisis Game With Perfect Information.

Player 1 has two pure strategies, $S_1 = \{E, \sim E\}$, and player 2 has four pure strategies: $S_2 = \{(e,e), (e, \sim e), (\sim e, e), (\sim e, \sim e)\}$. We shall use s_1 to denote an arbitrary pure strategy for player 1. Similarly, we shall use s_2 to denote an arbitrary pure strategy for player 2. For example, $s_2 = (e, \sim e)$ would be one such strategy. Let's find the best responses for both players.

We begin with player 1. What are the best response strategies that he has? To find the set of best response strategies, we have to analyze all strategies of player 2. We begin with (e, e): that is, player 2 would escalate regardless of whether player 1 does. What is the best response to (e, e)? If player 1 chooses the strategy E, then the outcome will be disaster, and his payoff will be -5. If, on the other hand, player 1 chooses $\sim E$, the outcome will be victory for player 2, and the payoff for player 1 will be -1. Because $-1 > -5$, player 1 prefers to choose $\sim E$. That is, the strategy $\sim E$ is a best response to player 2's strategy (e, e).

What is the best response to $(e, \sim e)$? Playing E produces disaster, with a payoff of -5, while playing $\sim E$ produces the status quo with a payoff of 0. Since $0 > -5$, the best response (again) is to play $\sim E$.

What is the best response to $(\sim e, e)$? Playing E produces victory by player 1, with a payoff of 1. Playing $\sim E$ produces victory by the opponent, with a payoff of -1. Thus, the best response here is E because it gets $1 > -1$.

Finally, what is the best response to $(\sim e, \sim e)$? Playing E produces victory by player 1 with a payoff of 1. Playing $\sim E$ produces the status quo outcome, with a payoff of 0. Because $1 > 0$, the best response to this strategy is to play E.

Hence, the best responses by player 1 are:

$$\text{BR}_1(s_2) = \begin{cases} E & \text{if} \quad s_2 \in \{(\sim e, e), (\sim e, \sim e)\}, \text{ and} \\ \sim E & \text{if} \quad s_2 \in \{(e, e), (e, \sim e)\}. \end{cases}$$

Let's now look at the best responses for player 2. Again, we have to consider all possible strategies for player 1. What is the best response to E? Consider the information set following this action by player 1. If player 2 chooses e at this information set, the outcome is disaster, with a payoff of -5. If she chooses $\sim e$, the outcome is capitulation by herself, with a payoff of -1. Since $-1 > -5$, she prefers to play $\sim e$.

Is then $\sim e$ a best response to $s_1 = E$? No! The reason is that $\sim e$ is not a strategy for player 2! Recall that player 2's strategies have two components each. What we conclude from this is that the best response strategy for player 2 must involve playing $\sim e$ at the first information set. There are *two* such strategies: $(\sim e, \sim e)$ and $(\sim e, e)$, and either one of these will work. Thus, we see that it is quite possible to have more than one best response to a particular strategy.

What about the best response to $\sim E$? At this information set, choosing e produces victory by player 2, with a payoff of 1. Choosing $\sim e$, on the other hand, produces the status quo with a payoff of 0. Since $1 > 0$, player 2's best response must involve choosing e at this information set. There are two strategies that do this: (e, e) and $(\sim e, e)$.

Hence, the best responses by player 2 are:

$$\text{BR}_2(s_1) = \begin{cases} (e, e) \text{ or } (\sim e, e) & \text{if} \quad s_1 = \sim E, \text{ and} \\ (\sim e, \sim e) \text{ or } (\sim e, e) & \text{if} \quad s_1 = E. \end{cases}$$

We immediately see something very interesting about player 2's best response. The strategy $(e, \sim e)$ is not a best response to anything player 1 might possibly do. In other words, there is no action by player 1 that player 2 could optimally respond to with this strategy. This means that we should expect player 2 to *never play this strategy in this game.*

Let us now examine the strategy profiles. For example, one possible strategy profile in this game is $\langle E, (e, \sim e)\rangle$. In this strategy profile, player 1's strategy is E, and player 2's strategy is $(e, \sim e)$.

Consider first strategy E. From $\text{BR}_2(E)$, we know that it has the best responses $(\sim e, e)$ and $(\sim e, \sim e)$. To which of these two, if any, is E itself a best response? Looking at $\text{BR}_1(\sim e, e)$, we see that E is a best response. So, one Nash equilibrium is $\langle E, (\sim e, e)\rangle$.

Is E a best response to $(\sim e, \sim e)$? Sure it is, and so $\langle E, (\sim e, \sim e)\rangle$ is another Nash equilibrium. The equilibrium outcome is again capitulation by player 2.

Comparing the two solutions, we see that they have one thing in common: if player 1 expects player 2 to submit following escalation, then player 1 will always escalate because this gives him the best possible outcome (with the highest payoff). Conversely, if player 2 expects player 1 to escalate, then she prefers to submit because doing so, although resulting in capitulation, is still better than escalating in turn and ending in disaster. In both of these cases, the equilibrium outcome is victory by player 1. Remember this point, we shall return to it.

Continuing with our examination of profiles, consider now $\sim E$. We know that $\text{BR}_2(\sim E)$ is either (e, e) or $(\sim e, e)$. To which of these is $\sim E$ itself a best response? We have already seen that the best response to $(\sim e, e)$ is E, so the profile $\langle \sim E, (\sim e, e)\rangle$ cannot be Nash equilibrium because player 1's strategy is not a best response to player 2's strategy. However, $\text{BR}_1(e, e)$ is $\sim E$, and so the profile $\langle \sim E, (e, e)\rangle$ is yet another Nash equilibrium. The see the outcome produced by this strategy profile, note that in it player 1 submits, and player 2 escalates. The equilibrium outcome is victory by player 2.

In this equilibrium, player 1 expects that player 2 is going to escalate no matter what. Given such a strategy, player 1's optimal course of action is to submit because even though this would result in capitulation, it would avoid the worst outcome of disaster. Given that player 1 is expected to submit, player 2's strategy of escalating is also optimal.

Thus, the game of perfect information has three Nash equilibria, $\langle E, (\sim e, e)\rangle$, $\langle E, (\sim e, \sim e)\rangle$, and $\langle \sim E, (e, e)\rangle$. Let's think now carefully about what these profiles mean.

1.2 Incredible Threats in Equilibrium?

Consider the profile $\langle \sim E, (e, e)\rangle$. Player 1 chooses $\sim E$ because if he chooses E, player 2 is going to play e, and the outcome will be disaster. That is, not escalating is optimal because of player 2's **threat** to escalate if player 1 escalates. However, is this threat **credible**?

It is not. Think about what would happen if player 1 deviated and actually played E instead. Faced now with escalation, player 2 will prefer to play $\sim e$ and avoid disaster. The threat to escalate in response to player 1's escalation is not credible. This means that a rational player 1 should not believe it. In other words, player 2's strategy (e, e) cannot be optimal *if* player 1 actually escalates. Player 1 should believe that player 2 would respond to escalation with submission. But given that player 2 would capitulate following escalation, then player 1 can do better than submitting: he can actually escalate. Thus, the original strategy $\sim E$ cannot be optimal in turn! In other words, *because player 2 observes player 1's move before making her choice, player 2 cannot possibly threaten to escalate in response to escalation.*

What just happened here? What happened is the shortcoming of Nash equilibrium, which does not consider optimality during the game, only at the beginning of the game. The profile $\langle \sim E, (e, e)\rangle$ is a Nash equilibrium because it is optimal at the beginning of the game. If player 1 thinks that player 2 would play (e, e), then it is optimal to play $\sim E$. And since player 1 does not escalate, player 2's threat to escalate if player 1 does is never tested. Because player 2 never gets to respond to escalation, playing e there is optimal. But, of course, player 1 knows all this. He also knows that *if* he actually escalates, player 2 will back down for sure. Unfortunately, Nash equilibrium does not allow us to examine these kinds of conditional statements.

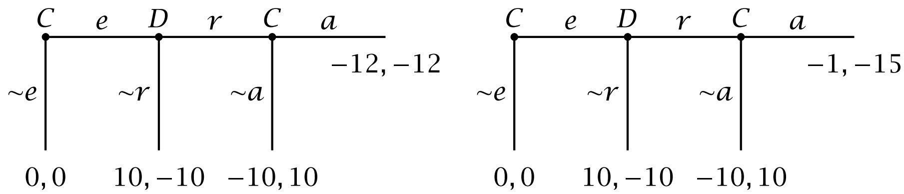
In fact, $\langle \sim E, (e, e)\rangle$ is not the only suspicious solution to the game. Consider $\langle E, (\sim e, \sim e)\rangle$. We already know that when player 1 escalates, $\sim e$ is the optimal response for player 2. Suppose now that player 1 does not escalate: is $\sim e$ an optimal response there? It is not because it yields a payoff of 0 as opposed to a payoff of 1, which player 2 could get by escalating instead. Thus, the **promise** not to escalate if player 1 does not escalate is not credible either! Again, *because player 2 observes player 1's move before making her choice, player 2 cannot possibly promise not to escalate in response to no escalation.*

In other words, player 2 can neither threaten to punish escalation by escalating, nor can she promise to reward restraint by not escalating. Player 2 cannot *credibly commit* to these courses of action. Our original solution concept of Nash equilibrium does not take into account the sequence of moves, and is not strong enough to rule out as unreasonable strategy profiles that include such incredible threats. What we really want is a stronger solution concept that would prune these profiles. Incidentally, this also demonstrates why we needed to include all information sets in the specification of a strategy: it allows us to identify profiles in which optimality may depend on unreasonable play at some information set.

1.3 The Escalation Game

Recall our complete information escalation game between a weak defender and a weak (irresolute) challenger, as shown in Figure 2(a), or between a weak defender and a tough (resolute) challenger, as shown in Figure 2(b).

Let's find the pure-strategy Nash equilibria for these two games. We can derive the best responses but instead we shall take a short-cut to simplify our task.



(a) Weak Challenger (b) Tough Challenger

Figure 2: Sequential Escalation Games with a Weak Defender.

Observe the following (these hold for both games):

1. If the defender plays $\sim r$ in some equilibrium, then $\sim e$ cannot be a part of any best-response by the challenger. If the defender is sure to capitulate, then any strategy involving escalation by the challenger would yield him a payoff of 10 whereas any strategy involving $\sim e$ yields a payoff of 0, which is strictly worse. Therefore, we conclude that $\sim r \Rightarrow e$ in any equilibrium.

2. If the defender plays r is some equilibrium, then e cannot be a part of any best-response by the challenger. If the defender is sure to resist, then escalation will lead either to war (with payoffs -1 if the challenger is tough and -12 if he is weak) or capitulation (with a payoff of -10). In all cases, the payoff is negative whereas any strategy with $\sim e$ yields 0, which is strictly better. Therefore, we conclude that $r \Rightarrow \sim e$ in any equilibrium.

3. If the challenger plays $\sim a$ in some equilibrium, then $\sim r$ cannot be a best-response by the defender. If the challenger is sure to capitulate when resisted, then the defender will get either 10 (the payoff from the challenger's capitulation if he also escalates) or 0 (the payoff from the status quo if he does not). If the challenger plays e, resisting is strictly better than capitulating, which would yield -10. If the challenger plays $\sim e$, then the defender is indifferent because she would get 0 whether she resists or not. However, notice that by (1), we know that $\sim r \Rightarrow e$ in any equilibrium, which implies that if she chose $\sim r$ predicated on her indifference given $\sim e$, then the challenger would not actually pick $\sim e$. Therefore, we conclude that $\sim a \Rightarrow r$ in any equilibrium.

These three implications are sufficient to give us the solutions to the games. Observe that from (3) and (2) we obtain $\sim a \Rightarrow r \Rightarrow \sim e$. In other words, in any equilibrium in which the challenger plays $\sim a$, the defender will play r, which in turn means that the challenger must be playing $\sim e$. Therefore, we have our first PSNE: $\langle(\sim e, \sim a), r\rangle$.

All the remaining solutions must involve the challenger playing a at his second information set. There are only two possibilities: (e, a) and $(\sim e, a)$. If he plays (e, a), then the defender cannot be playing r because by (2) if she did so, the challenger's best response must involve $\sim e$. Therefore, she must be playing $\sim r$, and by (1) this means the challenger is playing e. Therefore, we have our second PSNE: $\langle(e, a), {\sim}r\rangle$.

Finally, if the challenger plays $({\sim}e, a)$, then the defender cannot be playing ${\sim}r$ because if she did, (1) tells us that the challenger's best-response would have to involve e. Therefore, she must be playing r, and by (2) this means the challenger is playing ${\sim}e$. Therefore, we have our third, and final, PSNE: $\langle({\sim}e, a), r\rangle$.

Notice now two disturbing things about these solutions. First, the PSNE are the same for both games. It appears that it does not matter whether the challenger is weak or tough, Nash equilibrium leads to the same predictions for rational play. This is a bit odd since we know that a weak challenger would capitulate if resisted but the tough one will not. Intuitively, we would expect this to matter for the solution but at least in PSNE here it does not.

Second, some of the solutions appear suspect. For example, in the PSNE $\langle({\sim}e, a), r\rangle$, the challenger is successfully deterred from escalating by the defender's threat to resist. However, this threat cannot be credible if the defender really believes that the challenger would attack when resisted, as specified by his strategy. Hence, the defender is relying on an incredible threat: if attack is sure to follow her resistance, then she would capitulate instead, and, knowing this, the challenger would escalate.

The other two PSNE are also suspect depending on whether the challenger is tough or not. For example, in $\langle(e, a), {\sim}r\rangle$, he obtains the defender's capitulation by threatening to attack if resisted. However, the weak challenger would never do so, which means that when the defender knows that he is weak, she will not believe such a threat. But if she does not believe the threat, then she would certainly resist, in which case the challenger would not escalate in the first place. Analogously, in $\langle({\sim}e, {\sim}a), r\rangle$, the challenger is deterred by the defender's threat to resist. This threat is credible because the challenger is expected to capitulate for sure when resisted. However, we know that the tough challenger would never do that, which means that if the defender knows that he is tough, she would certainly capitulate. But if she does that, then the tough challenger will escalate instead.

Again, PSNE have undesirable properties, and they all hinge on the solution failing to pick out the problem that the strategies involve actions that players would not want to carry out when the contingency arises. We need a stronger solution concept that takes care of this problem.

2 Perfect Equilibrium

The revision of Nash equilibrium that can handle credible commitments properly is due to Reinchard Selten (co-winner with Nash of the Nobel Prize) and is technically known as **subgame perfect equilibrium**. Intuitively, it requires that the strategies are not only best responses to each other considered in their entirety, but also best responses to each other at any point in the game. That is, the strategies form a Nash equilibrium in every subgame. A *subgame* is any part of the game tree that (a) begins with a singleton information set, (b) includes all subsequent parts of the game tree, and (c) never splits information sets. Instead of defining subgames more precisely and messing with technicalities, let's just think about what this means in our example.

Consider player 2's decision following the action by player 1 in Figure 1. If player 1 has escalated, the subgame consists simply of player 2's choice. In this subgame there is only one player, so the optimal solution consists of this player's action. Because escalation yields -5 and no escalation yields -1, player 2's optimal response is to play $\sim e$; that is, $\sim e$ is the Nash equilibrium in this trivial game without an opponent.

Consider player 2's decision following non-escalation by player 1. In this subgame (again) there is only one player, and so Nash equilibrium is going to consists of its best action. Because escalation yields 1 and no escalation yields 0, player 2's best action is e; that is e is the Nash equilibrium in this trivial game without an opponent.

Thus, the perfection requirement that the strategies form Nash equilibria at all points in the game demands that player 2's strategy is $(\sim e, e)$ because these actions are the only possible best responses at the corresponding information sets.

What is then player 1's best response given that player 2 will play $(\sim e, e)$? We know that $\text{BR}_1(\sim e, e) = E$. And we further know that among the best responses to E we have $\text{BR}_2(E) = (\sim e, e)$. Therefore, $\langle E, (\sim e, e) \rangle$ is an equilibrium. Of course, we already knew that from our original analysis. The point is that it is the **only perfect equilibrium of the crisis game**. In other words, it is the only equilibrium in which all players' threats and promises are credible. All perfect equilibria are necessarily Nash equilibria but the converse (as we have just seen) is not true: there are many Nash equilibria that are not perfect.

By applying the perfection criterion, we have eliminated the unreasonable solutions $\langle \sim E, (e, e) \rangle$ and $\langle E, (\sim e, \sim e) \rangle$. That is, we should not expect rational players to ever play these strategies. We have concluded that $\langle E, (\sim e, e) \rangle$ is the only reasonable expectation we should have about the way the game is going to be played.

This immediately demonstrates the importance of **strategic moves**: player 1 can guarantee its most preferred outcome by **compelling** player 2 to back down. Conversely, player 2 cannot **deter** player 1 from escalating because it cannot credibly escalate in return. If player 1 commits to escalation, player 2 has no choice but back down. Interestingly, player 1's position here is enhanced by leaving the ultimate choice to its opponent! We shall have more to say about strategic moves later. Let's first learn a simple method for solving complete information games.

3 Backward Induction

If you have a game of perfect information, then you don't have to go through the best-response exercise to find the perfect equilibria. There is a much easier method called **backward induction**, which is really a way to look forward and reason backward.[1] That is, to determine what you should do today, you need to look at the consequences your various actions will have tomorrow, and then pick the one with the best consequences. More to the point, it involves looking forward to determine if your opponent will carry out the action he is threatening/promising to and deciding whether the threat/promise is credible. The easiest way to understand the process is through an example.

Consider the sequential escalation game with the weak challenger in Figure 2(a). Determining the credibility of an action boils down the determining whether the player would actually want to choose it if given a chance. For example, would player D prefer to resist if C actually escalated. Obviously, the answer would depend on whether C would prefer to attack if D actually resists. Thus, to determine the credibility of the promise to resist by D, we have to evaluate the credibility of C's threat to attack. This is why the process is called "looking forward." It is also called "reasoning backward" because once we've determined the credibility of C's threat to attack, we can go back and figure out the credibility of D's promise to resist, and so on.

As you can see, in this process you always have to go to the end of the game and start from there, working your way up to the beginning. Let's now do this in our example. The last move in the game belongs to C. If he attacks, his payoff will be -12, and if he does not attack, his payoff will be -10. Therefore, it is optimal not to attack. We conclude that in any perfect equilibrium, the action specified by C's strategy must include playing $\sim a$ at his last information set. This is illustrated in Figure 3, where the a action has been crossed out, and an arrow points to the outcome following the $\sim a$ (optimal) action.

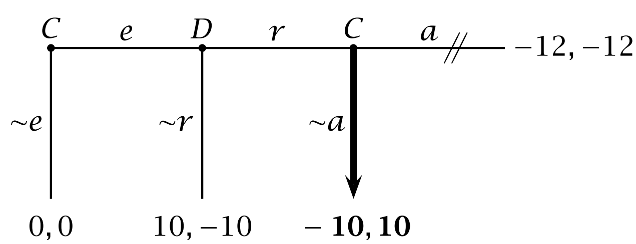


Figure 3: The Game After C's Optimal Attack Decision.

We now go back to D's decision at her information set. Playing $\sim r$ yields a payoff of -10. Playing r, on the other hand, leads to C's attack decision. We have already seen that the only rational choice there is $\sim a$, which means that playing r really leads to the outcome, where D's payoff is 10. Therefore, given that player C cannot credibly commit to attacking, player D's optimal course of action is to resist if challenged. Thus, in any perfect equilibrium, player D's strategy must specify r as the action at her information set. Figure 4 shows the resulting game

[1]Some authors use the term "rollback" in place of backward induction because it is simpler. The two are equivalent, but "rollback" has another meaning for national security and the use of force, so we shall stick with the older term even if it is a bit unwieldy.

tree, where the ~r action has been crossed out, and an arrow points toward C's information set. Thus, if D resists, the outcome is capitulation by C.

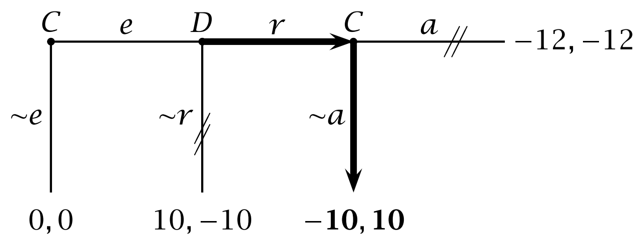


Figure 4: The Game After D' Optimal Resistance Decision.

We now move back up the tree to C's initial decision to escalate or not. If he does not escalate (plays ~e), then the outcome is the status quo with a payoff of 0. If he escalates, on the other hand, D will get to choose her course of action. We have already seen that the optimal choice is to resist, leading to the second information set for C, where he will inevitably back down. Thus, escalation is equivalent to having to capitulate later on in the game, with a payoff of −10. Therefore, C would choose ~e at the beginning of the game, avoiding the crisis (and therefore the need to capitulate when resisted) altogether.

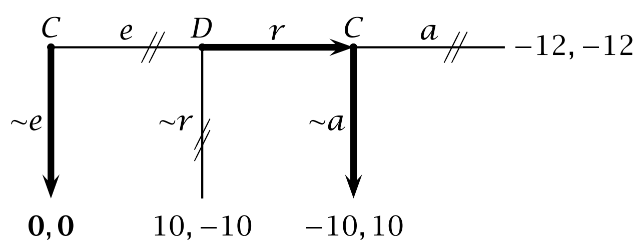


Figure 5: The Perfect Equilibrium of the Game with a Weak Challenger.

Thus, in any perfect equilibrium, player C's strategy must specify ~e as the action at the first information set. We can put all this together, as shown in Figure 5, which also demonstrates the perfect equilibrium of the game.

We conclude that the perfect equilibrium is $\langle(\sim e, \sim a), r\rangle$. That is, player C chooses not to escalate at the first information set and not to attack at the second information set, and player D chooses to resist if challenged. The equilibrium outcome is the status quo.

Note how this differs from the PSNE: we have eliminated two of them. Consider $\langle(\sim e, a), r\rangle$: the outcome here is also the status quo, and there's also a threat by D that causes C not to escalate. However, the defender's threat is incredible if she really thought C would attack. Furthermore, C's threat to attack itself is not credible if he is weak. Subgame-perfection eliminates this solution because it involves incredible moves (by both players actually). Consider now $\langle(e, a), \sim r\rangle$. As I noted before, this equilibrium is only reasonable if C has a credible threat to attack when resisted, which is not the case when he is weak. In this case, D would not believe such a threat and would not play $\sim r$. Again, subgame perfection picks out this problem and prunes the PSNE. The only remaining PSNE is admissible because it involves only credible threats by both players.

This solution illustrates a profoundly important point about the optimality of strategies. It demonstrates *why we need to consider the actions at all information sets, including the ones not reached if the strategy is followed*. In our case, player C's strategy is $(\sim e, \sim a)$, and one may wonder why we need to worry about what C would do if D resists given that if C plays this very same strategy, D would never get to resist in the first place.

The answer should now be obvious from the backward induction: the only reason it is optimal for C not to escalate is because he expects D to respond by resisting, in which case C will be forced not to attack. That is, the optimality of the action at the first information set depends indirectly on the optimality of the action at the last information set through player D's optimal strategy. One action cannot be considered without the other. That is precisely why we need for strategies to list actions at all information sets. Otherwise, we would not be able to determine whether they are optimal.

The other important insight here is that optimal behavior critically depends on credibility of future actions. This is the topic which we shall begin investigating in depth next week. For now, note that because D can **compel** her opponent to capitulate, she can **deter** him from escalating in the first place. As we shall often see, the success of deterrence will mostly rest on the credibility of the threat to resist challenges; that is, on the defender's ability to compel C to undo his action. We shall cover these ideas in depth and at length when discussing Cold War history.

4 Move Order Advantages

In the escalation game, C's problem is that he cannot credibly threaten to attack at the last stage, which is why he cannot compel D to back down when challenged. This is very similar to our original crisis game in that D derives an advantage by saddling her opponent with the last move in the game. Again, relinquishing the initiative seems like a winning strategy because it leaves the opponent to make the painful choice.

You may be tempted to generalize that it is always better to relinquish the choice to your opponent, but this would be a false generality. To see why, let's consider the same escalation game, this time against a resolute challenger. Recall that this challenger would rather fight than capitulate.

We begin with C's last move. Given resistance by defender, attacking yields -1 while not attacking yields -10. Therefore, C's optimal action is to attack. In the perfect equilibrium, C's strategy must specify a for this information set. Given that C will attack if resisted, D's payoff from resisting is -15, while the payoff from capitulating is -10. Therefore, player D would prefer to capitulate. In other words, in the perfect equilibrium, D's strategy must be $\sim r$. Finally, we turn to C's initial decision. Not escalating yields a payoff of 0 from the status quo.

Escalation, however, leads to D's capitulation with a payoff of 10. Therefore, C would strictly prefer to escalate. In other words, in the perfect equilibrium, C's strategy must specify e as the optimal action at the first information set.

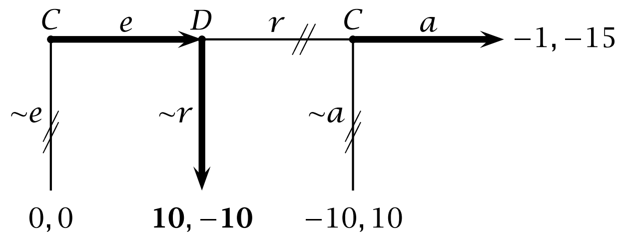


Figure 6: The Perfect Equilibrium of the Game with a Tough Challenger.

The perfect equilibrium, $\langle (e, a), {\sim}r \rangle$, is illustrated in Figure 6. Again, note the importance of specifying the actions at all information sets. If C's strategy did not include what to do when resisted, we would have no way of evaluating the optimality of D's strategy, and in turn, no way of telling whether escalation was optimal at the outset of the game. The equilibrium outcome is capitulation by the defender.

Further, unlike the previous game, D cannot benefit from relinquishing the initiative to C. In this case, the result (war) is much worse for D than capitulation. Because D cannot compel C to capitulate (since war is better for C than capitulation), D is forced to capitulate herself (because for her war is worse). Because of this, D cannot deter C from escalating in the first place. Thus, **the lack of a credible compellent threat leads to failure of deterrence**.

Observe now that this is the unique solution to this game: again we have eliminated two PSNE. As before, one of them was the blatantly improbable one $\langle ({\sim}e, a), r \rangle$ where D's behavior is inconsistent with her expectations about what would happen if she resists. The original solution fails to pick out this problem because D is never called upon to execute the threat to resist: if the strategies are followed, the game ends with the status quo. The other problem PSNE is the one that was subgame-perfect when the challenger is weak: $\langle ({\sim}e, {\sim}a), r \rangle$. A tough C does have a credible threat to attack when resisted, so D should not expect her to capitulate, which means D should not be able to deter C by threatening to resist. Again, the original solution did not pick this out because D's threat is never actually executed: if the strategies are followed, the game ends with the status quo.

This is actually the fundamental issue: Nash equilibrium does not detect incredible moves when they occur off the path of play (at information sets that are never reached if players follow their strategies). If all information sets are reached with positive probability, then Nash equilibrium will work just fine and all Nash equilibria will also be subgame-perfect. However, as we have seen what is not supposed to happen can have profound influence of what does happen, as Sherlock Holmes knew quite well when he talked about the dog that did not bark in the night.

Now we have a unique reasonable solution for each of the two possible scenarios we have considered. If the challenger is weak, he has no credible threat to go to war if resisted, which means that even a weak defender can take advantage of that and threaten credibly to resist any escalation. This, in turn, deters *C* from challenging the status quo in the first place. Peace prevails and the status quo persists. When the challenger is tough, he does have a credible threat to go to war if resisted, which causes *D* (who is weak) to capitulate if challenged. This causes deterrence failure. However, peace still prevails even though the status quo is revised in challenger's favor. In either case, war never occurs in equilibrium when the defender knows whether the challenger's threat to fight is credible or not.

And therein lies the problem, for in most real-life situations, the defender cannot be certain whether the challenger is going to fight or not. The fact that the perfect equilibrium takes a different form depending on that knowledge is both intuitive and worrisome: we need to analyze the situation while taking *D*'s uncertainty into account. As we shall see, this will dramatically alter our conclusions.

5 Coming Up...

The two examples with the basic escalation game show that depending on what we assume about the challenger, the optimal strategies and the equilibrium outcomes will be different. This is useful in itself for thinking about credible commitments, but does not tell us a whole lot about real crises, in which the defender is likely not to know whether the challenger is resolute or not. Next time we shall put these two games together and analyze the resulting game of incomplete information.

For now, chew on this: in both escalation games, the equilibrium outcomes (status quo with irresolute challenger and capitulation by defender with resolute challenger) involve zero chance of war. From what we've seen already, is this reasonable for models of crisis?