Distribution Agreement

Mutual Optimism and First Strike Advantages

By

Tyler Shuman
Master of Arts

Political Science

Shawn Ramirez
Advisor

Jeffrey Staton
Committee Member

David Zureick-Brown
Committee Member

Andrew Francis
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Mutual Optimism and First Strike Advantages

By

Tyler Shuman

Advisor: Shawn Ramirez, PhD

An abstract of

a thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Arts

in Political Science

2014.

Abstract

Mutual Optimism and First Strike Advantages

By Tyler Shuman

This paper uses a game theoretic model to answer the question of whether first strike advantages can be a mechanism that causes war due to mutual optimism, and whether that may be a reason why a weaker state might launch a surprise first strike against a stronger state. The motivation behind this paper is the case of the 1973 Arab-Israeli War, in which Egypt instigated war by launching a surprise attack on Israel, which had previously defeated Egypt in 1967. I theorize that war due to mutual optimism from a first strike might account for this attack.

I find that belief in a first strike advantage can be a mechanism for war due to mutual optimism, in which one state believes it will win the war because of its first strike advantage, and the other believes it will win because it has greater strength and information. The model rejects the hypothesis that war due to mutual optimism causes weaker countries to attack stronger ones, since by definition if one country is the victim of a surprise attack, it is not sufficiently optimistic to attack. However, a sufficiently high belief in a first strike advantage can cause a rational actor to attack a stronger opponent.

Mutual Optimism and First Strike Advantages

By

Tyler Shuman

Advisor: Shawn Ramirez, PhD

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Arts

in Political Science

2014.

Table of Contents

Introduction

On the Jewish holiday of Yom Kippur in 1973, after many years of tension, Egypt launched a surprise attack on Israel, beginning what came to be known as the Yom Kippur War, the October War, or the 1973 Arab-Israeli War. Feeling humiliated by a defeat by Israel in 1967, Egypt hoped to reclaim lost territory and restore its honor in the process. However, its efforts were in vain. Israels military forces managed to beat back the Egyptian troops, beyond the 1967 borders, and claim a decisive victory in less than a month.

Six years earlier, in 1967, Israel launched a surprise strike on the Egyptian airfields, destroying most of the Egyptian air force and disabling Egypts defensive capability. With that done, Israeli troops were able to move in until Egypt called for a ceasefire six days later, giving the conflict the moniker The Six-Day War. While one might attribute all of Israel's success to its surprise attack, it is important to remember that Israel was fighting a war on three fronts. Syria and Jordan surrendered as well, a fact that the Arab nations certainly could not overlook. Thus, a puzzle emerges: Given Israels impressive show of power in 1967, why did Egypt choose to instigate conflict with a clearly more powerful adversary?

I hypothesize that Egypt chose to attack Israel because it believed it would gain enough of an advantage from a surprise first strike to overcome Israel's military advantage. I investigate the possibility of war due to mutual optimism in this case, although I find that it is not appropriate for this particular case study. To solve this puzzle, I construct a game theoretic model in which the variable of interest is whether or not a weaker country has a first strike advantage when attacking a stronger country, and show equilibrium outcomes in which a belief in a first strike advantage causes the weaker country to attack the stronger, and in which war due to mutual optimism from belief in a first strike advantage can occur.

Literature

Power Parity

Inus Claude was the most influential proponent of the theory that a great imbalance of power between countries causes or is a major contributing factor of war. In his book, *Power and International Relations* (1962), he argues war breaks out when one country has substantially more power than another. The more powerful state will use its power to exploit the less powerful state and attack in order to gain concessions. This theory, called the power preponderance theory, has been the subject of much discussion in the field of international relations.

However, modern scholars seem to be in consensus in favor of the opposite, the power parity theory (Bremer, 1980; DiCicco and Levy,1999; Lemke and Werner, 1996; Moul, 2003). As one might expect, power parity hypothesizes that states that are relatively equal in power are more likely to go to war than states with a large difference in power. If two states are of similar strength, the logic goes, the outcome of war is uncertain, so disputes can only be resolved by fighting, whereas if there is a great difference of power, the weaker state will concede without having to waste resources on a costly and fruitless war. Both sets of logic make sense, but empirical evidence favors power parity.

Jonathan DiCicco and Jack Levy discuss the theoretic side of the power parity argument in “Power Shifts and Problem Shifts: The Evolution of the Power Transition Research Program.” (1999) They perceive states, and intrastate organizations, as a hierarchy of power. States lower down on the hierarchy may want to attack those at the top, but lack the means to do so. As lower states gain power, they may take advantage of this in order to attack states that were formerly above them.

Another cause of war might be that the higher states see the lower states rising, and preemptively declare war in order to fight from the most advantageous position, since they are losing relative power and will be relatively weaker, and therefore less likely to be victorious, in the future. This theory takes the self-serving logic of the power preponderance theory, but uses it to argue for power parity instead. Regardless of the story behind it, as the two following papers show, there is empirical evidence in favor of the power parity theory.

Douglas Lemke and Suzanne Werner, in “Power Parity, Commitment to Change, and War” (1996), and William Moul, in “Power Parity, Preponderance, and War Between Great Powers, 1816-1989” (2003), all empirically test the theory of power parity and find that indeed, dyads that are relatively equal in power are more likely to go to war than those that are not.

Lemke and Werner set the standard for measuring power by using a combination of Correlates of War Composite Capabilities Index and GDP to create a variable that captures what most people think of when they talk about a states power. Due to the infrequency with which this data are reported, it is very difficult to establish a dyads power difference at a single point in time. Instead, Lemke and Werner track changes in power over time, finding that as states grow more equal in power, they are more likely to go to war.

However, Lemke and Werner do not clearly define what they mean by “parity.” Their analysis is on a sliding scale of more to less equal, with likelihood of war being positively correlated with greater equality. While this supports the theory of power parity, it does not help when determining whether two states are “equal enough” to have an increased likelihood of conflict.

Moul, criticizing the methods of many previous power parity scholars, is meticulous in his analysis. He finds that dyads of roughly equal power are between twice and over a hundred times more likely to go to war than very unequal dyads, depending on variables such as alliances and possession of nuclear weapons.

One of Mouls criticisms of previous scholars is a lack of clear definition of what constitutes "parity." This is indeed a difficult line to draw; if two states are considered of unequal power if their power ratio is two to one, why might they be considered equal if the ratio were 1.9 to one? Still, Moul establishes a reasonable baseline, citing Bremer (1980), who also uses the Correlates of War data and GDP to create a value for power, and who defines parity as occurring when two countries have a ratio of power that is less than or equal to 1.5.

Moul also points out another serious problem with the way previous scholars have addressed the power parity question. Other scholars have used a unit of analysis of dyad-years, one for each possible pair of states in the world for every year their data span. This results in some very odd dyads; Ecuador and Afghanistan are unlikely to go to war not for reasons due to relative power, but because they are far apart and have little interaction at all. However, Moul finds that the relation between power parity and war still holds even after controlling for distance.

Instead of being directly stated in political science papers, power differentials usually manifest as a probability of winning a war. Since we assume that a more powerful country is more likely to win a war, translating this power difference into probability is reasonable. I utilize this conception of power, indicating that one state will win a war with probability π. If $\pi < \frac{1}{2}$, then that state is the weaker state. Thus, the balance of power is instrumental to my model, and to many formal international relations models, although it is rarely stated directly.

It is true that the correlation between power parity and war is only probabilistic. Plenty of equally powerful countries remain at peace, and sometimes wars occur between countries with a great difference of power. However, when such cases do occur, it is important to ask why. Why did Egypt attack Israel in 1973, given that the power difference in favor of Israel was clear in light of the Six Day War? To answer that, it is necessary to look at how power differences may be changed by a first strike, as well as mutual optimism.

First Strike Advantages

Much of the current literature on first strike advantages has to do with the possibility of a nuclear attack. If a state can annihilate its enemy with nuclear weapons, or at least incapacitate its military so that it cannot strike back, the state has a clear advantage in any conflict.

The same logic applies to conventional warfare, although, since the consequences of such a first strike are not so catastrophic, first strike advantages under conventional warfare conditions are less frequently discussed. Still, they follow many of the same patterns.

Robert Schelling devotes two chapters of his book *The Strategy of Conflict* (1960) to surprise attacks and first strike advantages. According to Schelling, first strike advantages are a compounding spiral of fear. State A fears State B may think that State A is going to attack, in which case State B will attack first, in which case State A will attack first. A thinks that B thinks that A thinks that B thinks that A thinks... As Schelling conceives first strike advantages, they are both a problem of communication and commitment. The two states don't know what each other are thinking, and so are forced to guess. Thus, even if both states would prefer that no war occur, they cannot communicate this to the other. Barring clear communication, they resort to guessing, and if there is a substantial advantage from a surprise attack, they assume the other will want to take advantage of it, and will strike first in self-defense. Thus, for Schelling, the power of the first strike comes not necessarily from the offensive benefit of a surprise attack, but from the defensive incentive to eliminate an enemy before he can eliminate you.

Schelling models first strike advantages as a classic Prisoner's Dilemma. It may be that no war is preferable to both sides. However, attacking is preferable to being attacked, so in equilibrium both states will attack immediately and simultaneously, resulting in a war that neither prefers. As is the case with the Prisoners Dilemma, even if clear communication were possible, the outcome would not change. Each state has the incentive to attack regardless of what its opponent does, and so cannot credibly commit to any peaceful agreement.

It is interesting to note that in this model, no first strikes actually occur. Instead, war occurs due to *anticipation* of a first strike. This will continue to be true in my model, although my model is more complicated than Schelling's simple example.

Schelling's discussion of first strike advantages primarily applies to nuclear warfare. The spiraling of fear of a first strike is characteristic of Cold War-era Soviet-American politics. From the standpoint of conventional warfare, such a fear spiral is extreme and unlikely. The lack of total nuclear devastation ameliorates crises.

Robert Jervis addressed the problem of first strike advantages and the offense-defense balance in "Cooperation Under the Security Dilemma" (1978). Like Schelling, Jervis, writing in a Cold War context, was primarily concerned with first strike advantages as they apply to nuclear warfare.

For Jervis, when circumstances favor the offense, there is an advantage to being the first to strike. If you are better off being on the offense than the defense, you certainly want to attack first. If this feeling is mutual between states, a security dilemma occurs, wherein states cannot credibly commit to not attacking each other.

An important distinction between nuclear and conventional warfare with regards to first strike advantages is the existence of a second strike. Jervis explains that the existence of second strike capability is important in combating the commitment problem posed by the existence of a first strike advantage. If a state can put resources into second strike capability, ensuring that it can still retaliate even in the event of a nuclear first strike, the advantage of the first strike is diminished.

However, in conventional warfare, a second strike capability is assumed. Conventional weapons do not have the same devastating power as nuclear weapons, and thus it is nearly impossible to completely destroy an opponents ability to fight with a single first strike. This makes conventional first strikes less powerful than nuclear ones, but eliminates the mitigating effect that a nuclear second strike capability has, since a conventional second strike is similarly weakened. Thus, Jervis's suggestion of maintaining a second strike capability in order to deter a first strike is not applicable to conventional warfare. Regardless, Jervis had a good point when he said, "Incentives to strike first will turn crises into wars."

Andrew Kydd's "Game Theory and the Spiral Model" (1997) takes a different approach to the relation between security dilemmas and first strike advantages. Like Jervis, Kydd discusses war as a result of states' fear of attack. However, Kydd focuses on the spiral model, in which states increasingly escalate their weapon capabilities in response to similar increases from their opponents, each seeking to have the most and best weapons in the name of security. At extreme points, states possess such large quantities of weapons and are so fearful of each other that they cannot control reactions and war can be started by a trigger-happy soldier in charge of a nuclear missile.

For Kydd, a first strike advantage doesnt cause additional insecurity. Indeed, a first strike advantage is one thing that can *stop* a spiral. A first strike advantage makes each state want to strike immediately instead of amassing arms over an extended time to the point where the situation becomes extreme and out of control. Thus, under Kydd's model, a first strike advantage would cut short the alternating process of weapon accumulation.

In "Rationalist Explanations for War," (1995) James Fearon formally incorporates the existence of a possible first strike advantage in his bargaining model of war as one factor that could potentially shrink the bargaining range. He talks about first strike advantages as causing a commitment problem. His model involves a state, A, with several probabilities of victory: p_f if it attacks first, p_s if it is attacked, and p if A and its adversary attack at the same time. The bargaining range only exists between p_f minus A's costs, and p_s plus B's costs, on the linear model. Thus as p_f increases and p_s decreases, the bargaining range shrinks or vanishes entirely.

The elimination of Fearon's bargaining range does not mean that there are no bargains that would be preferable to war. By definition, only one side can strike first, so there is always a theoretical bargain. However, when such a first strike advantage exists, as Jervis discussed previously, commitments to any such bargain will not be credible, because the incentives of a first strike advantage will always be superior to any bargain, although the bargain may be superior to a war outcome.

Fearon believes that first strike advantages are not instrumental to war as a whole, arguing that such an advantage is only key when leaders believe that "the first strike advantage is so great that regardless of how we resolve any diplomatic issues between us, one side will always want to attack the other in an effort to gain the (huge) advantage of going first." Rather, Fearon believes that first strike advantages may "exacerbate other causes of war by narrowing the bargaining range." While this is a reasonable conclusion, I will model in this paper a situation in which first strike advantages are indeed central, and provide a case study, that of the Arab-Israeli conflicts, that demonstrates such a situation historically.

Mutual Optimism

This paper builds primarily on two other works on mutual optimism: "Mutual Optimism and War," by Fey and Ramsay, and "Mutual Optimism as a Rationalist Explanation of War," by Branislav Slantchev and Ahmer Tarar.

In "Mutual Optimism and War," Fey and Ramsay argue that, assuming rational actors, there is no equilibrium in which war due to mutual optimism occurs. They claim that "if both sides are willing to fight, each side should infer that they have either underestimated the strength of the opponent or overestimated their own strength." (Fey and Ramsay, 2007) That is, both sides should immediately cease fighting once they have viewed their opponent's willingness to fight, thus leading to peaceful settlement.

There are a few important aspects of war omitted from this model. The first, and most relevant, is that Fey and Ramsay explicitly exclude the possibility of a surprise attack. In real life, it is not reasonable to assume that "neither the expected payoff to war or the expected outcome of negotiations depends on the choice of actions by the countries," as Fey and Ramsay state. While perhaps war due to mutual optimism is impossible under some very specific and limited circumstances, it is an oversight to neglect the choices of the countries involved when modeling war.

Furthermore, Fey and Ramsay assume that one actor can unilaterally stop the war by refusing to fight once it has observed its opponent's willingness and realized it has miscalculated the odds. This is not true, especially when these decisions are made "on the brink of war," when troops may already be mobilized and orders given. Even if plans aren't already in motion when this mutual observation of willingness occurs, by refusing to fight, one country leaves itself vulnerable. This will change the expected utility of war at that moment, making mutual war more likely and unilateral war a virtual certainty.

Slantchev and Tarar offer a rebuttal to Fey and Ramsay, specifically addressing the latter point. They show how it invalidates Fey and Ramsays model, and also point out other oversights in the way Fey and Ramsay conceive mutual optimism. Their largest disagreement is that Fey and Ramsay ignore "the causal mechanisms of war," focusing too much on what the opponents are thinking and not enough on what they are actually doing (Slantchev and Tarar, 2011).

Slantchev and Tarar also contest the standard reasoning behind why wars due to mutual optimism occur. Typically, it is assumed that wars due to mutual optimism occur when both sides demand too much, thus eliminating the bargaining range. However, Slantchev and Tarar argue that the bargaining range need not be eliminated for a war due to mutual optimism. Instead, the states must believe that they will gain more from fighting than from bargaining. Thus, a mutually optimistic war can be strategic.

In Slantchev and Tarar's model, the mechanism of war due to mutual optimism is that of risk and reward. One state, the satisfied state, chooses to offer one of two bargains. One will satisfy any opponent regardless of strength, and one will only satisfy a weak opponent. If the satisfied state offers a strong opponent the weak opponents bargain, the opponent will reject and a war will occur. The optimism of the satisfied state is shown by its smaller offer, and the optimism of the opponent is shown by rejecting the offer.

Slantchev and Tarar assert that Fey and Ramsay put unreasonable restrictions on their model, stating "It is no surprise that [Fey and Ramsay] find that mutual optimism cannot cause war; after all, they have ruled out the very mechanisms through which mutual optimism is theorized to do so." (Slantchev and Tarar, 2011) They also disagree with the assumption that one state can unilaterally impose peace on another, as mentioned above. In real life, this is usually not the case; one state may attack another, bringing about war, and the other may be powerless to stop it. However, this is then not a war due to mutual optimism, but just a regular war.

Slantchev and Tarar do not discuss the effects of first strike advantages either. For all their criticism of Fey and Ramsay for ignoring the "causal mechanisms" of war, this is somewhat of an oversight. Slantchev and Tarar propose one mechanism for war due to mutual optimism, that of risk and reward. I propose another: the potential for a first strike advantage causing a state to overestimate its own chances of success.

The Arab-Israeli Conflicts

The 1967 Arab-Israeli War: Origins and Consequences calls the Six Day War "the swiftest and most spectacular military victory of [Israel's] entire war-filled history." After launching a surprise first strike on the Egyptian air force, Israel managed to win a war on three separate fronts, conquering a large amount of politically-valuable territory in the process. While the victory in the Six Day War is a particularly striking event in Israels history, it is simultaneously indicative of relations and military strategy in that area of the world during Israel's lifetime.

The existence of Israel has been highly contentious since its creation. Israel claims to be peaceful, to fight only defensively, and only when there is no other choice. The neighboring Arab nations have a different view. They see Israel as an aggressor, forcing Palestinians off their land, constantly seizing territory and expanding its boundaries to the detriment of neighboring states. These opinions hold constant through the 1967 war.

Israel insists that it was fighting a war of self-defense. It knew that its neighbors were allying against it. When Egypt began mobilizing troops in the Sinai Peninsula, Israel reacted to the threat and launched a preemptive first strike on the Egyptian air force on June 5th, during a particularly vulnerable time of day, destroying nearly all the Egyptian planes (Shlaim and Louis, 2012). With Egypt militarily disabled, the Israel Defense Force was able to seize Gaza and force a retreat of the Egyptian forces that had been mobilized in the Sinai Peninsula. That done, the IDF turned its attention to Syria and Jordan, Egypt's allies, and conquered the West Bank and Golan Heights as well, before the Arab states accepted a ceasefire and ended the war on June 10th. All in the name of self defense, of course. There is no doubt that Egypt, Syria, and Jordan were preparing for war against Israel, and that Israel is safer with a "buffer zone" between itself and its enemies.

However, the Arab states interpret events differently. They see Israel as expansionist and war-hungry. They point out that Israel was already prepared for war in 1967, that Israel struck first, and maintain that the conquered territory was Israel's goal all along. Egypts mobilization of troops was meant to send a signal of resolve to other Arab states, not to provoke war with Israel (Shlaim and Louis, 2012). These attitudes will go on to influence actions during the 1973 war as well.

Primarily of note is Israel's use of a very successful first strike. By striking first when Egypt was vulnerable, Israel managed to wreak more havoc than anyone anticipated. That same day, Egypt pulled out of the Sinai Peninsula and agreed to a ceasefire that evening, meaning that Egypt's active participation in the Six Day War lasted less than twenty-four hours (Shlaim and Louis, 2012). The remaining five days were used to beat back Syria and Jordan.

Egyptian leaders were shocked by the effect of the attack. While they had anticipated war with Israel, they were unprepared for an attack of that magnitude, and as a result were unable to respond. (Shlaim and Louis, 2012). It is possible that Israel would still have won without the advantage of that first strike, but not that easily or quickly. Just looking at the bare facts, there is no doubt that Israel's use of a surprise first strike was advantageous.

The Arab nations felt humiliated by this defeat, and refused negotiations, so Israel kept the territory it had won, and the Arab nations, particularly Egypt, nursed wounded pride and a serious blow to the national ego (Aker, 1985). The Egyptian people began to think of their country as weak, technologically backwards, and strategically inferior. For a country with such a strong national identity, this was painful and disconcerting.

Egypt carried this shame and resentment throughout several smaller military skirmishes between 1967 and 1973 that together made up the War of Attrition, which ultimately resulted in stalemate. Recovering its pride was one motivation for Egypt in entering the 1973 war (Aker, 1985). Recapturing territory lost in 1967 was also a factor, but Egypt wanted to be seen as a serious military power again. For Egypt, this was not just about land and politics. It was about honor.

During this time, the Arab states were again preparing for war. Egyptian military strategists had been studying IDF tactics and strategies. The IDF had very offensive-focused strategies designed to easily penetrate and capture territory. A surprise first strike was a key strategy in their arsenal, as Egypt had seen in 1967 (Asher, 2003). Egypt and Syria, once again planning to go to war with Israel, were prepared to turn Israel's own tactics against it. The Soviet Union provided Egypt and Syria with newer, more advanced tanks, and an air defense system specifically designed to deal with the Israeli air force, which had proven so deadly in 1967. (Aker, 1985). However, they took great pains to hide this advantage from Israel. Reports that the Soviet weapons were of poor quality were leaked to Israeli intelligence, as were rumors that the antiaircraft weapons were poorly maintained, and that spare parts were in short supply. (Handel, 1977). This lulled Israel into a false sense of security. They fully believed the Arab states were too incompetent to launch an attack in 1973, although they anticipated an attack might occur a few years later. This belief was supported by previous declarations of war from Egypt, which it failed to follow through on. (Handel, 1977)

The date of the attack, October 6th, was also calculated to provide the benefit of surprise. The attack occurred on the Jewish holiday of Yom Kippur, when most of Israel would spend the day fasting and praying. This also happened to coincide with the Muslim holy month, Ramadan, during which the Arab states would not be expected to go to war. Egypt sent troops into the Israeli-controlled Sinai Peninsula, while Syria simultaneously entered the Golan Heights. For three days, Egypt and Syria made great progress in retaking the land. However, this advantage did not last. Israel marshaled troops, first driving the Syrians out of the Golan Heights, and then turning its attention to the Sinai. Israel divided the Egyptian troops, cutting them off from each other and from supplies and other aid. Several ceasefires were called, although Israel continued to advance, claiming each time that Egypt had been the one to violate the agreements. By October 25th, when the United States finally intervened to restrain Israel and a durable ceasefire was called, Israel had nearly completely encircled the Egyptian Third Army. Israel allowed the Egyptians to retreat, maintaining control of all its post-1967 territory, and peacekeepers were sent in to maintain order. (Aker, 1985.)

Egypt and Syria's motivations and beliefs are clear. The had a strong incentive to go to war in order to salvage their pride, wounded in 1967. That is to say, their benefits from war were higher, because they included intangible benefits from things like reputation and nationalistic pride. They had witnessed the strength of the IDF during the Six Day War. In addition to the surprise attack, the IDF routed ground troops in the Sinai Peninsula and Golan Heights. Still, they thought that their new weapons, provided by the Soviets, would leave them better equipped to deal with Israeli troops. However, I argue that the belief most instrumental to the decision to go to war was their belief in a first strike advantage. In their studies of IDF tactics, they came to believe that Israel's surprise first strike was largely responsible for its success in 1967. By adopting this tactic, the Arab states believed they would have enough advantage to overcome superior Israeli forces. This belief proved to be incorrect.

Below, I show, using a formal model, the conditions under which a weaker state such as Egypt would attack a stronger state, Israel. This model hinges on the belief in a first strike advantage, and shows both a mechanism by which wars due to mutual optimism can occur, and an explanation for first strikes by weaker states against stronger states.

Concepts

While both mutual optimism and first strike advantages have been discussed extensively in previous literature, definitions for these concepts are often vague and can differ between authors. Below are brief definitions of mutual optimism and first strike advantages as I will be using throughout this paper.

Mark Fey and Kristopher Ramsay define war due to mutual optimism as "war due to inconsistent beliefs." They later clarify this definition to say that "if [leaders] expectations are inconsistent in that both antagonists think their side will be better off fighting a war, the argument goes, then neither side would be willing to participate in a peacefully negotiated settlement." (Fey and Ramsay, 2007) This is an old argument, first introduced by Geoffrey Blainey in 1988, and it the one I will use. A war due to mutual optimism occurs when two states choose to fight instead of negotiate because each of them believes that they will win. This is distinguished from a war in which both sides choose to fight because they believe it is better than the alternative of risking being the victim of a surprise attack. In such a case, one or both sides might not be confident in their ability to achieve victory, but be hoping to increase that chance as much as possible. Such a war would then not be considered due to mutual optimism.

The term "first strike advantage" is most often used in discussions of nuclear warfare, in which a nuclear first strike may have the capacity to eliminate an opponent's ability to retaliate. Conventional weaponry is rarely so powerful. Thus, a first strike advantage is more limited than in nuclear warfare, and is usually based on the element of surprise, allowing one state to damage another's military, supplies, or infrastructure before a response can be formed. For the purposes of this paper, I define a first strike to be a surprise attack against an opponent that is not prepared for war. Firing the first shot on a field lined with soldiers from both states is not enough to qualify. The victim of the first strike may be in the midst of war preparations, but it must not be ready to launch an attack of its own. The degree of advantage from a conventional first strike depends on many factors. For instance, in a lengthy war due to political disagreements or religious differences, a first strike may initially cripple an enemy, but as time goes on, the adversary will have time to recover lost forces and again regain their position. However, a first strike that disables the opponent's defensive capabilities is likely to be very effective. Additionally, it is generally easier to defend territory than to capture it, so if attacking first allows a state to conquer territory while its opponent is still organizing troops, the first state may have an advantage when the second attempts to reclaim the territory.

Model Setup

This is a two player game with both simultaneous and sequential elements. The two players are the satisfied state, denoted S, and the dissatisfied state, denoted D. Nature is also a player, determining whether or not there exists a first strike advantage, denoted α, for D. To begin, Nature chooses whether or not there is a first strike advantage for D; that is, whether α is a number between 1 and $\frac{1}{\pi}$, and thus boosts D's chances of winning a war, or whether α equals 1, and therefore has no effect. I assume that, if a first strike advantage exists, it is large enough so that $\alpha\pi \geq 1-\alpha\pi$, meaning that the advantage is enough to make D more likely to win than S. S, being the satisfied state with presumably more resources and intelligence, knows whether or not such a first strike advantage exists for D. D does not know which state of the world it is in, but has prior beliefs A and 1-A that a first strike advantage exists and doesn't exist, respectively, regardless of any distribution. D chooses whether to initiate conflict with S and challenge the status quo, or not initiate and maintain the status quo. If D does not initiate, S receives a status quo payoff of 0, and D receives a small payoff ϵ, which is used to indicate that there is some cost to initiating conflict, even if both states ultimately decide not to attack.

If D chooses to initiate, both sides simultaneously decide whether or not to attack the other. If either state attacks, a war occurs. Each side has a probability of winning, which is π for D and $(1-\pi)$ for S, where $0 \leq \pi \leq 1$. For the purposes of this game, we will assume that the value of π, and therefore $(1-\pi)$, are known to both S and D.

If both states attack, a "fair" fight occurs, with neither side having a first strike advantage, and the utility for the two states is their predicted probability of success minus the cost of fighting, denoted $c > 0$. If one side chooses to attack and the other does not, the state that attacks receives a first strike advantage. This is α for D and β for S.[1] This number is multiplied by the attacker's probability of success, increasing it. By attacking when the opponent does not, states increase their probability of success, and therefore their utility, but they still have to pay the same costs of war as in the fair fight outcome. While the existence of a first strike advantage for D is dependent on the state of the world determined by Nature, I assume that, if it does exist, both states know it's value. That is, α is either equal to 1 or to some value between 1 and $\frac{1}{\pi}$, which is known to both states, although only S knows whether the first strike advantage exists or not. β is consistent in all states, in order to reduce uncertainty and extraneous variables. Thus it may be taken that the value of β is known.

If neither side attacks, war does not occur, and both sides receive a payoff of zero. The path of play and the payoffs for each outcome are best shown by figure 1 below.

[1]Because the probability of winning must always be between zero and one, α is bounded by $1 \leq \alpha \leq \frac{1}{\pi}$, and beta is bounded by $1 \leq \beta \leq \frac{1}{1-\pi}$.

(Insert figure 1 here.)

While the value of all of these parameters is known, they are still variable, and there are many things that could contribute to a certain variable having a higher or lower value. The initial probability of winning a war for either side depends on weapons, strategy, inside information, leadership, and aid from allies. The value of a first strike advantage depends on the type of attack used, the weapons and resources available, and the defenses the opponent has set up. Costs of war can also vary due to weapons, information, and alliances. Additionally, there may be intangible benefits to fighting a war that mitigate material costs. For instance, countries may fight to regain pride after a humiliating previous defeat, or they may receive a boost in morale and support for the leader by uniting the populace against a common enemy.

Results

As the more informed state, S can condition its actions on the state of the world. Thus, it has four possible pure strategies: always attack regardless of the state of the world, never attack, attack when D has a first strike advantage and don't attack when D doesn't have a first strike advantage, and don't attack when D has a first strike advantage and do attack when it doesn't. By fixing each of these four strategies, we can find all pure strategy Perfect Bayesian Equilibria. I use W to denote the probability that D will attack, and V_1 and V_2 to denote the probability that S will attack when D does and does not have a first strike advantage, respectively. Since I am only interested in pure strategy equilibria, W, V_1 and V_2 will have values of 1 or 0.

If S always attacks, then its utility from attacking must be greater than or equal to its utility from not attacking in both states of the world. $W(1-\pi-c)+(1-W)(\beta(1-\pi)-c) \geq W(1-\alpha\pi-c)$, and $W(1-\pi-c)+(1-W)(\beta(1-\pi)-c) \geq W(1-\pi-c)$. Simplified, S's incentive compatibility conditions are thus: $(1-W)(\beta(1-\pi)-c) \geq 0$, and $\alpha \geq 1 - \frac{1-W}{W\pi}(\beta(1-\pi)-c)$. This shows that α is greater than or equal to a number less than or equal to one, which is a premise of the model. Thus, D acquires no new information from S's strategy, and resorts to its priors A when determining the probability that it has a first strike advantage. With no new information, D compares its expected utility for attacking and not attacking, given S's strategy of attacking in either state of the world, chooses its best response. In this case, D's expected utility is $\pi - c$ for attacking and $1 - \beta(1-\pi) - c$ for not attacking, so attacking is D's best response. Plugging a 1 for W into S's incentive compatibility conditions yields $0 \geq 0$ and $\alpha \geq 1$, both of which are innately true, so Always Attack is a best response for S, and (Attack, Always Attack) is PBE for this subgame. Moving up the game tree, D must decide whether or not to initiate conflict depending on whether or not its utility at the (Attack, Always Attack) outcome is greater than the utility of not initiating. Thus, D will initiate if $\pi - c \geq \epsilon$ and not initiate otherwise.

Proposition 1: If $\pi - c \geq \epsilon$, D will Initiate and Attack with beliefs A and S will Aways Attack, and if $\pi - c \leq \epsilon$, D will not initiate.

Since both states are simultaneously attacking in this equilibrium, it may be tempting to label the resulting war as due to mutual optimism. However, this is not the case, as this outcome does not depend on the beliefs of the two states. The only limit is whether the costs of war are high enough that D will choose not to initiate. Both states will attack regardless of S's knowledge of the state of the world or D's belief in its own first strike advantage. This outcome is similar to Schelling's Prisoner's Dilemma-type model of first strike advantages, in which each side attacks because it fears being the victim of a surprise attack. Nothing is known about each state's estimation of its own or its opponent's probability of winning the war. A state may think that it is likely to lose, and would prefer negotiating, but it would prefer to increase its chance of winning as much as possible, knowing that it cannot trust its opponent to credibly commit to peace. Thus, simultaneous war occurs, but not due to mutual optimism.

If S never attacks, its incentive compatibility conditions are the same as when it always attacks, except with the opposite signs. $(1 - W)(\beta(1 - \pi) - c) \leq 0$, and $\alpha \leq 1 - \frac{1-W}{W\pi}(\beta(1 - \pi) - c)$. If the first inequality is true, than the right side of the second inequality is 1 if W=1 and is undefined when W=0, so D learns nothing from S's strategy and resorts to its priors, A. Next, D compares its expected utility for attacking to not attacking. If D attacks, it will get $\alpha\pi - c$ with probability A and $\pi - c$ with probability 1-A. If D does not attack, it will get 0, so D will only attack if $A(\alpha\pi - c) + (1 - A)(\pi - c) \geq 0$, or rather, if $A \geq \frac{-(\pi-c)}{\pi(\alpha-1)}$. The right side of this inequality is negative if $c < \pi$ and positive if $c > \pi$, so unless the costs of war are greater than the expected benefits, D will attack. However, plugging W=1 into S's incentive compatibility conditions yields $0 \leq 0$ and $\alpha \leq 1$, which is only true if D does not have a first strike advantage. If D does have a first strike advantage, S would prefer to attack than not attack, so Never Attack is not a best response for S if D attacks. However, if the costs of war are large enough that D does not attack, plugging in W=0 gives $\beta(1 - \pi) - c$, which is true by the original incentive compatibility conditions. The second incentive compatibility condition, rearranged, becomes $W\pi\alpha \leq W\pi - (1 - W)(\beta(1 - \pi) - c)$. Substituting W= 0 into this equation gives $0 \leq -(\beta(1 - \pi) - c)$, which is also true, and Never Attack is a best response for S if D does not attack. But in the case that neither side decides to attack, D will choose not to initiate, since initiating is costly. The payoff ϵ is greater than the equilibrium payoff of the simultaneous subgame, (0,0).

Proposition 2: If $A \leq \frac{-(\pi-c)}{\pi(\alpha-1)}$*, D will not Initiate and not Attack with beliefs A, and S will Never Attack.*

This is the outcome in which war is too costly for either side to attack. Furthermore, D's belief in its own first strike advantage is too low for it to risk going to war; if D does not have a first strike advantage, it would prefer not to attack, and D believes that it is unlikely that it has such an advantage. Therefore, both sides would prefer to refrain from attacking, rather than fighting a costly war. Given that the payoff for both sides for not attacking is 0, D will choose not to Initiate, since Initiating is costly, and by not Initiating, D receives the payoff $\epsilon > 0$.

While I am primarily interested in conventional warfare, this equilibrium occurs when the costs of war are very high, such as when nuclear weapons are an option. Thus, one possible way to introduce nuclear weapons into this model is as a mechanism that increases the cost of war. Such a discussion is beyond the scope of this paper, but I would be remiss to ignore it completely.

If S attacks only when D's first strike advantage exists, its incentive compatibility conditions are $(1-W)(\beta(1-\pi)-c) \leq 0$, and $\alpha \geq 1 - \frac{1-W}{W\pi}(\beta(1-\pi)-c)$. The second incentive compatibility condition is the same as when S always attacks, and therefore D cannot learn anything about α for the same reasons as when S's strategy is Always Attack, and so resorts to its priors, A. D will always get $\pi - c$ if it attacks, and will get $(1-\beta(1-\pi)-c)$ with probability A and 0 with probability 1-A if it does not. Attack is D's best response if $\pi - c \geq A(1-\beta(1-\pi)-c)$, or $A \geq \frac{\pi-c}{1-\beta(1-\pi)-c}$. The denominator is negative if $\pi > c$, so D will attack if the costs are low relative to the benefits. Plugging in W=1 into S's incentive compatibility conditions, we get $0 \leq 0$, and $\alpha \geq 1$, both of which are true, so if D attacks, S's strategy is still a best response. If the costs of war are substantially more than the benefits, α may be less than $\frac{\pi-c}{1-\beta(1-\pi)-c}$, in which case D will not attack. Substituting W=0 into S's incentive compatibility conditions gives $(\beta(1-\pi)-c) \leq 0$, and, rearranging the second condition as before, $0 \geq -(\beta(1-\pi)-c)$. The first is true by S's original compatibility conditions, but the second is not, so this strategy is not a best response for S if D does not attack. Moving up the game tree, initiating is a best response if the expected utility of initiating, given the equilibrium of the subgame, is greater than ϵ. If not, D will not initiate.

Proposition 3: If $\pi - c \geq \epsilon$ and $A \geq \frac{\pi-c}{1-\beta(1-\pi)-c}$ D will Initiate and Attack with beliefs A, and S will Attack only if $\alpha > 1$, and if $\pi - c \leq \epsilon$ and $A \geq \frac{\pi-c}{1-\beta(1-\pi)-c}$, D will not initiate.

This equilibrium is most interesting because it shows both outcomes of interest, depending on the state of the world. Mutual optimism is apparent when D attacks and S attacks only when a first strike advantage exists for D. If the first strike advantage exists, both states will have entered into war due to a belief that they will win: S believes it will win because it knows about the state of the world, which can be translated to better intelligence and information in the real world, and D believes it will win because of its first strike advantage. What distinguishes this equilibrium from that in which S always attacks is that if S always attacks, D's response is not dependent on A. If S only attacks when D has a first strike advantage, D will only attack if A is high enough. D's decision attack or not attack in the lower subgame are contingent on D's belief that it is in the world in which it has a first strike advantage. When D's decision is based on A, and it chooses to initiate and attack because it believes a first strike advantage exists, war due to mutual optimism occurs. However, it is important to note that in this equilibrium, *no actual first strike occurs.* If one side managed to attack first, then it would not be a war due to mutual optimism, because one state would not be optimistic enough to attack. Rather, it is the *belief* that such an advantage exists that drives D to attack when it otherwise would not do so. In this way, the belief in a first strike advantage, regardless of whether or not it materializes, drives war due to mutual optimism.

Thus, while the expectation of a first strike advantage can be a mechanism by which war due to mutual optimism occurs, it is impossible to observe in the real world. Perhaps, if one were to interview a state leader and ask why he or she entered a war against a certain opponent, one might be able to determine whether a belief in a first strike advantage had any effect, but political scientists rarely get the opportunity to question political leaders about their motivations. This is even less likely when the conflict in question occurred many years ago and the leaders are no longer available for comment. Thus, while theoretically interesting to political scientists, there is little external validity.

However, this outcome of war due to mutual optimism only occurs in one state of the world, that in which D has a first strike advantage, which causes S to choose to attack. The other outcome of this equilibrium occurs when D does not have a first strike advantage, but believes it does. S does not attack, and D does, although there is no benefit to D's first strike. This shows a situation in which a state, driven by the belief in a first strike advantage, attacks a more powerful adversary. An example of this is the 1973 Arab-Israeli War.

In the period between 1967 and 1973, Israel is the "satisfied state." Following the Six-Day War, Israel has achieved its goal of a buffer zone between itself and its Arab neighbors. It controls all of the contested territory in the region, and by achieving such swift, decisive victory in 1967, it does not believe it is in immediate danger of attack. Israel might possibly be able to achieve even better security by employing absolute war tactics in order to completely eliminate its adversaries, but such actions would have major repercussions from the international community. Israel proved it did not want to employ such tactics during the Six-Day War, so Israel has little to gain from war.

Meanwhile, any of the Arab states could be considered the "dissatisfied state" in this two-player game, but I will focus on Egypt, given its position as a particularly influential instigator in the Arab-Israeli conflicts. Egypt lost a lot in the Six-Day War. In addition to the loss of territory, Egyptians also lost feelings of pride, superiority, and accomplishment. This blow to the national ego was as important as the strategic and economic value of the Sinai Peninsula in Egypt's decision to go to war with Israel again. The Egyptians were fighting not just for territory, but also for honor.

In the years after the Six-Day War, Egypt thoroughly examined the Israel Defense Forces' strategies, incorporating these into their own military manuals. One of their most important discoveries was the IDF's belief that "the basic tactic for defeating the enemy is the offensive, especially in a surprise attack." (Asher, 2009) This belief became a foundation of Egyptian strategy during the 1973 war. Because a surprise attack had been so important to Israel's victory in 1967, disabling the Egyptian air force and limiting Egypt's defensive capabilities, Egypt believed that a successful surprise attack was vital in defeating Israel.

Israel, on the other hand, believed that Egypt would not go to war "until certain basic conditions were met, such as Arab air superiority and strategic pan-Arab cooperation." (Asher, 2009) Clearly, Israel did not believe that Egypt could have much to gain from a first strike. Indeed, having used the surprise attack strategy so effectively in 1967, the IDF would be prepared for a similar strategy from the Egyptians if they thought Egypt had anything to gain from it.

Thus, we have the game set-up. Israel is the satisfied state and Egypt the dissatisfied state. Israel, having superior information about first strike advantages due to having made use of them in the past, knows whether or not Egypt will get any benefit from striking first (and knows that it will not.) While Israel would logically attack if it believed Egypt would gain a significant advantage from striking first, it knows that Egypt would not, and so refrains from attacking despite Egypt's threats, which I consider "initiation" of the conflict. Egypt, meanwhile, believes that it will have a first strike advantage. That is, A is high. So Egypt initiates conflict by making threats, and then attacks Israel. Israel, following its strategy of only attacking if Egypt has a first strike advantage, does not attack, and is the victim of a surprise first strike during the Jewish holiday of Yom Kippur. However, this surprise attack does little to affect the overall war outcome. Israel, having a stronger and more tactically advanced military, still manages to defeat Egypt, as well as its ally, Syria.

The 1973 Arab-Israeli war is *not* an example of a war due to mutual optimism because of the belief of a first strike advantage. As stated above, if a surprise attack occurs, the war is not due to mutual optimism, because one state was not optimistic enough to attack, and mutually optimistic wars due to the belief in a first strike advantage are not outwardly observable. However, this war is another possible outcome of a model that takes first strike advantages into account, an observable outcome that supports the model.

S's final potential strategy is to attack only if D's first strike advantage does not exist. S's incentive compatibility conditions are $(1-W)(\beta(1-\pi)-c) \geq 0$, and $\alpha \leq 1 - \frac{1-W}{W\pi}(\beta(1-\pi)-c)$. The second incentive compatibility condition is the same as when S never attacks, so D learns nothing about α for the same reason and resorts to its priors, A. If D attacks, it gets $\alpha\pi - c$ with probability A and $\pi - c$ with probability 1-A. If D doesn't attack, it gets 0 with probability A and $(1-\beta(1-\pi)-c)$ with probability A. Therefore, D attacks when $A \geq \frac{1-\beta(1-\pi)-\pi}{\alpha\pi-\pi+1-\beta(1-\pi)-c}$ and doesn't attack when $A \leq \frac{1-\beta(1-\pi)-\pi}{\alpha\pi-\pi+1-\beta(1-\pi)-c}$. In the case when D does attack, substituting W=1 into S's incentive compatibility conditions gives $0 \geq 0$ and $\alpha \leq 1$. The first is true, but the second is true only if D does not have a first strike advantage. If D does have a first strike advantage, S would prefer a different strategy, so this strategy is not a best response for S if D attacks. If D doesn't attack, substituting W=0 into the rearranged incentive compatibility conditions gives $\beta(1-\pi)-c \geq 0$ and $0 \leq -(\beta(1-\pi)-c)$. The first is true by S's original incentive compatibility conditions, but in the second is not, so this strategy is never a best response for S. Therefore, there are no equilibria in which S only attacks if D does not have a first strike advantage.

Alternate Hypotheses

The 1973 Arab-Israeli War shows many attributes consistent with my model. It shows a weaker country, Egypt, launching a first strike against a stronger country, Israel, in a world in which Egypt did not have an advantage from striking first, but believed it did. In many ways, it answers the question of why a weaker country might attack a stronger one, using a belief in a first strike advantage as a mechanism.

However, it is not a perfect fit. To begin, it is difficult to determine what action or actions constituted "Initiating" on Egypt's part. Egypt did make threats to Israel between 1970 and 1973, but it repeatedly failed to follow through, and therefore future threats were ignored. Furthermore, Egypt and its allies went to great lengths to conceal its military activities and spread misinformation in order to increase the impact of its first strike. Therefore, it is hard to say that Israel saw Egypt's initiation and made a conscious choice not to attack, as the model specifies.

If the 1973 war fit my model perfectly, I should have seen a case in which Egypt made an explicit threat to Israel. Israel would then have ignored the threat and not prepared for war because it did not believe that it mattered whether it attacked Egypt or not, consistent with S's incentive compatibility conditions for proposition 3. The model requires a conscious choice by Israel that is not really present in the historical case.

Furthermore, while there is evidence that suggests that a surprise first strike was an integral part of Egypt's strategy for the 1973 war, it may not have been the driving force. Egypt felt humiliated after losing the 1967 war, and the 1973 war was an attempt to recover its pride as well as the territory it had lost. It's possible that these incentives on their own might have been enough to cause Egypt to attack, regardless of its belief in its first strike advantage.

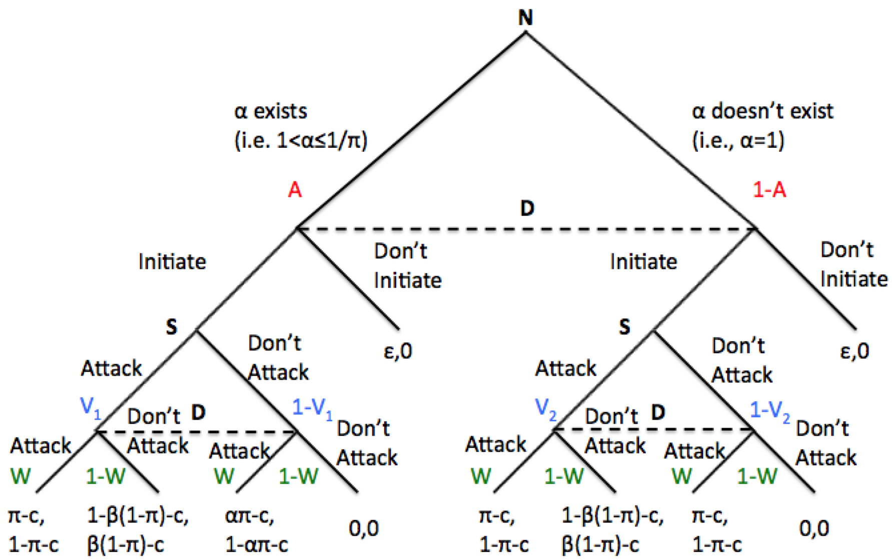
Conclusion

My central question was, "Does the belief in a first strike advantage encourage a weaker country to attack a militarily stronger opponent?" A subquestion was, "Can the belief in a first strike advantage create war due to mutual optimism?" The answer to both is yes, although the connection is not as clear as I originally anticipated. I originally thought that war due to mutual optimism because of a belief in a first strike advantage might explain a weaker country's attack. While the concepts are sound, the causality was flawed. No surprise first strike can occur in a war due to mutual optimism, because the very presence of a surprise attack means that one country was not sufficiently optimistic to go to war.

However, by modeling a situation in which the weaker country may or may not have a first strike advantage, I show that there exists a perfect Bayesian equilibrium in which either war occurs due to mutual optimism caused by the belief in a first strike advantage, or in which a weaker country attacks a stronger country because of its belief in a first strike advantage. The two outcomes are dependent on whether or not the first strike advantage actually exists, as well as the weaker country's belief in it. This provides a possible explanation for Egypt's attack on Israel in 1973.

There are other questions related to this topic that I would like to explore. One central point of my argument is that the weaker country's first strike advantage is a way to decrease the power imbalance between the two countries. However, in order to keep the model tractable, only the weaker country's first strike advantage is varied between states of the world. A more complete model would also vary the stronger country's first strike advantage.

Surprise attacks such as the ones that occurred during the Arab-Israeli wars are fairly uncommon in international conflict. It is more often the case that both states are preparing for war simultaneously. Why is this? Is it, as Schelling suggests, a result of a Prisoner's Dilemma in which both sides attack in order to avoid the "sucker's payoff" of being the victim of a first strike? If so, one would think wars would occur with more frequency than they do. Is it that first strikes are rarely advantageous? One can conceive any number of hypothetical situations in which a surprise attack would be beneficial. While academics such as Schelling, Jervis, and Fearon frequently theorize about first strike advantages, I believe the subject would benefit from more case studies and a model that accounts for variable first strike advantages on the part of both or all states involved in a conflict.



N
α exists
(i.e. 1<α≤1/π)
α doesn't exist
(i.e., α=1)
A
1-A
D
Initiate
Don't
Initiate
Initiate
Don't
Initiate
S
Don't
Attack
ε,0
S
Don't
Attack
ε,0
Attack
Attack
V1
Don't D
Attack
1-V1
Don't
Attack
V2
Don't D
Attack
1-V2
Don't
Attack
Attack
W
1-W
Attack
W
1-W
Attack
W
1-W
Attack
W
1-W
π-c,
1-π-c
1-β(1-π)-c,
β(1-π)-c
απ-c,
1-απ-c
0,0
π-c,
1-π-c
1-β(1-π)-c,
β(1-π)-c
π-c,
1-π-c
0,0

Figure 1

Appendix

There are four possible strategies for S: always attack, never attack, attack when a first strike advantage exists and don't attack when it doesn't, and attack when it doesn't exist and don't attack when it does. To find the pure strategy perfect Bayesian equilibria, I first fix each of S's strategies and then determine D's beliefs and optimal actions.

Strategy 1: S always attacks.

In this case, $V_1 = 1$ and $V_2 = 1$. Since S will attack if its utility from attacking is greater than its utility from not attacking, it must be the case that $W(1-\pi-c)+(1-W)(\beta(1-\pi)-c) \geq W(1-\alpha\pi-c)$, and $W(1-\pi-c)+(1-W)(\beta(1-\pi)-c) \geq W(1-\pi-c)$. Reduced, these two inequalities are S's incentive compatibility conditions:

(i) $(1-W)(\beta(1-\pi)-c) \geq 0$, and

(ii) $\alpha \geq 1 - \frac{1-W}{W\pi}(\beta(1-\pi)-c)$

If $(\beta(1-\pi)-c) \geq 0$, as it must be for S to always attack, (see (i)), then $\alpha \geq 1$, which is a premise of the model. Therefore, D learns nothing new and resorts to its prior beliefs, A.

Next, compare D's expected utility from attacking and not attacking, given S's strategy.

$$EU_D(Attack|SAlwaysAttacks) = \pi - c$$

$$EU_D(Don'tAttack|SAlwaysAttacks) = 1 - \beta(1-\pi) - c$$

From this, it is clear that Attack is D's best response. Next, verify that Always Attack is best response for S, given D's strategy of Attack. Therefore, W = 1. We can plug this into S's incentive compatibility conditions and see that:

(i) $(0)(\beta(1-\pi)-c) \geq 0$

$0 \geq 0$, and

(ii) $\alpha \geq 1 - \frac{0}{\pi}(\beta(1-\pi)-c)$

$\alpha \geq 1$

Both of these statements are true, so (Attack, Always Attack) is an equilibrium of this subgame. The final step is to determine whether or not D will initiate conflict, given the equilibrium outcome of doing so.

D will initiate if $W(A(V_1(\pi - c) + (1-V_1)(\alpha\pi - c)) + (1-A)(\pi - c)) + (1-W)(AV_1 + (1-A)V_2)(1-\beta(1-\pi)-c) \geq \epsilon$. Substituting 1 for V_1, V_2, and W, and simplifying, D will initiate if $\pi - c \geq \epsilon$.

((Initiate, Attack), (Always Attack)) with beliefs A is a PBE if $\pi - c \geq \epsilon$.

((Don't Initiate, Attack), (Always Attack)) with beliefs A is a PBE if $\pi - c \leq \epsilon$.

Strategy 2: S never attacks

In this case, $V_1 = 0$ and $V_2 = 0$. For never attacking to be the best strategy for S, it must be true that $W(1-\pi-c) + (1-W)(\beta(1-\pi)-c) \leq W(1-\alpha\pi-c)$, and $W(1-\pi-c)+(1-W)(\beta(1-\pi)-c) \leq W(1-\pi-c)$. Therefore, S's incentive compatibility conditions are:

(i) $(1-W)(\beta(1-\pi)-c) \leq 0$, and

(ii) $\alpha \leq 1 - \frac{1-W}{W\pi}(\beta(1-\pi)-c)$

If $(\beta(1-\pi)-c) \leq 0$, as it must be for S to never attack, (see (i)), then α is less than or equal to any number from 1 to (theoretically) infinitely large, which puts no limits on α

beyond the limits of the model. Therefore, D learns nothing new and resorts to its priors, A.

Comparing D's expected utilities:

$$EU_D(Attack|SNeverAttacks) = A(\alpha\pi - c) + (1 - A)(\pi - c)$$
$$EU_D(Don'tAttack|SNeverAttacks) = 0$$

Attack is D's best response if $A(\alpha\pi - c) + (1 - A)(\pi - c) \geq 0$, i.e. if $A \geq \frac{-(\pi-c)}{\pi(\alpha-1)}$. Don't Attack is D's best response if $A(\alpha\pi - c) + (1 - A)(\pi - c) \leq 0$, or $A \leq \frac{-(\pi-c)}{\pi(\alpha-1)}$. Plugging this into S's incentive compatibility conditions, if D attacks, W = 1:

(i) $(0)(\beta(1 - \pi) - c) \leq 0$

$0 \leq 0$, and

(ii) $\alpha \leq 1 - \frac{0}{\pi}(\beta(1 - \pi) - c)$

$\alpha \leq 1$

This is true if D's first strike advantage does not exist, but is not true if it does, so Never Attack is not a best response for S if A is high enough that D attacks. If D does not attack, W=0:

(i) $(1)(\beta(1 - \pi) - c) \leq 0$, and

(ii) $\alpha \leq 1 - \frac{1}{0 \cdot \pi}(\beta(1 - \pi) - c)$

$\alpha \leq 1 - \frac{\beta(1-\pi)-c}{0}$

The first statement is true by S's first incentive compatibility condition. The second, rearranged, is $W\pi\alpha \leq W\pi - (1-W)(\beta(1-\pi)-c)$. Substituting W=0 into that gives $0 \leq -\beta(1-\pi)-c$, so the second equation is also true, and Never Attack is a best response for S if D does not attack.

D will initiate if $W(A(V_1(\pi-c)+(1-V_1)(\alpha\pi-c))+(1-A)(\pi-c))+(1-W)(AV_1+(1-A)V_2)(1-\beta(1-\pi)-c) \geq \epsilon$. Substituting 0 for V_1, V_2, and W, and simplifying, D will initiate if $0 \geq \epsilon$. However, if there is a cost of initiating, the model stipulates that $\epsilon > 0$, so there is no PBE in which D initiates.

((Don't Initiate, Don't Attack), Never Attack) with beliefs A is a PBE if $A \leq \frac{-(\pi-c)}{\pi(\alpha-1)}$.

Strategy 3: S attacks if α exists and doesn't attack if α does not exist (strategy S*)

In this strategy, $V_1 = 1$ and $V_2 = 0$. S plays S* if $W(1-\pi-c)+(1-W)(\beta(1-\pi)-c) \geq W(1-\alpha\pi-c)$, and $W(1-\pi-c)+(1-W)(\beta(1-\pi)-c) \leq W(1-\pi-c)$. Therefore, S's incentive compatibility conditions are:

(i) $(1-W)(\beta(1-\pi)-c) \leq 0$, and

(ii) $\alpha \geq 1 - \frac{1-W}{W\pi}(\beta(1-\pi)-c)$

If $(\beta(1-\pi)-c) \leq 0$, as it must be for S to play S*, (see (i)), then $\alpha \geq 1$, which is a premise of the model, so D learns nothing new and resorts to its priors, A.

D's expected utilities if S plays S* are:

$$EU_D(Attack|SplaysS*) = \pi - c$$
$$EU_D(Don'tAttack|SplaysS*) = A(1-\beta(1-\pi)-c)+(1-A)(0)$$

Attack is D's best response when $\pi - c \geq A(1-\beta(1-\pi)-c)$, i.e. when $A \geq \frac{\pi-c}{1-\beta(1-\pi)-c}$, ($1-\beta(1-\pi)-c$ is negative, so the $\geq$ becomes a $\leq$ when you divide by it), and Don't Attack is D's optimal action when $\pi - c \leq A(1-\beta(1-\pi)-c)$, or $A \leq \frac{\pi-c}{1-\beta(1-\pi)-c}$. If D attacks, W=1:

(i) $(0)(\beta(1-\pi)-c) \leq 0$

$0 \leq 0, and$

(ii) $\alpha \geq 1 - \frac{0}{\pi}(\beta(1-\pi)-c)$

$\alpha \geq 1$

Both of these are true, so (Attack, S*) is an equilibrium to this subgame.

If D doesn't attack, W = 0:

(i) $(1)(\beta(1-\pi)-c) \leq 0$

(ii) $\alpha \geq 1 - \frac{1}{0 \cdot \pi}(\beta(1-\pi)-c)$

$\alpha \geq 1 - \frac{(\beta(1-\pi)-c)}{0}$

The first is true by S's first incentive compatibility condition. Rearranging the second inequality as above gives $0 \geq -\beta(1-\pi)-c$, which is not true, so S* is not a best response to Don't Attack.

D will initiate if $W(A(V_1(\pi-c)+(1-V_1)(\alpha\pi-c))+(1-A)(\pi-c))+(1-W)(AV_1+(1-A)V_2)(1-\beta(1-\pi)-c) \geq \epsilon$. Substituting 1 for V_1 and W and 0 for V_2 and simplifying, D will initiate if $\pi - c \geq \epsilon$.

((Initiate, Attack), (S*)) with beliefs A is a PBE if $\pi - c \geq \epsilon$ and $A \geq \frac{\pi-c}{1-\beta(1-\pi)-c}$.

((Don't Initiate, Attack,), (S*)) with beliefs A is a PBE if $\pi - c \leq \epsilon$ and $A \geq \frac{\pi-c}{1-\beta(1-\pi)-c}$.

Strategy 4: S attacks if α doesn't exist and doesn't attack if α does exist (strategy S)**

In this strategy, $V_1 = 0$ and $V_2 = 1$. S plays S** if $W(1-\pi-c)+(1-W)(\beta(1-\pi)-c) \leq W(1-\alpha\pi-c)$ and $W(1-\pi-c)+(1-W)(\beta(1-\pi)-c) \geq W(1-\pi-c)$. S's incentive compatibility conditions are:

(i) $(1-W)(\beta(1-\pi)-c) \geq 0$

(ii) $\alpha \leq 1 - \frac{1-W}{W\pi}(\beta(1-\pi)-c)$

If $(\beta(1-\pi)-c) \geq 0$, as it must be for S to play S**, (see (i)), then α is less than or equal to any number from 1 to (theoretically) infinitely large, which puts no limits on α beyond the limits of the model. Therefore, D learns nothing new and resorts to its priors, A.

D's expected utilities if S plays S** are:

$$EU_D(Attack|SplaysS**) = A(\alpha\pi - c) + (1-A)(\pi - c)$$
$$EU_D(Don'tAttack|SplaysS**) = A(0) + (1-A)(1-\beta(1-\pi)-c)$$

Attack is D's best response when $A(\alpha\pi-c)+(1-A)(\pi-c) \geq (1-A)(1-\beta(1-\pi)-c)$, i.e. when $A \geq \frac{1-\beta(1-\pi)-\pi}{\alpha\pi-\pi+1-\beta(1-\pi)-c}$ and Don't Attack is D's best response when $A(\alpha\pi-c)+(1-A)(\pi-c) \leq (1-A)(1-\beta(1-\pi)-c)$, or $A \leq \frac{1-\beta(1-\pi)-\pi}{\alpha\pi-\pi+1-\beta(1-\pi)-c}$. If D attacks, W = 1:

(i) $(0)(\beta(1-\pi)-c) \geq 0$

$0 \geq 0$

(ii) $\alpha \leq 1 - \frac{0}{\pi}(\beta(1-\pi) - c)$

$\alpha \leq 1$

The first is true. The second is true only if D does not have a first strike advantage, so S** is not a best response for S if D attacks. If D does not attack, W=0:

(i) $(1)(\beta(1-\pi) - c \geq 0$

(ii) $\alpha \leq 1 - \frac{1}{0 \cdot \pi}(\beta(1-\pi) - c)$

$\alpha \leq 1 - \frac{(\beta(1-\pi)-c)}{0}$

The first is true by S's first incentive compatibility condition. Rearranging the second inequality as above, $0 \leq -\beta(1-\pi) - c$, which is not true. S** is never a best response for S.

All PBEs

In summary, the three pure strategy perfect Bayesian equilibria are

- If $\pi - c \geq \epsilon$, D will Initiate and Attack with beliefs A, and S will Always Attack. If $\pi - c \leq \epsilon$, D will not Initiate.
- ((Don't Initiate, Don't Attack), Never Attack) with beliefs A is a PBE if $(1-\pi)-c \leq 0$ and $A \leq \frac{-(\pi-c)}{\pi(\alpha-1)}$
- If $\pi - c \geq \epsilon$ and $A \geq \frac{\pi-c}{1-\beta(1-\pi)-c}$, D will Initiate and Attack with beliefs A and S will play S*. If $\pi - c \leq \epsilon$ and $A \geq \frac{\pi-c}{1-\beta(1-\pi)-c}$, D will not Initiate.

Works Cited

Aker, Frank. *October 1973: The Arab-Israeli War.* Hamden, Connecticut: Archon Books, 1985.

Arreguin-Toft, Ivan. "How the Weak Win Wars: A Theory of Asymmetric Conflict." *International Security.* 1 (2001).

Asher, Dani. *The Egyptian Strategy for the Yom Kippur War.* Jefferson, North Carolina: McFarland & Company, Inc., 2009.

Claude, Inis L. *Power and International Relations.* New York: Random House, 1962.

DiCicco, Jonathan M., and Jack S. Levy. "Power Shifts and Problem Shifts: The Evolution of the Power Transition Research Program." *Journal of Conflict Resolution.* no. 43 (1999): 675-704.

Fearon, James D. "Rationalist explanations for war." International organization 49 (1995): 379-379.

Handel, Michael I. "The Yom Kippur war and the inevitability of surprise." International Studies Quarterly (1977): 461-502.

Jervis, Robert. "Cooperation Under the Security Dilemma." World Politcs 30 (1978): 167-214.

Kydd, Andrew. "Game theory and the spiral model." World Politics 49 (1997): 371-400.

Fey, Mark, and Kristopher Ramsay. "Mutual Optimism and War." *American Journal of*

Political Science. 51 (2007): 738-754.

Lemke, Douglas, and Suzanne Werner. "Power Parity, Commitment to Change, and War." *International Studies Quarterly.* 2 (1996): 235-260.

Moul, William. "Power Parity, Preponderance, and War Between Great Powers, 1816-1989." *Journal of Conflict Resolution.* 47 (2003): 468-489.

Schelling, Thomas C. *The strategy of conflict.* Harvard university press, 1980.

Shlaim, Avi, and William Roger Louis, eds. *The 1967 Arab-Israeli War: Origins and Consequences.* Vol. 36. Cambridge University Press, 2012.

Slantchev, Branislav, and Ahmer Tarar. "Mutual Optimism as a Rationalist Explanation of War." *American Journal of Political Science.* 55 (2011): 135-148.