When to Appease and When to Punish: Hitler, Putin, and Hamas☆

David K. Levine[2], Lee Ohanian[1]

Abstract

Much has been written about deterrence, the process of committing to punish an adversary to prevent an attack. But in sufficiently rich environments where attacks evolve over time, formulating a strategy involves not only deterrence but also appeasement, the less costly process of not responding to an attack. This paper develops a model that integrates these two processes to analyze the equilibrium time paths of attacks, punishment, and appeasement. We study an environment in which a small attack is launched and can be followed by a larger attack. There are pooling and separating equilibria. The pooling equilibrium turns the common intuition that appeasement is a sign of weakness, inviting subsequent attacks, on its head, because appeasement is a sign of strength in the pooling case. In contrast, the separating equilibrium captures the common intuition that appeasement is a sign of weakness, but only because deterrence in this equilibrium fails. We interpret several episodes of aggression, appeasement, and deterrence: Neville Chamberlain's responses to Hitler, Putin's invasion of Ukraine, Israel's response to Hamas, Turkey's invasion of Cyprus, and Serbia's attacks in Kosovo.
JEL Classification Numbers: C7, D8

Keywords: social mechanisms, national defence, deterrence, appeasement

☆First Version: April 11, 2022. We would like to thank Rohan Dutta, Andrea Mattozzi, Salvatore Modica and Dan Treisman. We gratefully acknowledge support from the MIUR PRIN 2017 n. 2017H5KPLL_01.

*Corresponding author Lee Ohanian

Email addresses: `david@dklevine.com` (David K. Levine), `lee.ohanian@econ.ucla.edu` (Lee Ohanian)

[1]Department of Economics, RHUL

[2]Department of Economics, UCLA

"Deterrence is the art of producing in the mind of the enemy—the fear to attack."
Dr. Strangelove

1. Introduction

Much has been written about the policy of *deterrence*, the process of committing to a punishment in order to prevent an attack or other unwanted action from others.[3] But in sufficiently rich environments in which attacks may evolve over time in size and scope, formulating a strategy fundamentally involves not only deterrence but also the act of appeasement, the less costly process of not responding to an attack. Relatively less has been written on models that include both deterrence and appeasement.

This paper develops a model that integrates these two processes to analyze the equilibrium time paths of attacks, punishment, and appeasement. We create an environment that is relevant for analyzing attacks by a challenger that evolve over time, in which an initial, relatively small attack is launched, followed by subsequent attacks that are larger, such as Russia's invasion of Urkaine following its invasion of Crimea, and Germany's invasion of Poland, following its invasion of Czechoslovakia.

We call the initial attacks *probative* attacks, in which an attacker first launches a relatively small attack that can reveal information about the incumbent's strength. A potential second attack, which is larger, and which we call a *primary attack*, may be launched after the probative attack.

We embed these two types of attacks in a two-stage game, in which a small (probative) attack can occur in the first stage and is potentially followed by a larger (primary) attack in the second stage. The Incumbent - the Western Alliance, for example – can be either a weak or a strong type as measured by the cost of punishing the Challenger.

A key element of our model is that the intention of the challenger is uncertain and that it can be revealed over time. It therefore will be informationally advantageous to delay punishment of a probative attack until the second stage, at which time the incumbent can make a better-informed decision on whether to punish.

Prior to the first stage the incumbent makes a commitment to punishing either the probative attack, the primary attack, both attacks, or neither of the attacks. To capture the impact of possible changes in political preferences over time, we model this commitment to punish as one that can be revoked at a cost to the incumbent.

We report two main findings that provide a deeper and more nuanced understanding of the process of punishment and appeasement, and how these choices relate to understanding strong and weak incumbents. Absent the need to signal intentions, we find there is a pooling equilibrium in which it is always better to appease the challenger in the sense of delaying punishment for the first stage until after the second stage. In the pooling equilibrium, appeasement is a signal of strength,

[3] See, for example, Schelling (1960), Kahn (1960) and Fearon (2018).

thus turning on its head the conventional view that appeasing an invader is a sign of weakness that invites subsequent attacks.

We also find conditions under which equilibrium appeasement does reveal a weak incumbent, and thus provides new insights about the standard view of appeasement. Depending on the costs of the tradeoffs involved in punishing, there can be an optimal separating equilibrium with signaling in which the strong type punishes at the end of the first stage, thus signaling strength, and in which the challenger does not launch a primary attack. In contrast, the weak type revokes their commitment to attack in this equilibrium. The weak type appeases the probative attack, and thus suffers a primary attack in the second stage as the challenger correctly expects they will not be punished if they launch a primary attack.

A separating equilibrium reveals weakness because in this equilibrium, the challenger always attacks in the first stage, and sometimes in the second stage. For other parameter values, the optimal equilibrium is a pooling equilibrium in which appeasement always occurs in the first round, with the strong type punishing in the second round. In this equilibrium, the strong type is sufficiently committed to punish that the challenger chooses not to attack, even in the first round.

In the pooling equilibrium, deterrence is indeed very powerful. But in the separating equilibrium, deterrence is ineffective, at least in the first stage. A truly strong type has no need to signal their strength because they are known ex-ante to be strong, and this avoids a first period attack which would otherwise be launched to probe for weakness, with an eye to a possible second period primary attack.

The paper is organized as follows. Following our literature review, section 2 presents the model. Section 3 shows that absent signaling it is best to appease and wait for more information. Section 4 presents the main results on equilibrium outcomes with signalling involving punishment, appeasement, and the effectiveness of deterrence in the separating and pooling equilibriums. Section 5 discusses the implications of the model for case studies of Neville Chamberlain's responses to Hitler, Putin's invasion of Ukraine, Israel's response to Hamas, Turkey's invasion of Cyprus, and Serbia's attacks in Kosovo. Section 6 concludes.

Literature Review

This paper advances the literature by analyzing the informational advantages of waiting to punish and the finding that appeasement can be a sign of strength, and not just weakness. Moreover, the use of mechanism design principles leads to just two equilibria, rather than many, both of which shed light on the key aspects of incumbents and aggressors in understanding attacks.

There are earlier formal models studying appeasement and deterrence. Hirshleifer, Boldrin and Levine (2009) argue that absent indivisibility conflict should be avoided through appeasement. They do not, however, study when deterrence may be desireable.

The paper most closely related to the model here is in Treisman (2004) who focuses on the issue of punishing a challenger in the first period in hopes of deterring a different challenger in the second period. He gives as an example Britain in the late 19th Century facing a first period challenge from the US and second period challenge from France. This contrasts with our model of an incumbent facing an initial probing attack from a challenger such as with Hitler or Putin. This contrast drives some key differences between the structure of the models.

First, in our model there is an informational advantage of waiting so absent reputational issues it is never a good idea to punish in the first period. Second, in our setting the second period is more important (as measured by utility) than the first, while in Treisman (2004) the two periods are of equal importance.

Third, in our setting it is natural to consider different types of challenger as well as incumbent. As a result Treisman (2004) is led to study pooling equilibria in which there is never a second period attack, including those in which a first round attack and first round punishment takes place. By contrast in our pooling equilibria there is entry only by the committed challenger and the incumbent never punishes in the first round.

In addition, unlike Treisman (2004) separating equilibria play a key role in our narrative. Roughly speaking, while Treisman (2004) argues that appeasement can be rational in the face of future challenges, in our setting appeasement can in fact be an indication of strength. Finally, because we view the problem as one of mechanism design with partial commitment by the incumbent we are led to consider which equilibria are best for the incumbent and avoid the problem in Treisman (2004) that there are many equilibria.

Finally, the paper draws heavily on ideas from the reputation literature on uncertain types that originate in the work of the gang-of-four: Kreps and Wilson (1982) and Milgrom and Roberts (1982).

2. The Model

Two players, an incumbent player 1 and a challenger player 2 play a two period game $t = 1, 2$. In the first period the challenger may *attack* or *exit*. If they *exit* they continue to *exit* in period two. If they attack and attempt to exit in period 2 they succeed with probability $0 < 1 - \lambda < 1$, otherwise the attack continues.

Both the incumbent and challenger have two types. The challenger can be a *committed* behavioral type who attacks in the first period and attempts to exit in the second period or a *normal* type. It is common knowledge that the probability of a committed type is $0 < \pi < 1$. The incumbent is either *strong* s or *weak* w. It is common knowledge that the probability of a strong incumbent is $0 < \mu < 1$

If the challenger exits the incumbent and normal challenger get a basic payoff of zero. In the first period of an attack the incumbent receives a base payoff of -1 and the normal challenger receives a base payoff of a. In the second period of an attack the incumbent receives a base payoff of $-c < 0$. A normal challenger who continues to invade in the second period gets a benefit of 1, while if the challenger attempts to exit they get 0.

The incumbent has the ability to punish the challenger with utility penalties P. The cost to the incumbent for a punishment is $\psi_k P > 0$ with $k \in \{w, s\}$ and $\psi_s \leq \psi_w$. That is, the strong incumbent has a lower cost of imposing punishments on the challenger.

The timeline of play is as follows.

- The incumbent announces a punishment scheme P_1, P_2.
- The players privately learn their types.
- The incumbent may privately revoke their commitment to the punishment scheme at a utility cost of $R > 0$, otherwise the commitment is binding.
- Period 1: the challenger attacks or not. If there is an attack the incumbent punishes the challenger with $\tilde{P}_1 \geq 0$; if the incumbent is committed then $\tilde{P}_1 = P_1$.
- Period 2: the challenger continues to attack or attempts to exit. If the challenger attacks the incumbent punishes the challenger with $\tilde{P}_2 \geq 0$; if the incumbent is committed then $\tilde{P}_2 = P_2$.

The notion of equilibrium is that the incumbent may choose any Nash equilibrium of the game. This is equivalent to allowing the incumbent to choose the actions of the challenger in an incentive compatible way given a credible commitment plan.

We make one key assumption concerning parameters: we assume that a is not too large. If a is large then a normal challenger might prefer to attack in period one and attempt to exit in period two. This case is not of interest to us, so we assume specifically that $a < \lambda/(1-\lambda)$. This means that a may be positive (or zero, or negative) but cannot be too large. Hence the normal challenger can get a benefit from entering in period 1 but we will show that in equilibrium the normal challenger never enters in period 1 with the definite intention of leaving in period 2.

Types of Equilibria

It is convenient to distinguish three types of equilibria according to optimality: by an *equilibrium* we simply mean an Nash equilibrium of the game. By an *optimal equilibrium* we mean an equilibrium of the game that is optimal for the incumbent within the class of all equilibria. Finally, a *candidate equilibrium* is an equilibrium of the game within which the the optimal equilibria must lie: it satisfies necessary conditions for optimality but need not satisfy sufficient conditions.

We can also describe equilibria according to equilibrium play. There are two types of equilibria: *pooling equilibria* in which $\tilde{P}_1$ is independent of type and whether or not the commitment was revoked, and *separating equilibria* in which different incumbents choose different values of $\tilde{P}_1$. In separating equilibria a challenger who enters in period 1 may decide whether or not to enter in period 2 based on $\tilde{P}_1$. A pooling equilibrium in which $\tilde{P}_1 = 0$ we refer to as an *appeasement* equilibrium. An appeasement equilibrium in which $\tilde{P}_2 = 0$ we refer to as *trivial* - in this equilibrium the challenger always attacks.

3. The Informational Advantage of Waiting

As indicated if we assume that a is not too large regardless of the incumbent strategy the normal challenger never enters in period 1 with the definite intention of attempting to exit in period 2. Specifically

Proposition 1. *If $a < \lambda/(1-\lambda)$ there is no equilibrium in which a normal challenger who enters in period 1 attempts to exit for sure in period 2.*

Proof. Suppose it is optimal for an normal challenger to enter in period 1 and exit when feasible in period 2. Conditional on a positive probability$\tilde{P}_1$ let the expected punishment for continuing the attack be EP_2. Attempting to exit gives the challenger $a - \tilde{P}_1 - \lambda EP_2$ while attacking gives the challenger $a - \tilde{P}_1 + (1 - EP_2)$ hence $-\lambda EP_2 \geq 1 - EP_2$ so that $EP_2 \geq 1/(1-\lambda)$. Exiting in the fierst period gives 0 so $a - \tilde{P}_1 - \lambda EP_2 \geq 0$ or using $EP_2 \geq 1/(1-\lambda)$ and $\tilde{P}_1 \geq 0$ this implies

$$a \geq \lambda/(1-\lambda)$$

contradicting the assumption. □

This result highlights the role of the assumption that if the normal challenger attempts to exit but fails they lose the prize of the second period invasion. First, it is clear that the result goes through if they lose a positive fraction of the prize: we assume they lose the entire prize to avoid introducing an additional parameter. Second, if they do not lose the prize at all then the incumbent could deter entry followed by exit either by punishing in the first period or by punishing following exit in the second period. The latter is due to the better information technology available in period two as indicated in Proposition 2 below. By ruling out obtaining the entire prize on exit it is unnecessary for the incumbent to punish following exit and eliminating this possibility as we have done simplifies the analysis without losing any insight into the solution.

We next show how the informational advantage works by showing that equilibria with a single type never punishes in the first period: it is always better to wait and hope that some committed types exit.

Proposition 2. *Suppose that $\psi_s = \psi_w = \psi$ so there is only one type of incumbent. Then there are two types of optimal equilibria: if*

$$\psi < \frac{1-\pi}{\pi\lambda}\frac{c+1}{a+1}$$

and

$$\psi \leq \frac{1}{\pi\lambda}\frac{R}{a+1}$$

then the incumbent chooses $P_1 = 0$ and $P_2 = a+1$ does not revoke the commitment and gets a utility of $-\pi - \pi\lambda(c + \psi(a+1))$ because the challenger stays out. If either inequality strictly fails then the equilibrium is the trivial one in which the incumbent gets $-1 - (1 - \pi(1-\lambda))c$.

Proof. Utility for the incumbent in the trivial equilibrium is $-1 - (1 - \pi(1-\lambda))c$.

If a committed equilibrium is used it must yield higher utility and the punishment must be sufficiently small as not to be revoked. By Theorem 1 the normal challenger chooses between attacking both periods and exiting in period 1. If the normal challenger exits in perid 1 they get 0. If they attack in both periods they get $a + 1 - P_1 - P_2$. Hence the incentive constraint for the normal challenger to stay out can be written as $a + 1 \leq P_1 - P_2$. If the normal challenger attacks the utility of the incumbent is $-\psi(1+P_1+(1-\pi(1-\lambda))P_2)$. To maximize incumbent utility we see that the constraint must hold with equality so incumbent utility is

$$-\psi(P_1-(1-\pi(1-\lambda))(a+1-P_1).$$

Differentiating with respect to P_1 gives $-\psi\pi(1-\lambda)<0$ so that $P_1=0$ meaning that the P_2 information technology is superior and should be used with $P_2=a+1$. If the normal challenger stays out incumbent utility is then $-\pi-\pi\lambda(c+\psi(a+1))$.

We conclude that the incumbent strictly prefers deterence when

$$1+(1-\pi(1-\lambda))c>\pi+\pi\lambda(c+\psi(a+1))$$

and that this is credible if and only if $\pi\lambda\psi(a+1)\leq R$. □

4. Characterization of Equilibria

Theorem 1. *A non-trivial optimal pooling equilibrium is an appeasement equilibrium but the normal challenger stays out for certain. In an optimal separating equilibrium $P_1>0$, the normal challenger strictly prefers to attack, the weak type of incumbent revokes and chooses $\tilde{P}_1=0$ and the strong type never revokes. The normal challenger continues to attack against the weak type of incumbent and tries to exit against the strong type of incumbent.*

It is important also to know that all three types of equilibria are optimal for feasible parameter values.

Theorem 2. *Fix $a<\lambda/(1-\lambda),\pi$. For any R if $a<\lambda/(1-\lambda),R,\pi$. Suppose that*

$$\psi_s<\frac{R(1-\lambda)}{\pi\lambda}$$

$$\mu<\frac{\psi_s\pi\lambda(a+1)}{R}$$

$$c>\frac{\psi_s}{1-\lambda}$$

$$\psi_w>\frac{R}{\pi\lambda(a+1)},\frac{\mu(1-\pi)c}{a+(1-\mu)+\mu\lambda/(1-\lambda)},\frac{R(1-\lambda)}{\pi\lambda}$$

then the optimum equilibrium is separating. For any ψ_w if $R>\psi_w\pi\lambda(a+1)/(1-\lambda)$ the optimal equilibrium is pooling if

$$c>\frac{\pi}{1-\pi}\left((1-\mu)\psi_w+\mu\psi_s\right)\lambda(a+1)-1$$

and trivial if the reverse inequality holds.

These are propositions 3 and 4 in the Appendix. Roughly, if c is small, not surprisingly, the simple equilbrium is best. On the other hand large R favors pooling equilibrium as it is hard to get the weak incumbent to revoke, while large ψ_w and μ makes it desirable to avoid the high cost of having the weak incumbent attempt to prevent an invasion while low ψ_s makes it desirable for the strong incumbent to separate.

5. Applications

In interpreting these results it is important to realize that there are *de facto* three types of challenger: the normal type and two subtypes of the committed type depending on whether or not exit is successful in the second period. A committed type who enters and exits we may think of as a *sincere type*: a German leader who is strongly committed to protecting the German speaking Sudetenland but with no aspirations towards the rest of Czechoslavakia and Poland (the second period). By contrast the committed type who enters twice we may think of as a *hard type*.

In this vein we may consider the appeasement equilibrium, for example when the R/ψ_w is large, that is when the revocation cost is high or the cost of punishment to the weak type low. In this case the normal type stays out while both the sincere and the hard type enter. First period appeasement is good in both cases because neither type can be deterred from attacking in the first period. The sincere type then exits and the incumbent suffers only the small loss from the probatory attack. Should the incumbent have the ill luck to face the hard type they suffer a second period attack, and some cost of punishment. From Lemma 5 below the strong incumbent always punishes, while the weak incumbent punish for higher values of R/ψ_w but not for smaller values.

An example of a sincere type can be found in Treisman (2004)'s discussion of the incumbent Britain and the challenger US in the late 19th Century. Here the US was strongly committed to being the sole power in the Americas, but had no aspirations towards Britain's global interests - no interest, for example, in invading India.

A more modern example is that of Turkey. In July of 1974 Turkey launched a full scale invasion of Cyprus, an independent state with rule disputed between Greek and Turkish inhabitants. The circumstances were not dissimilar to those preceding the Russian invasion of the Donbass region and Crimea in 2014. The Russian invasion was ostensibly in response to the Maidan movement in Ukraine which removed the elected President Viktor Yanukovych. Similarly, in Cyprus, the Turkish invasion was ostensibly in response to a coup attempt against the elected President Makarios III by the Greek military junta. As was the case with the Donbass/Crimean invasion, there was negligible response: this probative attack was appeased. However, the Turkish government proved to be a sincere type: in the subsequent 49 years they made no effort to launch a primary attack either against the Greek portion of Cyprus or against Greece.

As NATO appeased Putin's probative attack against the Donbass and Crimea then fought against his invasion of Ukraine according to theory this means that Putin is a committed type, and since he entered twice, in fact a hard type. We discuss the appeasement of the hard types Hitler and Putin, in the conclusion. Here we examine the less well known case of the appeasement of Pakistan by the Soviet Union.

In Afghanistan in April 1978 the communist party took over and the Soviet Union effectively became the incumbent in Afghanistan. In October of 1978 a rebellion backed by the challenger Pakistan began and while achieving substantial success did not result in overthrowing the Soviet backed government. This we view as a probative attack. There was little Soviet response: a few helicopters and technicians were dispatched, but that was all. As a result the rebellion by December 1979 has grown into a primary attack threatening to bring down the Soviet backed government. The Soviet Union responded with a full scale invasion at the end of the month. In other words, the Soviet Union, despite their appeasement was committed to defending Afghanistan, and as deterence failed, we conclude that Pakistan was in fact a hard type: the behavior of the ISI over the years being generally consistent with this view.

The model also can be applied to the history of the Hamas-Israel conflict, which has been ongoing since 1989. Prior to October 7, 2023, Hamas had launched many attacks that were followed by Israeli military responses. We interpret the pre-October 7 pattern of Hamas attack-Israel response as Hamas probative attacks followed by Israeli punishment of probative attacks.

The ongoing sequence of attack-response suggests that deterrence was ineffective in dissuading Hamas from launching probative attacks, and that Hamas and Israel are operating within the separating equilibrium in which Hamas is the hard type of challenger who attacks, and Israel is the strong type of incumbent who responds to a probative attack.[4]

We consider the October 7, 2023 Hamas attack as a primary attack, as it is clearly much larger than earlier Hamas attacks. The model suggests that prior to October 7, Hamas never initiated a primary attack because Israel revealed themselves to be a strong incumbent – one with a relatively low cost of responding to a primary attack - and would punish accordingly if a primary attack was ever launched. But survey data suggests that the cost to Israel of responding strongly to a primary attack may have gone up, particularly if this cost is considered to involve its relationship with the U.S., as the U.S. provides financial support that represents about 10 percent of Israel's annual defense budget.

Gallup polling[5] shows that U.S. voter support for Israel has dropped substantially recently, particularly among registered Democrats and independent voters. In 2013, independent voters supported Israel by a 76 percent to 24 percent margin over Palestinians, and registered Democrats supported Israel by a 67 percent to 33 percent margin. But by March 2023, these positions had changed significantly. Independent support had declined to about 58 percent for Israel and increased to about 42 percent for Palestinians, whereas support among registered Democrats had flipped to about 56 percent for Palestinians and about 44 percent for Israel in March, 2023.

Given rising political support within the U.S. for Palestinian causes, Hamas may have mistakenly believed that Israel would not respond nearly as strongly as they did to the October 7 primary

[4] To connect our simple two-period model to the long history of Hamas-Israeli conflicts, assume that the game is being played independently over time and by a sequence of political leaderships for both players, such that Hamas continues to probe Israel for weakness, and that Israel responds since they are the strong type of incumbent.

[5] https://news.gallup.com/poll/472070/democrats-sympathies-middle-east-shift-palestinians.aspx

attack. Within the model, this can be modelled as either a change in Hamas's belief of Israel's punishment cost parameter, or a change in their perception of the size of Israel's response, or both.

While there may indeed be several factors involved in the Hamas decision to launch a much larger attack than they ever launched previously, the model suggests the possibility of mistaken beliefs regarding Israel's likely response playing an important role. Put differently, the model highlights the view that Hamas may have made a different decision on October 7 had they known Israel would launch such an enormous response to a primary attack.

Finally: while we know that deterence works - the story of the Cold War is pretty clear here - it also works with respect to probative attacks. One example is that of the Kosovo crisis. In the late 1990s Serbia had designs on the autonomous region of Kosovo. Gambling that the EU and NATO were weak they launched a series of probative attacks and provocations culminating in the Račak massacre in January of 1999: there was a separating equilibrium so deterrence failed. Nevertheless the EU and NATO proved to be the strong type and responded with six week bombing campaign against Serbia. As Serbia was not a committed type, no primary attack followed, and Kosovo has remained autonomous since.

6. Conclusion

We close by using the model to reinterpret Neville Chamberlain, who is considered by many to represent the epitome of appeasement. Well known is Chamberlain's speech after the Munich Accord, in which Germany agreed to respect the integrity of the remainder of Czechoslovakia after Hitler had annexed the German speaking portion of that country. Chamberlain stated "My good friends, for the second time in our history, a British Prime Minister has returned from Germany bringing peace with honour. I believe it is peace for our time."

But less well known is Chamberlain's subsequent speech regarding Hitler: "This morning the British Ambassador in Berlin handed the German Government a Final Note stating that unless we heard from them by 11 0'clock that they were prepared at once to withdraw their troops from Poland a state of war would exist between us. I have to tell you now that no such undertaking has been received, and that consequently this country is at war with Germany."

This analysis provides a very different interpretation of Chamberlain. Specifically, the game between Chamberlain and Hitler falls within our model's pooling equilibrium, in which it is optimal for Chamberlain not to have responded to Hitler's first-period attack in Czechoslovakia, but to respond to Hitler's second period attack in Poland. We therefore interpret Chamberlain not as an appeaser but as an incumbent committed to punish a second period attack. Unfortunately, Chamberlain was revealed to be the weak type of committed incumbent, and Hitler was revealed to be the hard type of challenger who continued to attack, rather than the normal type of challenger who would have stopped attacking after the Munich Accord.

Our model suggests the same pooling equilibrium interpretation regarding NATO's lack of response after Vladimir Putin annexed Crimea, with an apparent self-contained goal of making Crimea part of Russia. However, following Putin's 2022 attack, Putin revealed himself to be a hard type with much broader goals. And like Chamberlain after Hitler's attack in Poland, NATO responded to Putin's 2022 primary attack, which has included providing Ukraine with military and financial support and sanctions against Russia. The difference in interpreting the outcomes of the Hitler versus Putin primary attacks thus far is that either NATO is a strong type compared to Chamberlain, and/or Hitler was much better equipped to carry out a primary attack than Putin.

References

Hirshleifer, J., M. Boldrin, M. and D. K. Levine (2009): "The slippery slope of concession," *Economic Inquiry* 47: 197-205.

Dutta, R., D. K. Levine and S. Modica (2021): "Interventions with Sticky Social Norms: A Critique," *Journal of the European Economic Association*, forthcoming

Fearon, J. D. (2018): "Cooperation, conflict, and the costs of anarchy," *International Organization* 72: 523-559.

Kahn, H. (1960): "The nature and feasibility of war and deterrence," Rand Corporation.

Kreps, D. and R. Wilson (1982): "Reputation and Imperfect Information," *Journal of Economic Theory* 50: 253-79.

Levine, David and Salvatore Modica (2016): "Peer Discipline and Incentives within Groups", *Journal of Economic Behavior and Organization* 123: 19-30.

Mery, Aschkan (2021): *The Opium of the People: Essays on Counter-narcotics Efforts in Afghanistan*, EUI PhD Thesis.

Milgrom, P. and J. Roberts (1982): "Predation, reputation , and entry deterrence," *Journal of Economic Theory* 27: 280-312.

Olson Jr., Mancur (1965): *The Logic of collective action: public goods and the theory of groups*, Harvard Economic Studies.

Ostrom, Elinor (1990): *Governing the commons: The evolution of institutions for collective action*, Cambridge university press.

Schelling, T. (1960): *Strategy of Conflict*, Harvard University Press.

Townsend, Robert M. (1994): "Risk and Insurance in Village India," *Econometrica* 62: 539-539.

Treisman, D. (2004): "Rational appeasement. International Organization," 58: 345-373.

Appendix: Proof of the Theorems

Pooling Equilibria

Lemma 1. *In any candidate equilibrium if the commitment is revoked then* $\tilde{P}_2 = 0$*; if the normal challenger exits in period 1 for certain then also* $\tilde{P}_1 = 0$*, otherwise* $\tilde{P}_1 \in \{0, P_1\}$.

Proof. Attack always has positive probability so play must be optimal following an attack. Since $\tilde{P}_2$ is chosen after all moves by the challenger and is costly if the commitment has been revoked it must be chosen equal to zero in that case. If $\tilde{P}_1 \neq P_1$ then a normal challenger knows that the incumbent has revoked so will not punish in period 2 so attacks in period 2 with probability 1 so $\tilde{P}_1 \neq P_1$ has no deterence effect. Since with probability π a committed type does enter choosing $\tilde{P}_1 = 0$ minimizes cost. If the normal challenger exits in period 1 then it is merely costly to set $\tilde{P}_1 = P_1 > 0$ as there is no effect on the behavior of the committed challenger. □

Lemma 2. *In a non-trivial candidate pooling equilibrium the normal challenger does not strictly prefer to attack.*

Proof. Suppose the challenger strictly prefers to attack. By Theorem 1 the challenger does not strictly prefer to exit in period 2. If the normal challenger strictly prefers to attack in period 2 then the trivial equilibrium is better than the pooling equilibrium. Hence we may assume that the normal challenger is indifferent to attacking in period two. Hence the incentive constraint binding so $-\lambda EP_2 = 1 - EP_2$ or $EP_2 = 1/(1-\lambda)$. Since the challenger strictly prefers to enter $a + 1 - P_1 - EP_2 > 0$. Putting these together

$$a + 1 > \frac{1}{1-\lambda} + P_1$$

so

$$a > \frac{\lambda}{1-\lambda}$$

contradicting the condition $a < \lambda/(1-\lambda)$ that a not be too large. □

Lemma 3. *In a candidate equilibrium the challenger does not enter with positive probability when indifferent to entry.*

Proof. We will show that there is a better equilibrium. Consider keeping the strategy of the incumbent fixed and having the normal challenger exit. This is incentive compatible for the normal challenger. It might not be credible for the incumbent since the reduced punishment costs might make is strictly desirable for an incumbent type that previously was revoking to not revoke. Switch the strategy of such an incumbent to not revoking. This also increases the utility of the incumbent - and since by Lemma 1 it does not decrease the punishments issued to the challenger so will not cause the challenger to strictly prefer to enter. Hence this is an equilibrium that is strictly better for the incumbent than the equilibrium in which the challenger enters with positive probability. □

Lemma 4. *If the normal challenger exits for certain then in any candidate equilibrium $P_1 = 0$. In particular this is the case in a candidate pooling equilibria and conversely any equilibrium in which this is true is pooling.*

Proof. Since the challenger exits if the incumbent revokes then the incumbent must play $\tilde{P}_1 = \tilde{P}_2 = 0$ by 1. Suppose that $P_1 > 0$ and consider choosing instead $\hat{P}_1 = 0$, $\hat{P}_2 = P_1 + P_2$. This does not lower the incentive for the normal challenger to stay out but it does lower the cost to a non-revoking incumbent from $P_1 + (1-\pi(1-\lambda))P_2$ to $(1-\pi(1-\lambda))(P_1+P_2)$. It may also cause an incumbent who is revoking to strictly prefer not to revoke: if so not revoking further increases the utility of the incumbent and further increases the incentive of the challenger to stay out. Finally, it reduces the set of feasible strategies the normal challenger can use after a first period attack since it is pooling rather than separating, also reducing the incentive of the challenger to attack. □

Lemma 5. *In a candidate pooling equilibrium $\psi_s \pi\lambda(a+1)/\mu - R \leq 0$ or $\psi_w \pi\lambda(a+1) - R \leq 0$. If $\psi_w(a+1) > R$ the weak type revokes, the strong type does not and the punishment is $P_2 = (a+1)/\mu$ and incumbent utility is $-(1-\mu)R - \psi_s\lambda(a+1)(a+1) - \pi(1+\lambda c)$. If $\psi_w\pi\lambda(a+1) \leq R$ neither type revokes, the punishment is $P_2 = a+1$ and incumbent utility is $-\left((1-\mu)\psi_w + \mu\psi_s\right)\pi\lambda(a+1) - \pi(1+\lambda c)$.*

Proof. From Lemma 4 $P_1 = 0$. As the normal challenger stays out the normal challenger gets 0 and if the normal challenger enters the challenger gets $a+1$ from which we see that the expected punishment is $EP_2 = a+1$. Let ϕ_k be the fraction of the incumbents who are type k and do not revoke: the cost to the incumbent is $C = (1-\phi_s-\phi_w)R + (\phi_w\psi_w + \phi_s\psi_s)\pi\lambda P_2$ and the expected punishment is $(\phi_w+\phi_s)P_2$ giving $P_2 = (a+1)/(\phi_w+\phi_s)$. The revocation constraints are therefore

$$\Pi_k \equiv \frac{\psi_k}{\phi_w+\phi_s}\pi\lambda(a+1) - R \leq 0$$

if $\phi_s > 0$ while the objective is

$$C = (1-\phi_s-\phi_w)R + \frac{\phi_w\psi_w + \phi_s\psi_s}{\phi_w+\phi_s}\pi\lambda(a+1).$$

The revocation constraints cannot fail for both since we cannot have $\phi_s = \phi_k = 0$. If the revocation constraint fails for the strong type it fails for the weak type, so either it holds for both types or it fails only for the weak type.

If

$$\psi_s\pi\lambda(a+1) - R > 0$$

then both constraints fail for any feasible ϕ_k and there is no equilibrium in which the challenger stays out. If

$$\psi_w\pi\lambda(a+1) - R > 0$$

then for any feasible ϕ_k the weak type revokes, so $\phi_w = 0$ and if $\psi_s\pi\lambda(a+1)/\mu - R > 0$ the constraint must then be violated also for the strong type, so again there is no equilibrium in which the challenger stays out. Hence we may suppose that $\psi_s\pi\lambda(a+1)/\mu-R \leq 0$ and $\psi_s\pi\lambda(a+1)-R \leq 0$.

Differentiating the objective we see that

$$\frac{\partial V}{\partial \phi_k} = -R + \frac{\psi_k\left(\phi_w + \phi_s\right) - \left(\phi_w\psi_w + \phi_s\psi_s\right)}{\left(\phi_w + \phi_s\right)^2}\pi\lambda(a+1) = \Pi_k - \frac{\phi_w\psi_w + \phi_s\psi_s}{\left(\phi_w + \phi_s\right)^2}\pi\lambda(a+1).$$

If the revocation constraint is satisfied this is strictly negative. In case $\psi_w\pi\lambda(a+1) - R > 0$ so that $\phi_w = 0$ we have $\phi_s = \mu$. If $\psi_w\pi\lambda(a+1) - R \leq 0$ then $\phi_w = 1-\mu$ and $\phi_s = \mu$. □

Separating Equilibrium

In a separating equilibrium the probability distribution over $\{0, P_1\}$ is not the same for both types.

Lemma 6. *In a separating equilibrium the challenger strictly prefers to attack, there is positive probability of the weak type revoking and playing* 0 *in the first period. If the weak type plays* 0 *with probability less then* 1 *then the strong type plays* P_1 *with probability* 1.

Proof. By Lemma 3 if the normal challenger is indifferent they exit. If they strictly prefer exit out they do so. If the challenger exits by Lemma 4 $P_1 = 0$ so the the equilibrium is not separating. Hence it must be that the challenger strictly prefers to attack.

For there to be a separating equilibrium we must have $P_1 > 0$ and some type revokes and plays $\tilde{P}_1 = 0$ with positive probability. By Lemma 1 a revoking incumbent must choose $\tilde{P}_2 = 0$ so there are two revoking strategies: $\tilde{P}_1 = 0$ and $\tilde{P}_1 = P_1$.

Suppose the strong type revokes with positive probability. The weak type suffers higher costs from not revoking so strictly prefers to revoke. If the weak type chooses $\tilde{P}_1 = P_1$ then the strong type who has lower cost from this than the weak type must not revoke with $\tilde{P}_1 = 0$, hence plays $P_1$with probability 1. Otherwise the weak type plays $\tilde{P}_1 = 0$ with probability 1. □

Define θ to be the fraction of challengers playing $\tilde{P}_1 = P_1$ in period 1 who have not revoked.

Lemma 7. *An optimal separating equilibrium exists only if*

$$\psi_s\pi\lambda/(1-\lambda) \leq R \leq \psi_w\pi\lambda/(1-\lambda)$$

and

$$\frac{c}{\psi_w} < \frac{a + (1-\mu) + \mu\lambda/(1-\lambda)}{\mu(1-\pi)}.$$

In this case $\theta = 1$

$$P_1 = (1-\pi)c/\psi_w$$

and

$$P_2 = 1/(1-\lambda).$$

The challenger stays in when $\tilde{P}_1 = 0$ and stays out when $\tilde{P}_1 = P_1$; the weak type always revokes with $\tilde{P}_1 = 0$ and the strong type never revokes. The utility from the separating equilibrium is better than the simple equilibrium when $((1-\pi)(1-\psi_s/\psi_w)+\pi\lambda)\, c \geq \psi_s \pi\lambda/(1-\lambda)$.

Proof. If $\tilde{P}_1 = 0$ then the normal challenger attacks in period 2 for sure knowing there will be no punishment. If the normal challenger always attacks on $\tilde{P}_1 = P_1$ then they always attack and this implies a simple equilibrium rather than a separating one. Hence on $\tilde{P}_1 = P_1$ the probability the normal challenger attacks is $\gamma < 1$ and moreover, the normal challenger cannot strictly prefer to attack in period 2.

If the normal challenger exits in period 2 they get $-\lambda EP_2$ and if they enter they get $1 - EP_2$ so it must be that $\theta P_2 = EP_2 \geq 1/(1-\lambda)$.

The weak type must be willing to revoke and play $\tilde{P}_1 = 0$. This gives utility $-1-(1-\pi(1-\lambda))c - R$. Revoking and playing $\tilde{P}_1 = P_1$ gives utility $-1-((1-\pi)\gamma+\pi\lambda)c - R - \psi_w P_1$. Not revoking gives utility $-1-((1-\pi)\gamma+\pi\lambda)c-\psi_w P_1-\psi_w((1-\pi)\gamma+\pi\lambda)P_2$. Hence the two credibility constraints are

$$P_1 \geq (1-\pi)(1-\gamma)c/\psi_w \tag{6.1}$$

$$(1-\pi)(1-\gamma)c + R \leq \psi_w P_1 + \psi_w((1-\pi)\gamma+\pi\lambda)P_2 \tag{6.2}$$

with utility $-1-(1-\pi(1-\lambda))c - R$.

The strong type must be willing not to revoke. This gives utility $-1-((1-\pi)\gamma+\pi\lambda)c-\psi_s P_1 - \psi_s((1-\pi)\gamma+\pi\lambda)P_2$. Revoking and playing $\tilde{P}_1 = 0$ gives utility $-1-(1-\pi(1-\lambda))c-R$; revoking and playing $\tilde{P}_1 = P_1$ gives utility $-1-((1-\pi)\gamma+\pi\lambda)c - R - \psi_s P_1$. Hence the two credibility constraints are

$$\psi_s P_1 + \psi_s((1-\pi)\gamma+\pi\lambda)P_2 \leq (1-\pi)(1-\gamma)c + R \tag{6.3}$$

$$\psi_s((1-\pi)\gamma+\pi\lambda)P_2 \leq R \tag{6.4}$$

with utility $-1-((1-\pi)\gamma+\pi\lambda)c - \psi_s P_1 - \psi_s((1-\pi)\gamma+\pi\lambda)P_2$.

We claim that inequality 6.2 cannot bind. If it holds with equality then consider having all the weak not revoke observing that they did not do so originally. Since we showed $\gamma < 1$ the challenger weakly prefer to stay out when $\tilde{P}_1 = P_1$. Hence switching the weak to not revoking implies they strictly prefer to stay out, hence they will stay out in period 1 as well. This makes the incumbent better off and reduces the incentive to revoke so is a better equilibrium for the incumbent.

We claim that $\theta P_2 = 1/(1-\lambda)$. If not $\theta P_2 > 1/(1-\lambda)$ implying $\gamma = 1$ contradicting the fact that we proved that $\gamma < 1$.

We claim that $P_1 = (1-\pi)(1-\gamma)c/\psi_w$, that is inequality 6.1 must bind. If not we can lower P_1. This will not violate inequality 6.2 because we just showed it does not bind, and does not violate any other constraint. It does not effect the challenger who is entering in period 1 in any case, but it does lower cost for the incumbent.

Plugging $P_1 = (1-\pi)(1-\gamma)c/\psi_w$ into inequality 6.3 we see that

$$\psi_s((1-\pi)\gamma+\pi\lambda)P_2 \le (1-\pi)(1-\gamma)(1-\psi_s/\psi_w)c+R \tag{6.5}$$

which cannot bind by inequality 6.4. For the normal challenger only $\theta P_2 = 1/(1-\lambda)$ matters, and so 6.4 becomes

$$\psi_s((1-\pi)\gamma+\pi\lambda)/(\theta(1-\lambda) \le R \tag{6.6}$$

Notice that the LHS is increasing in γ so lowering γ helps with the constraint. The utility of the weak type does not depend upon γ, the utility of the strong type is decreasing in γ so we conclude that $\gamma = 0$.

Since the weak type is indifferent to revoking and $\tilde{P}_1 = 0$ the weak type gets $-1-(1-\pi(1-\lambda))c$. Since the strong type is willing not to revoke, the strong type gets $-1-P_1-\pi\lambda(c+\psi_s P_2) = -1-P_1-\pi\lambda(c+\psi_s/(\theta(1-\lambda))$ from which we see that θ should be chosen as large as possible.

We see that inequality 6.6 can be satisfied for $\theta \le 1$ only if $R \ge \psi_s((1-\pi)\gamma+\pi\lambda)/(1-\lambda)$. If we can find θ so that inequality 6.2 holds with equality or is violated this implies the existence of a better pooling equilibrium. Hence it must be possible to take $\theta = 1$ meaning that the weak types choose $\tilde{P}_1 = 0$ and the strong types do not revoke.

The overall utility of the incumbent is then

$$-1-(1-\mu)(1-\pi(1-\lambda))c-\mu\left(c(1-\pi)\psi_s/\psi_w+\psi_s\pi\lambda/(1-\lambda)\right)$$

while utility from the simple equilibrium is $-1-(1-\pi(1-\lambda))c$. Hence the separating equilibrium is better when

$$\left((1-\pi)(1-\psi_s/\psi_w)+\pi\lambda\right)c \ge \psi_s\pi\lambda/(1-\lambda).$$

Finally we must check that the normal challenger strictly prefers to enter in period 1. Staying out gets 0. Entering gets $a+(1-\mu)-\mu P_1-\mu\lambda P_2$ so entering is strictly better when

$$a+(1-\mu)-\mu(1-\pi)c/\psi_w+\mu\lambda/(1-\lambda) > 0.$$

□

Proposition 3. *Fix $a < \lambda/(1-\lambda), R, \pi$. Suppose that*

$$\psi_s < \frac{R(1-\lambda)}{\pi\lambda} \tag{6.7}$$

$$\mu < \frac{\psi_s\pi\lambda(a+1)}{R} \tag{6.8}$$

$$c > \frac{\psi_s}{1-\lambda} \tag{6.9}$$

$$\psi_w > \frac{R}{\pi\lambda(a+1)}, \frac{\mu(1-\pi)c}{a+(1-\mu)+\mu\lambda/(1-\lambda)}, \frac{R(1-\lambda)}{\pi\lambda}. \tag{6.10}$$

Then the optimum equilibrium is separating.

Proof. Sufficient conditions for the optimum to be a separating equilibrium are that there is no pooling equilibrium from Lemma 5

$$\psi_w \pi\lambda(a+1), \psi_s \pi\lambda(a+1)/\mu > R \tag{6.11}$$

and from Lemma 7 that there be an optimal separating equilibrium

$$\frac{c}{\psi_w} < \frac{a+(1-\mu)+\mu\lambda/(1-\lambda)}{\mu(1-\pi)} \tag{6.12}$$

$$\psi_s \pi\lambda/(1-\lambda) \le R \le \psi_w \pi\lambda/(1-\lambda)) \tag{6.13}$$

that gives higher utility than a simple equilibrium

$$\psi_s\left((1-\pi)(1-\psi_s/\psi_w)+\pi\lambda\right)c \ge \psi_s \pi\lambda/(1-\lambda). \tag{6.14}$$

The first inequality in 6.11 is the first inequality in 6.10. The second inequality is inequality 6.8. Inequality 6.12 is the second inequality in 6.10. The first inequality in 6.13 is inequality 6.7. The second inequality is the third inequality in 6.10. Finally inequality 6.14 is implied by 6.9. □

Existence of Pooling and Trivial Equilibria

Proposition 4. *Suppose that $R > \psi_w \pi\lambda(a+1)/(1-\lambda)$. If*

$$c > \frac{\pi}{1-\pi}\left((1-\mu)\psi_w + \mu\psi_s\right)\lambda(a+1) - 1$$

then the optimal equilibrium is pooling while the reverse strict equality implies that the optimal equilibrium is trivial.

Proof. From Lemma 7 $R > \psi_w \pi\lambda/(1-\lambda)$ implies no separating while from Lemma 5 $R > \psi_w \pi\lambda(a+1)$ implies there is no revocation in a pooling equilibrium and that utility from a pooling equilibrium is $-\left((1-\mu)\psi_w + \mu\psi_s\right)\pi\lambda(a+1)(a+1) - \pi(1+\lambda c)$. Since trivial equilibrium utility is $-1-(1-\pi(1-\lambda))c$ the second condition is that utility from separating is strictly greater than trivial. □