Draft: 31 January 2014

Fear, Hope, and War: Positive Inducements Help Win Wars

TYSON BELANGER *Harvard University*

How do states win wars against other states? We have three explanations. By selection effects, states choose more winnable wars. By warfighting, states use negative inducements so enemies fear fighting. And by peacemaking, states use positive inducements so enemies hope for settling. This article investigates peacemaking. It theorizes that states optimally produce war influence only if they efficiently combine both warfighting negative and peacemaking positive inducements. It measures positive inducements by law of war compliance, where compliance is their presence and noncompliance means their absence, so it hypothesizes that compliance improves outcomes. It tests this by estimating average compliance effects on interstate outcomes from 1899 to 1991, in four models, with multiple specifications, and over nine issue areas. It finds that compliance likely on average causes better immediate military and final political outcomes. To win, states should be prudent by selection, fierce in warfighting, and principled enough for peacemaking.

Imagine a state that surprise attacks other states, betrays truces, tortures prisoners of war, targets Red Cross hospitals, torpedoes ships without warning, sprays nerve gas, fire bombs cities, massacres women and children, and devastates holy places. Such a state would be terrifying. But would it win? To resist it, do other states need to outdo atrocities, reciprocate to stay even, or comply with the laws of war? Moreover, what are the consequences if resisting states unleash righteous wrath, take an eye for an eye, or turn the other cheek and stay the hand of vengeance?

This article investigates the peacemaking explanation of how states win interstate wars. The peacemaking explanation argues that winning partly depends on using positive inducements to uphold or improve how current and potential enemy states and their sub-groups value settling. This article begins by reviewing views on peacemaking effects. It next theorizes that states must efficiently combine warfighting negative inducements and peacemaking positive inducements to optimally produce war influence. It measures positive inducements with law of war compliance, as coded by Morrow and Jo (2006), and hypothesizes that compliance likely improves outcomes. It tests this by statistically estimating the average effects of overall compliance on outcomes from 1899 to 1991, first in a baseline model, then in replications using Lake (1992), Reiter and

Stam (1998), and Downes (2009), and for nine separate law of war issue areas. It concludes by estimating average effects, considering alternative explanations, and proposing future research.

The tests discover numerous significant estimated effects. Overall compliance likely on average improves state outcomes, enemy compliance likely harms them, and relative compliance *very likely* improves them. The baseline model also estimates that broad relative compliance likely on average improves eventual outcomes in eight law of war issue areas. These include issuing formal declarations of war, respecting prisoners of war, caring for enemy wounded, obeying high seas standards of conduct, not using chemical and biological weapons, refraining from bombing cities, safeguarding civilians, and protecting cultural properties. One issue area, armistice compliance, likely has negative effects. These results provide unprecedented empirical support for the peacemaking positive inducement explanation of how states win interstate wars.

Reviewing Literature

To explain how states win interstate wars, most scholars use selection effects and warfighting arguments. Fearon (1995), and Bueno de Mesquita with Siverson (1995), argue that prudent leaders select into more winnable wars. Among scholars who emphasize warfighting, Morgenthau (1956), Waltz (1979), and Mearsheimer (2001) argue that brute force capabilities and threats decide high stakes international outcomes. Within wars, Pape (1996) compares coercive uses of airpower, and Biddle (2004) investigates modern battlefield tactics.

Few scholars study peacemaking positive inducements. For convenience, I conceive and group four peacemaking views among classical and contemporary authors. One view is *Optimistic*. Sun Tzu writes that wise leaders wield moral authority, leverage bargaining plus intelligence to win with less fighting, and recruit from captured enemy troops. Schelling (1960) uses game theory to mathematically show how sides can gain relative advantages by committing to better outcomes for enemies. More recently, Reiter and Stam (2002) hypothesize, but do not test, that states win by respecting prisoners and motivating enemy surrenders.

A second peacemaking view is *Pessimistic*. Machiavelli argues we should be both feared and loved, observes it is difficult to be both, and concludes it is “much safer” to be feared, if we avoid being hated. Clausewitz emphasizes advantages from using maximum force and fewer positive inducements than enemies. He writes, “war is such a dangerous business that the mistakes which come from kindness are the very worst,” and if “one side uses force without compunction, undeterred by the bloodshed it involves, while the other side refrains, the first will gain the upper hand” (Clausewitz 1976, 75-76). Outside the realist tradition, Walzer (1977) argues that war poses dilemmas of effective but immoral policies, verses costly but moral ones.

A third view is *Skeptical*. Thucydides, as a historian, event-by-event analyzes the Peloponnesian War and avoids sweeping causal inferences. Grotius and Kant describe the norms and laws of war as mutually beneficial, not as relatively advantageous. Contemporary war termination theorists, including Goemans (2000), often explain concessions by interacting opposing expected utilities for fighting, so positive inducements have little independent effect.

A fourth view is *Pluralist*. It argues that positive inducement effects differ by conditions, relations, and issues. For example, see Axelrod (1984). In his iterated prisoner’s dilemma, the best strategy is Tit for Tat. Like an *Optimist*, it begins with cooperation and never defects first. However, A’s worst payoff is from cooperating as B defects, and A’s best payoff is defecting as B cooperates. These payoffs suggest, as *Pessimists* believe, that “cheaters” can gain advantages. In this context, Tit for Tat protects itself, as a *Pessimist* might, by getting even. It reciprocates in kind, with one defection per enemy defection, and seeks zero relative cooperation.

THEORIZING INDUCEMENTS

I theorize that using peacemaking positive inducements, as compared to using none, causes better interstate war outcomes. My theorizing has four assumptions. The first, from the bargaining theory of war, is that all sides must value settling over fighting before peace becomes possible. Enemy evaluative gaps therefore define necessary quantities of wartime influence that states and allies must produce, while they also produce influence to keep potential foes neutral.

The second assumption, drawing from Schelling, is that states can uphold and increase enemy values for settling by peacemaking positive inducements (unconditional offerings and conditional promises) or suppress and reduce enemy values for fighting by warfighting negative inducements (unconditional brute force and conditional threats). These two inducement types produce wartime influence that bridges enemy evaluative gaps from complementary directions.[1]

The third assumption, from microeconomics producer theory, is that optimized output production depends on efficiently combining and substituting inputs, as if like labor and capital, by marginal benefit to cost ratios. Inputs gain maximize if marginal benefits exceed marginal costs. They loss minimize if marginal costs exceed marginal benefits, yet their ratios are higher than other inputs or inaction. They are inefficient if marginal costs exceed marginal benefits and if other inputs or inaction have higher ratios. States willingly add gain maximizing input levels, reluctantly add necessary loss minimizing levels, and mistakenly add inefficient levels.

The fourth assumption, from modern marketing, is that producers improve net returns by exploiting asymmetric stakes, using arbitrage, differentiating products, and discriminating prices to target niche demands and resources among consumer subgroups. In war, states, with their

[1] Brute force can kill, so states do not need to "bridge" all initial enemy evaluative gaps. Also, deaths reshape preferences among surviving enemy peoples. Nevertheless, modern wars have always ended in state-to-state or state-to-sub-group bargaining. Given baseline model's COWv4 data, the highest battle deaths as percent of prewar population is 5.1%, for WWII Germany.

diplomats and militaries, encounter and can leverage opportunities to implicitly and explicitly bargain with enemy allies, leaders, soldiers, and civilians, especially those that become isolated, seek surrender, or hold dissident preferences. Micro-bargaining improves macro-outcomes.

I call this, "War Influence Producer Theory," because states win by producing influence. The first assumption sets necessary outputs.[2] The second provides two inputs. The third reasons that efficiently combining inputs is better than not. The fourth capitalizes on how wars increase the complexity and interdependence of state-to-enemy relations. Together, these assumptions suggest states that produce necessary influence by efficiently and selectively combining negative and positive inducements are more likely to optimize outcomes, as compared to states that only use negative aggregate inducements. Mono-negative macro-bargaining states can still win, principally in the short-term, by favorable selection, warfighting, or chance effects, and before prewar peacemaking effects dissipate. However, they more likely lose as selection effects fade, foes recover and adapt, and chance averages out, but especially if enemies use more positive inducements in relations that deepen and multiply over time.

Positive inducements have several mechanisms. Directly to enemies, states use positive inducements for limited trades and loss minimizing over negative inducements. They also use them to motivate reconcilable enemies to shirk, surrender, or turn against less reconcilable and irreconcilable enemies. Toward potential enemies, states use positive inducements to retain their neutrality or realign sympathies. With friendly and allied states, positive inducements to enemies satisfy audience interests, earn diplomatic reputations, and reinforce allied positive inducements. For a state's own domestic subgroups, positive inducements to enemies affirm

[2] States may gain maximize or loss minimize by exceeding necessary influence, especially to hedge against backsliding and uncertain enemy evaluations.

inclusiveness, rally moral support, and inspire confidence among peacemaking *Optimists*. Finally, by these same mechanisms but conversely, states can suffer from enemy peacemaking.

This article compares general peacemaking views and motivates future research. From the above theorizing, it draws four *Optimist* hypotheses. First, (H_1) unilaterally using at least minimal positive inducements causes better outcomes than using none, whatever enemies do. Meanwhile, (H_2) unilateral at least minimal enemy positive inducements harm state outcomes. Next, consider the net assessment (Rosen 1991), so (H_3) states gain from higher relative positive inducements. Lastly, (H_4) diverse types of positive inducements provide advantages.

Pessimists, *Skeptics*, and *Pluralists* disagree. *Pessimists* believe positive inducements are often inefficient, particularly if naïve, idealistic, or legalistic states use too much and let them interfere with warfighting. *Pessimists* therefore hypothesize that states suffer from unilateral positive inducements, gain from enemy uses, suffer from relative uses, and suffer across types. In contrast, *Skeptics* doubt if positive inducements affect outcomes. Adversaries lack credibility, so states should not believe or use them. Selection, warfighting, and chance decide outcomes. So, *Skeptics* hypothesize that unilateral, enemy, relative, and multi-type positive inducements all have insignificant effects. *Pluralists* reject generalized views. They hypothesize that positive inducements have differing conditional unilateral, enemy, relative, and multi-type effects.

MEASURING INDUCEMENTS

In practice, positive and negative inducements are interdependent, so it is hard to measure one without the other, and many measures cannot distinguish their effects.[3] Measures have

[3] Here are seven types of inducement interdependence with examples to enemy troops. First, restraining one can increase the other, so not harming prisoners may require safeguarding them, while not protecting them can result in their harm. Second, scarce resources can pose trade-offs, so states may allocate troops to fight or manage prisoners, and feed prisoners or their own troops.

“positive inner correlations” for inducements if their positive and negative effects rise and fall together. This confounds interpretations. For example, democracy and wealth may cause wins by either inducement, so tests of them cannot distinguish warfighting from peacemaking effects.

I prefer measures with “negative inner correlations.” Their inducement effects rise and fall inversely. With them, we can compare high negative and low positive inducement levels against lower negative and higher positive inducement levels. If such measures affect outcomes, the effect is likely from the predominant inducement. For example, in World War II, when Allied bombing killed over 600,000 German civilians, this reduced German expected values for both fighting and the post-war peace (Roberts 2011). Bombing thereby meant both a higher negative and lower positive inducement effect. So, if we test bombing and it improves outcomes, we can infer bombing negative inducements improve outcomes. Or, if bombing harms outcomes, we can infer upholding positive inducements, by not bombing, helps outcomes.

As another virtue of negative inner correlated measures, they make hard tests for finding positive effects for positive inducements. *Pessimists* and *Skeptics* believe negative inducements improve outcomes, and positive inducements are harmful or insignificant, so they highly expect negative inner correlation measures to show negative inducements win. For example, they may believe surprise attacks help win wars. Like them, Japanese planners preferred to surprise attack Pearl Harbor. However, they ignited “a deadly firestorm of revenge, racism, embarrassment, and

---

Third, by blowback, excesses in one can reduce the other, so intimidating atrocities may discredit future promises, or enemies may misinterpret generosity as signaling weakness and fight harder. Fourth, one can enhance or complement the other, so attacks may enable dissident enemies to surrender, while granting quarter may help small mobile forces capture larger fortified ones. Fifth, by externalities, an inducement to one foe can mean something different to others, so receiving surrenders harms enemies who still fight, just as harming tyrants liberates their troops. Sixth, inducements can be setups, so states may swindle foes to surrender today then neglect them tomorrow, just as states may shock whole armies into surrender to sooner recruit them. Seventh, people have different reference points, so sufficient food for some may be torturously inadequate to others, while sporadic sniping may feel like a reprieve from prior artillery shelling.

idealism in the heart of the American people" (Murray and Millet 2000, 168). When Churchill heard of it, he declared the Allies had just won. If, across many wars, formal declarations likely have better effects than surprise attacks, this should doubly persuade *Pessimists* and *Skeptics*.

Negative inner correlation measures have two weaknesses. First, setups confound their interpretations. Negative inducements may weaken enemies to support later winning positive inducements, or positive inducements may lull foes to support winning negative inducements. For example, states may gain by partially fulfilling Armistice agreements before betraying them. This means significant effects for one inducement cannot rule out enabling effects by the other. As another weakness, powerful but contrary effects can appear nonexistent. For example, naval blockades, like Britain's World War I starvation blockade of Germany, which caused hundreds of thousands of civilian deaths, may gain strong negative but lose comparable positive effects. So, enemy evaluative gaps may shift but be unchanged in size, and blockades with enormous consequences may appear statistically insignificant for outcomes, if models do not parse effects.

Negative inner correlations occur among many measures of law of war compliance. The laws of war, as norms and standards of conduct, require diplomatic, humane, respectful, and sometimes risky or costly deeds, often conditional on enemies ceasing or not joining fighting, while they prohibit betrayals, frightfulness, indiscriminant terror, and irreversible destruction. Compliance, especially overall across issue areas, thereby upholds or increases enemy values for settling, so it directly and proxy measures positive inducement use, credibility, and propensity. In contrast, noncompliance reduces enemy values for settling, because it embitters relations, discredits bargaining, and provides little or no hope to escape violence, so it directly and proxy measures the absence of positive inducement use, credibility, and propensity. Noncompliance also involves negative inducements, so it makes hard tests against positive compliance effects.

Check these negative inner correlations for the laws of war. States retain diplomatic credibility by declaring wars and fulfilling armistice agreements, or they exploit sneak attacks and betrayals. States treat prisoners and captured wounded well to motivate surrenders, or they neglect them to save resources, torture them for intelligence, and outright kill them. States meet high seas standards of conduct, abstain from chemical and biological weapons, and not bomb cities, or they transgress these taboos and horrify enemies. Moreover, states safeguard civilians and protect cultural properties, or they punitively massacre civilians and devastate landmarks.

Therefore, rates of compliance with these law of war issue areas are useful for testing positive inducement effects. If unilateral, relative, and multi-issue compliance helps outcomes, while enemy compliance harms them, this supports *Optimist* peacemaking hypotheses. But if unilateral, relative, and multi-issue compliance harms outcomes, while enemy compliance helps them, this supports *Pessimists*. If compliance has no effects, this supports *Skeptics*. Finally, if compliance has conditional or issue dependent effects, this supports *Pluralists*.

Treatment Variables

My treatment variables are historically observed, non-experimental, measures of law of war compliance. I derive them from a peer reviewed dataset, funded by the National Science Foundation and published by Morrow and Jo in 2006. The dataset defines interstate wars as disputes between recognized governments, usually with over 1,000 battle deaths. It organizes wars into directed dyads, with one state row against each enemy state, but combines rows for states under unified commands, because they have interdependent compliance. It codes compliance starting with troops fighting and ending when this ceases. Morrow added to the dataset and studied the causes of noncompliance for his 2007 article. Rows in his 2007 dataset have A's compliance to B and B's compliance to A. For my baseline model, I use Morrow's 2007 format, with its 222 directed dyads rows, and group them in 48 wars from 1899 to 1991.

Morrow and Jo code compliance by Magnitude (M) and Frequency/Extent (F) of violations. They use these ordinal scales:

> Magnitude (M): 1=No violations at all; 2=Minor violations only; 3=Some major violations; and 4=Many major violations such that compliance doesn't matter.
> Frequency/Extent (F): 1=No violations at all; 2=Single or occasional violations; 3=Recurrent or common violations, with the standards still observed on many occasions; and 4=Massive violations to the point where the standard is ignored.

Magnitude=4 and Frequency=4 (M4&F4) means true "noncompliance" and other combinations are low, high, or full compliance. Standards do not differ over time or if states ratified treaties.

Morrow and Jo assess nine issue areas. Some have missing codes, because states lacked capabilities or opportunities for noncompliance. This is appropriate, so we compare actual and realistic counterfactual compliance. Where states had opportunities but Morrow and Jo found no historical evidence, they assign codes by typical rates of conduct. I describe the issues below, by their positive requirements and major violations, plus a prominent M4&F4 example. In brackets, I note the number of true noncompliance (M4&F4) / total observations by issue area.

1. DECLARATIONS OF WAR [16/222]: States should participate in diplomacy and issue formal war declarations, not launch massive sneak attacks. Egypt began the Yom Kippur War by surprise attacking Israel on 6 October 1973. Within 18 hours, Egypt sent 90,000 troops plus 850 tanks across the Suez Canal, and Israel lost 187 of its 290 front line tanks (Bar-Joseph 2000).

2. ARMISTICE [0/71]: States should honor flags of truce and satisfy ceasefire or armistice agreements, not betray them. In the Vietnam War, North Vietnam had two M4&F3 Armistice observations. In 1959, it began violating the Geneva Agreement by enlarging the Ho Chi Minh trail. In 1964, the year before the interstate war, it sent 10,000 troops south (Karnow 1991).

3. PRISONERS OF WAR [19/222]: States should accept surrenders, provide healthy holding conditions, allow relief society visits, and share information on prisoners, not overwork, starve, grossly punish, torture, or kill them. Germany surprise attacked the Soviet Union on 22 June 1941. In the first eight months, Germany shot, gassed, or neglected to death over 2 million Soviet prisoners. By war's end, it killed 3.3 million of 5.7 million it captured (Hastings 2012).

4. WOUNDED [3/222]: States should heal prisoners, respect enemy medical personnel, and recognize sanitary cities, not use them in ruses. On 1 July 1941, Soviet troops shot 165 German prisoners, including many wounded (De Zayas 1989). In 1943, at Stalingrad, they took 91,000 prisoners, gave no medical care, and almost half died before spring (Beevor 1998).

5. HIGH SEAS [9/162]: States should allow return passage for shipping when wars begin, meet formal standards if imposing blockades, use correct ship markings, warn before attacking, and rescue shipwrecked peoples, not indiscriminately mine, bombard civilians, hire privateers, take hostages, or attack civilian and hospital ships. Britain entered World War I on 5 August 1914. That same day, Germany mined British ports. On 7 May 1915, a German U-boat sank the *Lusitania* without warning and killed 1,201 civilians, including 128 Americans (Preston 2002).

6. CHEMICAL/BIOLOGICAL WEAPONS (CBW) [8/205]: States should use conventional weapons, not toxins, bacteria, viruses, large nuclear attacks, or some incendiaries. The Iran and Iraq War lasted from 1980 to 1988. Iraq began using mustard gas in August 1983 and nerve gas in March 1984. Iraqi chemical attacks killed 20,000 Iranian soldiers (Harris and Paxman 2002).

7. AERIAL BOMBARDMENT [2/169]: States should grant civilians and civil installations immunity from aerial, rocket, and missile attacks, not target them. In World War II's final five months, U.S. bombing killed over 300,000 Japanese civilians, wounded hundreds of thousands, and made 15 million homeless by destroying 20% of Japan's housing structures (Frank 1999).

8. CIVILIANS [18/222]: States should safeguard and aid civilians, not take hostages, mass or non-judicially punish, deport, conscript, blockade, plunder, abuse, bombard, and kill them, or destroy infrastructure and use unmapped mines. On 25 March 1971, Pakistani troops began massacring Bengalis in East Pakistan. They killed an estimated 3 million, raped 200,000, and displaced 40 million (Jahan 2009). India intervened in December. The war lasted two weeks.

9. CULTURAL PROPERTY [2/222]: States should protect holy, educational, artistic, and historic places, with their personnel, not destroy or loot them. In early 1900, China's Boxers burned churches, massacred converts, and murdered missionaries. World powers demanded that China punish them. China refused, declared war, and cultural attacks increased (Esherick 1987).

As a tenth area, I compute OVERALL compliance, across all areas. No state had M4&F4 in all nine issues, but three states had four: World War I Germany against France, World War II Germany against the Soviet Union, and World War II Japan against the United States. I use overall compliance to test (H_1) unilateral, (H_2) enemy, and (H_3) relative effects. I later use individual issue areas to test (H_4) multi-issue effects.

With Morrow and Jo's codes, I operationalize compliance in two ways. The first is *direct* mass-squared $(M*F)^2$ compliance. For this, I multiply Magnitudes (M) and Frequencies (F). Both M and F measure positive inducements, and M*F is Morrow 2007's dependent variable. Next, I square (^2) these products, so upper compliance levels differ far less than lower levels. This models the ease of compensating across upper levels and difficulty across hard-to-replace lower levels. For overall compliance, I add the nine $(M*F)^2$ issue values in each directed dyad, divide by the maximum value of $9*(4*4)^2$, and multiply by -1. Dividing scales coefficients but does not change significance. Multiplying by -1 aligns coefficients, so positive A values suggest A's compliance improves its outcomes, and negative B coefficients mean B's compliance harms

A's outcomes. I next compute relative overall compliance by subtracting B from A, so positive coefficients mean higher A helps A and higher B harms A. Finally, I create by-issue variables. For this, I divide $(M*F)^2$ issue values by $(4*4)^2$ and multiply by -1. I then compute relative by-issue compliance and drop relative values if A or B have missing observations.

Direct compliance has two limitations. It misses post-dyadic compliance. For example, World War II Germany's direct compliance to the Netherlands misses May 1940 to May 1945. These years mattered. Germany never gained total Dutch capitulation, so Japan could not bypass Western embargoes by entering the oil-rich Dutch East Indies, like it entered Vichy Indochina. As another limit, direct compliance misses policies outside dyads. For example, Germany's direct compliance to the Netherlands is blind to German policies against other countries.

To account for post- and extra-dyadic in-war compliance, I compute *wide* mass-squared $(M*F)^2$ compliance. This begins at dyadic starts and counts all compliance until war's end. By issue, I aggregate A's highest magnitudes and frequencies for A to B with A's other same day start or later dyads. If A complied to prisoners with B but noncomplied (M4&F4) in same day or later dyads, A's wide prisoners compliance to B is M4&F4. For B, I aggregate its magnitudes and frequencies to A with B's other dyads beginning that day or later. Wide magnitudes and frequencies range 1 to 4. I then use the same calculations as for direct mass-squared compliance.

OUTCOME VARIABLES

My dependent variables are interstate war outcomes. Theorists often analyze them in nested levels. Tactical outcomes describe the immediate results of limited encounters or battles. Operational outcomes describe groups of tactical results. Strategic outcomes describe groups of operations to large political objectives. Grand strategic outcomes describe groups of strategies, by their largest and most lasting political consequences. Ideally, states succeed at every level.

However, levels have different causal dynamics. This can generate multi-level dilemmas and prompt profound disagreements about how states win wars. In a famous post-war conversation by American and North Vietnamese Colonels, the American said, “You know, you never defeated us on the battlefield.” The Vietnamese Colonel replied, “That may be so, but it is also irrelevant” (Summers 1982). For international relations, we are most interested in large and lasting political consequences, so I explain strategic and grand strategic outcomes.

Morrow’s 2007 article provides one set of peer reviewed strategic outcomes. He bases them on the *Correlates of War Project* dataset (COWv3) and recodes for immediate outcomes. For example, COWv3 codes France as winning World War II, but Morrow codes it losing to Germany and Italy, because of its 1940 defeat. Morrow made his changes, because he theorized that impending military losses reduce compliance, and he used these outcomes as controls for explaining noncompliance. I make two changes.[4] My Morrow 2007 based strategic outcomes have 103 wins, 16 draws, and 103 losses. By convention, they are symmetric, so A’s win is B’s loss, and draws are mutual. Also, outcomes describe relative gains. Wins can be pyrrhic and more costly in duration or deaths than draws and losses. World War I Serbia has three wins, but it fought over four years, and 62.5% of its men aged fifteen to sixty-five died (Hastings 2013).

To measures grand strategic outcomes, I base them on COWv4. COWv4 codes winning, losing, or indecisive relative final outcomes by side, so it codes World War II France as winning. France fought alongside the Allies, governed a German post-war occupation zone, and gained a United Nations Security Council permanent seat. Similarly, Poland, Norway, the Netherlands, Belgium, Yugoslavia, and Greece lost militarily, but their elites foresaw they might eventually

[4] Morrow codes the 1919-1921 Franco-Turkish War as a draw. I code France as winning against the Ottoman Empire. Morrow codes the 1938 Changkufeng War as Japan winning. I code the Soviet Union as winning. For details, see the Annex.

triumph, and some pursued political resurrection. Concurrently, Germany knew its early gains might be reversed, and this insecurity affected its occupations and foreign relations. Ultimately, by May 1945, Germany had a worse outcome than states it earlier militarily defeated. COWv4 does well for such eventual relative outcomes. I make four changes.[5] My COWv4 based grand strategic outcomes have dyadic symmetry, with 99 wins, 24 draws, and 99 losses.

COWv4 based grand strategic codes differ from Morrow 2007 based strategic codes in 32 directed dyads, about 14% of the total, and all in World War I or II. They also differ by available treatment variables. Direct compliance can explain either outcome, but wide compliance includes post-dyadic measures, so it can only explain grand strategic outcomes.

Control Variables

For my baseline model, I considered over 56 controls, most with prewar and predyad year variations, coded by lead states for unified commands. To choose from them, I required they plausibly affect compliance, plausibly affect strategic and grand strategic outcomes, avoid quasi-separation, reduce residuals, and minimize Cooks values. I generally found that relative values, side A minus B, performed best. Differences explain differences. I use these 14 controls:

Capabilities: Material power reduces compliance (Morrow 2007) and improves outcomes. With the National Military Capabilities Dataset v4.0, I make two composite measures. The first adds prewar urban population and .001*predyad military spending. The next adds prewar total population and prewar energy resources. I then compute relative values.

[5] COWv4 codes Russia as winning World War I. I recode it to draws verses Germany, Austria-Hungary, Bulgaria, and the Ottoman Empire. COWv4 codes the Franco-Turkish War of 1919-1921 as a draw. I recode France as winning. COWv4 codes Poland and the Soviet Union as winning World War II. I recode Poland as losing verses the Soviet Union. COW v4 codes the 1975-1979 Vietnamese-Cambodian and 1977-1978 Ethiopian-Somali wars as transitioning into other war types. I code Vietnam and Ethiopia as winning. For details, see the Annex.

Democracy: Democracies have less noncompliance (M4&F4), more often have positive overall relative compliance, and are more likely to win. I code states as democracies if they have prewar year Polity IV polity scores above 6 and use relative values. Prewar democracies fought each other only once, in the 1974 Turco-Cypriot War, which began 5 days after a Cyprian coup.

Democratic Ratification: Democracies that ratify recent and relevant treaties comply more (Morrow 2007). Also, ratifications may indicate or affect state power, alliances, and outcomes. I replicate Morrow, code prewar democratic ratifications with International Red Cross data, average the nine issue areas for overall compliance rows, and use relative values.

Domestic Political Inclusiveness: Politically inclusive states may comply more and are much likelier to win. I measure this by A's prewar domestic winning coalition or "W," minus B's, as coded by Selectorate Theory (Bueno de Mesquita, Smith, Siverson, and Morrow 2003).

Recent Wars & Outcomes: After recent interstate wars, states may comply and lose more. Wars and demobilizations weaken states. Wins and draws may also leave them overconfident or overextended. I code if A or B concluded a COWv4 war 5 years prewar, add one point if the latest one ended in a grand strategic win or draw, and use relative values.

After 1945: Cold War balancing deterred noncompliance and increased draws. Norms, technology, institutions, and nationalism also mattered. I use post-1945 war dummy variables.

Prior Noncompliance: States with prior massive noncompliance may have institutions, beliefs, or traditions that reduce compliance, and those with this material power more likely win. I code 1's for any M4&F4 observation since 1899 or massively targeting civilians since 1816 (Downes 2012), code 0's otherwise, and use relative values.

Ongoing Noncompliance: States with predyad violations likelier noncomply and lose. This is also true for violations against them, as they face multiple foes and often reciprocate. For A and B, I tally predyad first violations plus suffered first violations, then use relative values.

Allied Combatants: Sides with fewer allies and more enemies comply less and lose more. With Morrow 2007, I tally the predyad states fighting alongside A and against B, then subtract the number alongside B and against A.

Initiation: Initiators comply less, enter more winnable wars, and have more preparation. I use Morrow 2007 codes, which uses COWv3 and adds in-war front initiations. 50 observations have no initiators. I use relative values.

Revisionist War Aims: Revisionists may comply less and face different difficulties. With Militarized Interstate Dispute v3, I create three predyadic relative codes for revisionism plus government change aims, revisionism plus territory aims, and revisionism plus policy aims.

OVERALL COMPLIANCE

This section tests baseline overall compliance average effects. The baseline model uses logit for A's wins or draws (1) verses losses (0), and clusters robust standard errors by 48 wars. Supplemental Tables report complete results. Table 1 summarizes overall direct mass-squared $(M*F)^2$ compliance estimated effects on Morrow 2007 based strategic outcomes. Model 1 estimates A's unilateral compliance very likely improves A's outcomes. Model 2 estimates B's compliance very likely harms A's outcomes. Model 3 estimates A's relative compliance very likely helps A, with a 2.91 maximum Cooks value and standard residuals ranging -2.32 to +2.32. These small values are rare for highly predictive models of war outcomes. Together, these results support *Optimist* hypotheses for (H_1) unilateral, (H_2) enemy, and (H_3) relative effects.

TABLE 1. Baseline Overall Compliance Strategic Models			
	Strategic Outcome Models (Morrow 2007 Based)		
Overall Compliance	1. A Only	2. B Only	3. A Relative
A's Direct $(M*F)^2$	5.332 (1.534)***		
B's Direct $(M*F)^2$		-6.390(1.846)***	
A's Direct $(M*F)^2$ Relative			20.407(2.649)***

Notes: $^+$p<.10, *p<.05, **p<.01, ***p<.001. Tests are two-tailed logit of wins or draws (1) verses losses (0). N=222. Results are coefficients and robust standard errors clustered by 48 wars in parentheses. Supplemental Tables report complete results.

Table 2 reports baseline overall direct and wide mass-squared $(M*F)^2$ compliance effects on COWv4 based grand strategic outcomes. Again, Supplemental Tables report more results. Models 1-3 use overall direct $(M*F)^2$ compliance. They estimate A's unilateral compliance likely helps A, B's unilateral compliance very likely harms A, and A's relative compliance very likely helps A. Models 4-6 report estimated overall wide $(M*F)^2$ compliance effects. A's wide compliance likely helps A, B's wide compliance very likely harms A, and A's wide relative compliance likely helps A. These results affirm *Optimist* (H_1) unilateral, (H_2) enemy, and (H_3) relative hypotheses for both direct and wide compliance effects on grand strategic outcomes.

TABLE 2. Baseline Overall Compliance Grand Strategic Models			
	Grand Strategic Outcome Models (COWv4 Based)		
Overall Compliance	1. A Only	2. B Only	3. A Relative
A's Direct $(M*F)^2$	4.873 (1.913)*		
B's Direct $(M*F)^2$		-8.727 (1.778)***	
A's Direct $(M*F)^2$ Rel.			15.292(3.074)***
	4. A Only	5. B Only	6. A Relative
A's Wide $(M*F)^2$	5.963 (1.893)**		
B's Wide $(M*F)^2$		-10.201(1.325)***	
A's Wide $(M*F)^2$ Rel.			15.708 (6.355)*

Notes: $^+$p<.10, *p<.05, **p<.01, ***p<.001. Tests are two-tailed logit of wins or draws (1) verses losses (0). N=222. Results are coefficients and robust standard errors clustered by 48 wars in parentheses. Supplemental Tables report complete results.

The World Wars deserve special testing. They have all the strategic and grand strategic differences, 52 of the 77 M4&F4 observations, and comprise 41% of the dataset, with 56 WWI and 36 WWII rows. If we add dummy controls for them, compliance has similar or higher significance. If we drop WWI, A and B's unilateral direct strategic results lack significance, but direct relative strategic and all grand strategic results exceed 99% significance. If we drop WWII, all results exceed 99%, with 95% in A's unilateral direct grand strategic results, as above. If we drop the 92 World War rows, A's results are insignificant positive, B's are negative at 95% significance, and A's relative results are positive constants, in completely predictive models.

As another test, we might remodel outcomes. Leaders can disagree if some draws, by mitigating risks and costs, are better than wins. In Korea, Presidents Truman and Eisenhower accepted draws, but General MacArthur wanted victory. Still, I retest logit with wins (1) verses draws or losses (0). I also run ordered logit wins (2), draws (1), and losses (0). For both, results exceed 95% significance. I next drop draw rows, and this increases results significance.

Alternatively, we can remodel errors. The baseline model clusters robust standard errors by wars, because errors are in-war interdependent. Surprising wins and noncompliance by one dyadic or warring side often correspond to surprising losses and noncompliance by the other. However, if we use regular or robust errors, results still exceed 95% significance, with A's unilateral direct grand strategic result at 90%. Or, if we cluster by dyads, all results exceed 95%.

Next, we might re-specify exponents. If we use M*F with no exponent, results exceed 99% significance, with A's unilateral direct grand strategic result still at 95%. If we use four instead of two as an exponent, this increases unilateral significance, and all results exceed 99%.

For a simpler test, we can try binary specifications. For direct binary compliance, the 44 A's with one or more noncompliance (M4&F4) issues get 0's, and 178 others get 1's. For wide binary, the 59 A's with one or more noncompliance issue in that dyad, same day starts, or later dyads get 0's, and 163 get 1's. A's relative compliance equals -1, 0, or 1. The results are all over 99% significance, with A's unilateral direct and wide grand strategic results above 95%.

With binary compliance, we can match by forecasts. This weights observations to better compare those with similar forecasts but different compliance. To estimate forecasts, I drop compliance and predict probabilities of wins or draws. I then coarsen forecasts into groups above or below 50% and match on binary treatments. This generally increases significance to results above 99%, with A's unilateral wide grand strategic at 95%. Next, I group by 0-25%, 25-50%, 50-75%, and 75-100%. This generally increases significance, to all above 99%.

REPLICATION TESTING

My best estimates are the baseline results. They make full use of the original compliance data, with my most relevant and updated controls. Still, if compliance has such strong effects, we should find results elsewhere. To test this, I use three well-known models for democracy effects on outcomes. One is from Lake's 1992 *American Political Science Review* article, "Powerful Pacifists: Democratic States and War." Another is from Reiter and Stam's 1998 *American Political Science Review* article, "Democracy, War Initiation, and Victory." The third is from Downes' 2009 *International Security* article, "How Smart and Tough Are Democracies?"

The replication models differ from my baseline. Instead of directed dyads, they use one monadic state row per war. I add data, so side A monadic compliance equals A's aggregate compliance verses its monadic dataset foes and B's equals the aggregate compliance of A's foes. I next compute mass-squared variables and add them to primary models in each article. After this, the models describe wars from 1899 to 1982. Lake uses logit with regular standard errors, tests grand strategic outcomes, drops wars without democracies, drops COW draws, does not unify commands, and only controls for A's unilateral Polity democracy score (0 to 10), iron and steel production, and troop numbers. Reiter and Stam use probit with robust standard errors, organize wars by Stam 1996, test Stam 1996 strategic outcomes, drop draw rows, drop some non-leading states, and use twelve variables for polity (-10 to 10), initiation, military strategies, terrain, and allied capabilities. Downes uses ordered probit with robust standard errors clustered by 58 Stam (1996) sub-wars, uses Reiter and Stam rows but adds draws, uses their variables but reinterprets polity and initiation, and recodes several outcomes. Table 3 reports results.

TABLE 3. Lake, Reiter&Stam, and Downes^ Overall Compliance Models

	Strategic Outcome Models		
Overall Compliance	1a-c. Lake	2a-c. Reiter&Stam	3a-c. Downes^
A's Direct $(M*F)^2$	a. .854 (1.572)	a. 1.293 (1.421)	a. 1.487 (.773)$^+$
A's Direct $(M*F)^2$	b. 4.460 (2.254)*	b. 7.787 (2.174)***	b. 3.672 (1.350)**
B's Direct $(M*F)^2$	-4.718 (2.006)*	-9.250 (2.019)***	-3.388 (1.784)$^+$
A's D. $(M*F)^2$ Rel.	c. 4.636 (1.926)*	c. 8.033 (2.074)***	c. 3.524 (1.508)*
	Grand Strategic Outcome Models		
	4a-c. Lake	5a-c. Reiter&Stam	6a-c. Downes^
A's Direct $(M*F)^2$	a: 3.107 (1.768)$^+$	a. 1.899 (1.429)	a. .767 (.741)
A's Direct $(M*F)^2$	b: 9.050 (2.863)**	b. 5.337 (1.929)**	b. 3.435 (1.661)*
B's Direct $(M*F)^2$	-7.713 (2.571)**	-5.547 (1.861)**	-4.368 (1.784)*
A's D. $(M*F)^2$ Rel.	c: 8.213(2.496)***	c. 5.435 (1.680)**	c. 3.917 (1.701)*
	7a-c. Lake	8a-c. Reiter&Stam	9a-c. Downes^
A's Wide $(M*F)^2$	a: 3.107 (1.768)$^+$	a: 2.975 (1.452)*	a: 1.568 (.691)*
A's Wide $(M*F)^2$	b: 9.050 (2.863)**	b: 6.377 (1.704)***	b: 4.319 (1.631)**
B's Wide $(M*F)^2$	-7.713 (2.571)**	-5.137 (1.745)**	-4.368 (1.720)*
A's W. $(M*F)^2$ Rel.	c: 8.213(2.496)***	c: 5.667 (1.447)***	c: 4.346 (1.629)**

Notes: $^+$p<.10, *p<.05, **p<.01, ***p<.001. Results are coefficients with errors in parentheses. Tests are two-tailed. Lake is logit at N=103 with regular standard errors. Reiter and Stam is probit and robust standard errors for strategic N=127 and grand strategic N=131. Downes^ modifies Downes (2009). It uses probit for strategic N=157 with errors clustered by 58 Stam 1996 wars and grand strategic N=157 with robust errors clustered by 43 baseline wars. Supplemental Tables report complete results.

Lake has these results. Models 1a-c test overall direct mass-squared $(M*F)^2$ monadic compliance effects on Morrow based strategic outcomes. Model 1a suggests A's compliance has positive but insignificant effects. Model 1b adds B. It estimates A's compliance likely helps A, while B's compliance likely harms A. This loosens prescriptions against simultaneously testing treatment variables in one equation. So, consider Model 1c. It finds A's relative compliance likely helps A. Next, Models 4a-c and 7a-c estimate overall direct and wide mass-squared $(M*F)^2$ compliance effects on Lake's original COW grand strategic outcomes. Direct and wide compliance variables in Lake have the same numeric values and results. Lake grand strategic models estimate likely positive unilateral and relative effects for A, with negative effects for B. The Lake replications generally affirm baseline results and support *Optimist* hypotheses ($H_{1,2,3}$).[6]

Next, consider Reiter and Stam. Models 2a-c use overall direct mass-squared $(M*F)^2$ monadic compliance to explain Reiter and Stam coded strategic outcomes. Model 2a suggests A's compliance has positive but insignificant effects. Model 2b adds B. In it, A's compliance likely helps A, and B's compliance likely harms A. Model 2c estimates A's relative compliance very likely helps A. Next, Models 5a-c and 8a-c test overall direct and wide mass-squared compliance for COWv4 based grand strategic outcomes. In these, the original rows have concave results so, from Stam 1996, I add Sinai War Britain and France, plus World War II China verses Japan. Models 5a-c direct grand strategic results resemble direct strategic results. Model 8a estimates A's wide compliance likely helps A, whatever enemies do. Model 8b adds B, and Model 8c tests relative effects. In these, A's compliance very likely helps A, while B's likely harms A. These tests generally support *Optimist* hypotheses ($H_{1,2,3}$).[7]

Now consider Downes. Compliance has little significance in his original 2009 model. However, we find significant results if we change ordered probit to probit, control for offensive verses defensive military doctrine (Stam 1996), and control for post-1945 wars (Downes 2012). Models 3a-c test overall direct mass-squared $(M*F)^2$ compliance effects on Downes coded

[6] Adding compliance does not affect Lake's original estimated democracy effects.
[7] Adding compliance does not affect Reiter and Stam's estimated democracy effects.

strategic outcomes. Model 3a estimates compliance may help A. Model 3b adds B and estimates A's compliance likely helps A, while B's may harm A. Model 3c estimates A's relative compliance likely helps A. Models 6a-c and 9a-c test overall direct and wide $(M*F)^2$ compliance on COWv4 based grand strategic outcomes. In Model 6a, A's unilateral direct compliance result is positive but not significant. Model 6b adds side B, and Model 6c tests relative compliance. These find significant results. Next, Models 9a-c estimate A's wide compliance likely helps A, B's likely harms A, and A's wide relative compliance very likely helps A. With minor changes, the Downes replications support *Optimist* hypotheses ($H_{1,2,3}$).[8]

9 Issue Areas

The results so far support *Optimist* hypotheses for unilateral, enemy, and relative overall compliance. This section tests the *Optimist* hypothesis (H_4) that many positive inducement types improve outcomes. For each issue, it estimates direct and wide mass-squared $(M*F)^2$ relative compliance effects on strategic and grand strategic outcomes in the baseline, Lake, Reiter and Stam, and Downes models. Here, I summarize results. Supplemental Tables report more.

1. Declarations of War: 1/4 direct strategic, 3/4 direct grand strategic, and 3/4 wide grand strategic models estimate positive relative compliance effects above 95% significance. States likely gain by their declarations and enemy surprise attacks, mostly for grand strategy.

2. Armistice: 4/4 direct strategic, 4/4 direct grand strategic, and 4/4 wide grand strategic models estimate negative relative effects all above 95% confidence. 9/12 have negative constant results. But Morrow and Jo code just 71 observations. Also, no state has true noncompliance (M4&F4). Instead, all have some compliance, perhaps enough to set up betrayals.

[8] Adding compliance does not affect Downes' democracy results, but probit and the post-1945 controls change estimates. The results suggest, differing from Downes 2009, that higher polity states likely have better strategic and grand strategic outcomes. For details, see the Annex.

3. PRISONERS OF WAR: 4/4 direct strategic, 4/4 direct grand strategic, and 4/4 wide grand strategic models estimate positive relative compliance effects above 95% significance. Across multiple levels, states likely gain by respecting prisoners and suffering enemy violations.

4. WOUNDED: 1/4 direct strategic, 3/4 direct grand strategic, and 3/4 wide grand strategic models estimate positive relative compliance effects above 90% significance. States likely gain by healing prisoners, respecting medical personnel, and suffering enemy violations.

5. HIGH SEAS: 0/4 direct strategic, 0/4 direct grand strategic, and 1/4 wide grand strategic models estimate positive relative compliance effects above 95%. Notably, two direct strategic, three direct grand strategic, and two wide grand strategic results are negative but insignificant. High seas compliance might have powerful contrary negative and positive inducement effects, while relative wide compliance may help grand strategy, depending on how models parse effects.

6. CHEMICAL/BIOLOGICAL WEAPONS: 2/4 direct strategic, 2/4 direct grand strategic, and 4/4 wide grand strategic models estimate positive relative effects above 90% significance. At multiple levels, A likely gains by abstaining from these taboo weapons and by enemy violations.

7. AERIAL BOMBARDMENT: 1/4 direct strategic, 1/4 direct grand strategic, and 1/4 wide grand strategic models estimate positive relative effects above 90% significance. The baseline model has the three significant results. Meanwhile, the Reiter and Stam wide grand strategic model estimates negative relative effects above 95%. Like for High Seas, Aerial compliance might involve powerful contrary positive and negative effects that models parse differently.

8. CIVILIANS: 3/4 direct strategic, 2/4 direct grand strategic, and 4/4 wide grand strategic models estimate positive relative compliance effects above 90%. Across multiple levels, states likely gain from safeguarding civilians and suffering enemy violations.

9. CULTURAL PROPERTY: 4/4 direct strategic, 1/4 direct grand strategic, and 4/4 wide grand strategic models estimate positive relative compliance effects above 90%. On multiple levels, higher relative cultural protections likely help outcomes.

Now, consider these by-issue relative compliance effect patterns, using 90% significance. Across 108 results, 56 (52%) are positive, 39 (36%) are insignificant, and 13 (12%) are negative or concave. Of 27 baseline results, 19 (70%) are positive, 5 (19%) are insignificant, and 3 (11%) are negative. By issue area, across models, seven have at least one positive result, one (Aerial) has three positive and one negative results, and one (Armistice) is all negative. Among baseline direct strategic results, six are positive, two are insignificant, and one is negative. Of baseline direct grand strategic results, five are positive, three are insignificant, and one is negative. Of nine baseline wide grand strategic results, eight are positive and one is negative. These results best support the *Optimist* hypothesis (H_4) that many positive inducement types help outcomes.

EFFECT SIZES

I prioritize studying positive, negative, or insignificant effects, because causal directions best compare peacemaking positive inducement views. Still, large and frequent effects deserve more attention than small and infrequent ones. To help inform such triage, while remaining skeptical about obtaining exact causal estimates from observational data, this section estimates average effect sizes. It does this by holding control variables, mostly relative values, at their averages and estimating changes in probabilities of wins and draws, given varying increases or decreases from average relative compliance. This method suggests average effect sizes and 95% confidence intervals given average circumstances. To also help describe effect frequencies and counterfactual support, this section compares test changes to relative compliance distributions.

Table 4 reports average effect sizes for the baseline model and on average over the four models, given overall mass-squared relative compliance changes equal to one issue's difference between high (M3&F2) and low (M4&F3) compliance. These changes are numerically similar to changing from low (M4&F3) to noncompliance (M4&F4). At average relative compliance, probabilities of wins or draws range 60-74%. Test increases, if A complies more and/or B less, raises baseline probabilities from 12%±8 to 18%±3 and 4-model probabilities 10%±5 to 10%±6. Decreases, if A complies less and/or B more, decreases baseline probabilities 16%±14 to 23%±6 and four model probabilities 11%±8 to 12%±6. These effects are substantial, especially given their openness to in-dyad state decisions, unlike the control variables. Furthermore, the tests range just 21-40% of one standard deviation, so many wars likely have larger effects.

TABLE 4. Average Effect Sizes of M3F2↔M4F3 Changes For Overall Relative Compliance

Issue Area	Strat or GStrat	Side A $(M*F)^2$ Rel.	Average % Change Probability Win/Draw w/Test Changes Δ=M3&F2↔M4&F3 ±95%CI and Δas%of1StdDev					
			Baseline Model			Four Model Average		
			%if↑	%if↓	%SD	%if↑	%if↓	%SD
Overall	Strat	Direct	18±3	-23±6	40	10±5	-12±6	39
	GStrat	Direct	13±6	-17±8	40	10±5	-11±6	39
		Wide	12±8	-16±14	21	10±6	-11±8	33

We can further test smaller changes and by separate issue areas. Table 5 below estimates average relative compliance effect sizes by changes between M3&F2 and M3&F3, equaling between F2="Single or occasional violations" and F3="Recurrent or common violations, with the standards still observed on many occasions." This is numerically equal to changes from M2&F3 to M3&F3 or about double changes from M2&F2 to M3&F2 or M2&F3. At average relative compliance, initial probabilities of wins or draws range 58-71%, except for armistice. Armistice compliance has few observations and often fully predicted or concave results, so increases likely cause losses and decreases likely cause wins or draws. Across other issues, test increases change baseline probabilities -5%±15 to +34%±14 and 4-model probabilities -2%±7 to +24%±13. Decreases change baseline model probabilities -51%±16 to +5%±15 and 4-model probabilities -32%±24 to +1%±10. The test changes comprise 9-140% of one standard deviation, so many directed dyads feature larger effects, depending on the issue area.

TABLE 5. Average Effect Sizes of M3F2↔M3F3 Changes For Overall Relative Compliance

Issue Area	Strat or GStrat	Side A $(M*F)^2$ Rel.	Average % Change Probability Win/Draw w/Test Changes Δ=M3&F2↔M3&F3 ±95%CI and Δas%of1StdDev					
			Baseline Model			Four Model Average		
			%if↑	%if↓	%SD	%if↑	%if↓	%SD
Overall	Strat	Direct	8±2	-9±2	17	5±2	-5±2	17
	GStrat	Direct	6±3	-7±3	9	4±2	-4±2	16
		Wide	6±4	-6±5	9	4±3	-4±3	14
1. Declaration	Strat	Direct	3±7	-4±7	42	6±6	-7±7	36
	GStrat	Direct	9±7	-11±9	42	6±6	-7±6	36
		Wide	13±8	-15±11	34	7±6	-8±7	34
2. Armistice	Strat	Direct	L	W/D	113	L	W/D	102
	GStrat	Direct	L	W/D	113	L	W/D	102
		Wide	L	W/D	113	L	W/D	102
3. Prisoners	Strat	Direct	15±7	-17±10	57	10±7	-11±8	59
	GStrat	Direct	9±7	-11±9	57	10±7	-11±8	59
		Wide	13±8	-15±11	41	11±7	-13±9	53
4. Wounded	Strat	Direct	25±7	-35±13	101	6±8	-9±10	89
	GStrat	Direct	26±10	-51±16	101	15±10	-22±13	86
		Wide	24±9	-43±14	84	15±9	-21±12	82
5. High Seas	Strat	Direct	1±5	-1±5	66	-2±7	-1±7	67
	GStrat	Direct	-5±17	5±15	66	-2±11	1±10	68
		Wide	14±11	-19±17	47	2±9	-3±11	61
6. CBW	Strat	Direct	17±9	-23±20	93	10±10	-11±13	102
	GStrat	Direct	10±7	-12±11	93	14±10	-19±14	107
		Wide	12±6	-15±9	58	16±9	-21±13	89
7. Aerial	Strat	Direct	15±13	-17±18	103	1±11	-2±12	93
	GStrat	Direct	22±16	-29±30	103	4±12	-6±15	93
		Wide	22±12	-36±32	78	1±12	-4±16	84
8. Civilians	Strat	Direct	10±10	-11±12	59	7±7	-8±8	58
	GStrat	Direct	4±9	-4±10	59	6±7	-6±8	57
		Wide	10±5	-11±6	42	8±6	-8±7	52
9. Cultural	Strat	Direct	34±14	-47±27	140	20±13	-25±20	93
	GStrat	Direct	30±20	-49±55	140	23±18	-31±34	93
		Wide	24±8	-39±15	54	24±13	-32±24	71

Figure 1 illustrates ranges of overall direct and wide mass-squared relative compliance average effect sizes.

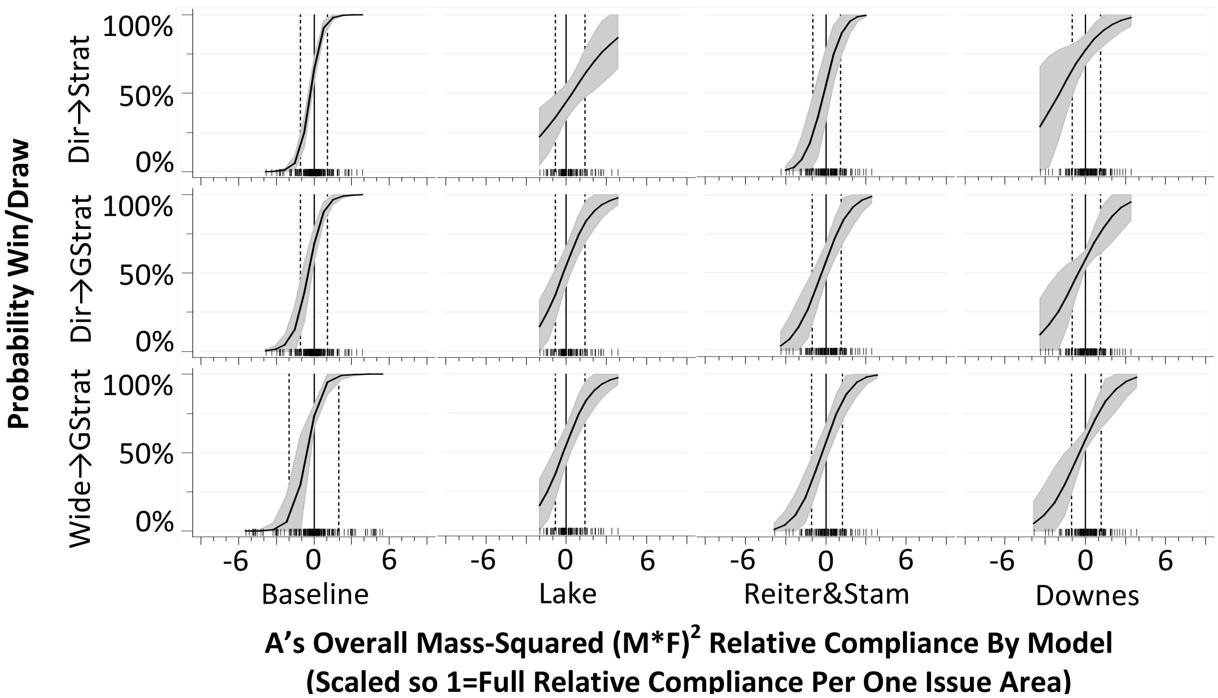

Figure 1. By-Model Average Probabilities of Wins or Draws Verses Losses by Side A's Overall Mass-Squared Relative Compliance (Holding Controls at Average Values)

Each X-Axis shows A's overall relative compliance, by direct strategic, direct grand strategic, or wide grand strategic models. Ticks equal relative differences between one side's noncompliance (M4&F4) against an enemy's full compliance in one issue. In theory, ticks range from -9 for all issues noncompliance against full compliance foes, up to 0 relative compliance at the vertical black line, then to 9 for full compliance against totally noncompliant foes. Short vertical black lines note actual observations. Dashed vertical black lines indicate one standard deviation above and below median relative compliance, all at or near 0. Each Y-Axis estimates probabilities of wins or losses, with controls at average values. S-curved black lines trace probabilities over the ranges of actual observations. Grey regions depict upper and lower 95% confidence intervals.

The graphs show how higher overall relative compliance climbs states up the S-curve to likely wins or draws, while lower relative compliance slides them down to likely losses.[9] No side A has under -6 or over 6 relative compliance ticks. Still, the narrow intervals at the lowest and highest observed values suggest more extreme relative values would have similar results.

How do selection effects compare? We might measure selection by Militarized Interstate Dispute initiator side, relative COWv4 war starters, and relative Morrow 2007 front initiators.[10] Figure 2 shows baseline effect sizes, with average controls, not counting compliance.

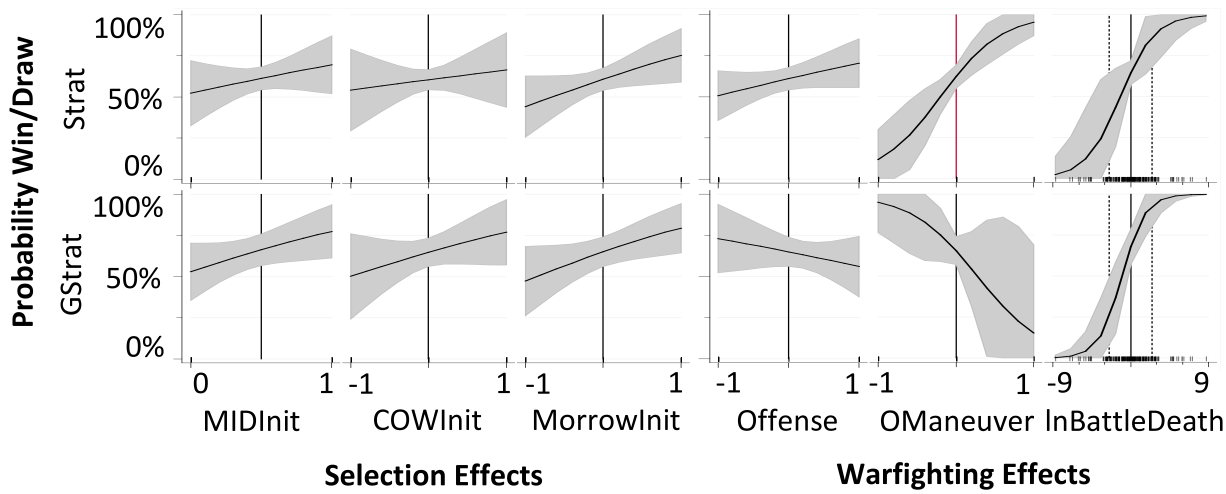

Figure 2. Baseline Model Average Probabilities of A's Wins or Draws Verses Losses by Selection and Warfighting Relative Variables (Holding Controls at Average Values)

Dispute initiator sides gain 17%±35 strategic and 24%±29 grand strategic probability points against target sides. War starters gain 12%±46 strategic and 27%±43 grand strategic points verses their foes. Front initiators gain 31%±33 strategic and 32%±32 grand strategic points verses enemies. Compliance compares well against these estimated selection effects.

[9] Along logit and probit S-curves, states with favorable circumstances may gain less from higher relative compliance and, if they fall off the flat S-top, lose more by negative relative compliance. States with unfavorable circumstances may lose less from lower relative compliance but, if they overcome the flat S-bottom, gain more from positive relative compliance. These asymmetries may explain differing A and B coefficients and differing increase or decrease test effect sizes.
[10] MID and COWv4 initiator variables include pre-control events.

What about warfighting? We may study relative codes for Stam 1996 offensive doctrine, Reiter and Stam 1998 offensive maneuver, and the natural log of COWv4 all-war battle deaths.[11] Figure 2 shows baseline sizes, not counting compliance. Initially offensive states gain 17%±27 strategic but lose 17%±36 grand strategic points. Offensive maneuver gains 84%±26 strategic but loses 80%±72 grand strategic points verses immobile defenders. Defensive maneuver gains 57%±68 strategic and 68%±59 grand strategic points verses low mobility attackers. One positive standard deviation of relative log battle deaths gains 52%±49 strategic and 69%±37 grand strategic points verses one negative standard deviation. Meanwhile, one positive standard deviation of overall mass-squared relative compliance gains 81%±10 direct strategic, 66%±21 direct grand strategic, and 91%±23 wide grand strategic points verses adversaries with one negative standard deviation. So, compliance compares well, especially to offensive maneuver, which likely historically has negative grand strategic effects. Plus, if we control for any of these selection or warfighting variables, relative overall compliance has similar or higher coefficients. For optimal outcomes, states likely should combine selection, warfighting, and peacemaking.

ALTERNATIVE EXPLANATIONS?

This article finds numerous significant correlations. Why? The simplest answer is that Sun Tzu, Schelling, Reiter, and Stam are right. The correlations are causal. Compliance and peacemaking positive inducements help win wars. Wise states provide alternatives to fighting. If we reject this, we should seek a better explanation. This section considers three possibilities.

First, endogeneity may explain the results if, with powerful and pervasive dynamics, winning causes compliance, or losing causes noncompliance, while compliance insignificantly

[11] COWv4 war battle deaths include pre-control and extra-dyadic deaths.

affect outcomes. *Optimists, Pessimists*, and most *Pluralists* would oppose this possibility.[12] Also, not all *Skeptics* would believe impending outcomes affect compliance, but some might.

We can explore this with forecast-as-treatment tests. These estimate forecasts for wins or draws, not counting compliance, then test if these positively cause compliance, controlling for other variables. In 141 side A models in Tables 1-3 or By-Issue tests, 75 forecasts are negative and 66 are positive for explaining compliance. At 90% significance, 14 are negative, 120 are insignificant, and 7 are positive. Forecasts likely do not on average positively affect compliance.

Next, we can try forecast-as-omitted-variable tests. We know forecasts positively correlate with wins and draws, so if they are controls, and this reduces compliance coefficients, this suggests forecasts positively cause compliance. However, in the 141 Table or By-Issue models, forecast controls cause 71 coefficient increases and 65 decreases, while 5 have concave results. Across 85 side A compliance results with significant positive coefficients, 4 increase in significance levels and 6 decrease. Furthermore, recall how matching forecasts on binary overall compliance little changed the results. In general, initial forecasts likely do not affect compliance.

We can next explore endogeneity with first-violations-verses-durations tests. States might begin with uncertainty and learn over time, so violations might begin later in wars, when violators are better informed about eventual outcomes. Morrow and Jo code 474 first violation dates, not counting declaration and armistice compliance, which have distinct timings. Most states violate one or more issues just 7 days or 4% into dyads. Across all first violations, most occur 30 days or 14% into dyads, noncompliance (M4&F4) observations on average begin

---

[12] If compliance improves outcomes, this affirms the *Optimist* explanation. If compliance harms outcomes, this cannot explain correlations between compliance and winning. If winning causes noncompliance or losing causes compliance, the results likely underestimate compliance effects.

earlier than more compliant observations, and states with different eventual outcomes have similar timings. Noncompliance likely does not strongly depend on learning from time.

Finally, consider first-violations-verses-battle-tides tests. Battle tides are start dates for battles after which one state won all remaining dyadic battles (Lo, Reiter, and Hashimoto 2008). If states learn from battles, first violations may occur after tides. Furthermore, if winning causes compliance, winners should not noncomply post-tide. We have 229 available first violation and tide comparisons, not counting declaration and armistice compliance. Most states have one or more violations 1.16 years pre-tide, with an average of 1.61 years. 74.3% of states have one or more pre-tide violation, 13.5% have on-tide first violations, and just 12.2% have them post-tide. Of all 229 comparisons, most first violations occur 1.48 years pre-tide and 1.81 years pre-tide on average. 77.2% of first violations are pre-tide, 5.7% are on-tide, and 17% are post-tide. Among the 36 M4&F4 comparisons, 94% began pre-tide, 0% on-tide, and 6% post-tide. Of 39 post-tide first violations, 74% had strategic wins and 26% had losses, while 51% had grand strategic wins and 49% had losses. Learning from battle tides likely does not cause first violations. In sum, across four exploratory tests, endogeneity likely does not alternatively explain the results.

As a second explanation, indiscipline may cause noncompliance and worse outcomes. However, recall that material power causes noncompliance (Morrow 2007). Strong militaries can easier substitute negative for positive inducements. Then, recall that violations, if they occur, often begin early, while states still best control their forces. Next, acknowledge how *Pessimists* believe noncompliance helps and prefer it. Now, consider this: Morrow and Jo's lowest compliance observations have their highest rates of centralized control and at-the-time clarity of illegality.[13] For example, in October 1977, Somalia fought Ethiopia and Cuba, and its Defense Ministry reportedly issued orders to take no prisoners, so thousands were summarily shot (de Waal 1991). Massive violations involve zealous or fearful discipline, not indiscipline.

As a third alternative, noncompliance may mean using inefficient negative inducements, while compliance may mean more efficient negative inducements. For example, if states use fewer bombers to attack civilians, they may use more for effective close air support. If so, aerial compliance might correlate with winning but not cause it. However, this does not explain other effects, like for war declarations, caring for enemy wounded, and protecting cultural properties. Furthermore, look closer at our aerial bombing research. It often argues, as in Pape 1996, that bombing hardens civilian will. This occurs if civilians have positive inducement expectations about wartime immunity and post-war peace. Bombing then diminishes these expectations, which widens evaluative gaps and thus hardens civilian will. Therefore, positive inducements help, and bombing sacrifices them. Coercive efficiency does not alternatively explain the results.

We need more research. This article uses observational data and extrapolations from it. Perhaps experimental data or future wars will have different results. Future research might also better manage in-dyad, in-war, and between-war interdependence.[14] Looking at the data, we may recheck Morrow and Jo's codings, code more years, study in-dyad variations, or further test

[13] Central state control of violations codes are 1="No violations at all," 2="Individual violations against state policy and which are punished by state policy," 3="Individual violations not punished by state policy," 4="Probable state decision to violate, and 5="Positive identification of state intent to violate." For all 77 noncompliance (M4&F4) observations, average centrality is 4.7 and median is 5. For 331 M4&F4, M4&F3, M3&F4, M3&F3, M4&F2, and M2&F4 observations, the average is 4 and median is 4. At-the-time noncompliance legal clarity codes are 1="No violations at all," 2="Legal Status of violation in clear dispute," 3="Probable violation, but not totally clear", and 4="Definite legal violation." For 77 noncompliance (M4&F4) observations, average clarity is 3.9 and median is 4. For 331 M4&F4, M4&F3, M3&F4, M3&F3, M4&F2, and M2&F4 observations, the average is 3.8 and median is 4.

[14] To reduce dyadic interdependence, I sampled baseline rows to count only one side, but this caused arbitrarily different coefficients and significance that depended on the sampling rules.

the control variables. In future models, we may study compliance interactions with regime type, military strategies, alliances, and soft power (Nye 1990). We may also deeper test mechanisms within issues and investigate tactical or operational effects. Or, we may test other dependent variables, such as allied entries, casualties, expenses, and durations. Beyond wars, we may test effects on territory changes, post-war regimes, international institutions, trade, rivalries, and propensities for future wars. Adjacent studies may test effects on militarized interstate disputes or for intrastate wars. This article opens several promising doors for future research.

CONCLUSION

The results best support *Optimist* views of compliance and positive inducement effects. States (H_1) likely on average benefit by using at least minimal overall compliance, whether or not enemies comply. States (H_2), whatever they do, likely on average suffer worse outcomes if enemies at least minimally comply. States (H_3) very likely on average gain by higher relative compliance. Plus, (H_4) higher relative compliance likely has positive effects across many issues. These results are robust across specifications and models, for both strategic and grand strategic outcomes. Altogether, this research provides substantial and unprecedented empirical support for the *Optimist* view that peacemaking positive inducements improve interstate war outcomes.

If wars shape international relations, peacemaking does too. If states compete, as if in a Hobbesian war of all against all, they succeed by their relative prudence, strength, and principles; the *Leviathan* needs its crown, sword, and crosier. If, as Machiavelli believes, state ends justify wartime means, relative overall compliance deserves such consequentialist justification, because it very likely on average improves outcomes. Moral high ground has realist military value.

These findings make several contributions to our literature. Scholars of selection effects and the causes of war gain ways to theorize rational war initiations. Warfighting scholars gain a control variable. If warfighting and peacemaking negatively correlate and both cause winning, we may underestimate warfighting if we omit peacemaking (King, Keohane, and Verba 1994). *Pessimist* scholars gain an opportunity to alternatively explain these results or update their views. *Skeptical* historians may afford more attention to peacemaking events. *Skeptical* scholars of international laws and institutions may consider if relative compliance causes relative gains in other areas. *Skeptical* war termination theorists may model positive inducements with greater independent weight. Lastly, Tit for Tat *Pluralists* may reconsider if exact in kind reciprocity improves outcomes and if the prisoner's dilemma accurately models wartime pay-offs. Schelling might instead recommend mixed strategy games involving three or more choices (1960, 134).

Given these results and contributions, why is *Pessimism* still so appealing? Maybe, unlike average effects, certain regime types, contexts, aims, and issues benefit by fewer positive inducements. Or, some positive prewar relations or war ends may outweigh harsh in-war means. Also, if positive inducements have diminishing returns, *Pessimists* may reject their upper uses, not their average uses. More psychologically, human perception might cause mistaken *Pessimist* intuitions (Kahneman 2011). Looking at individual wars, we face fundamental counterfactual uncertainties. Looking across wars, strong states noncomply and win more, so we may associate power, noncompliance, and wins. Moreover, interstate wars are irregular, and some have ambiguous results, so it is hard to form expert opinions about them. Tactical and operational outcomes are more regular and clear, but their lessons may ill-explain strategic and grand strategic outcomes. Even among supreme experts, misjudgments happen. Clausewitz is likely wrong about relative positive inducement effects. More viscerally, war involves killing, dying, hatred, dread, expressive acts, supreme sacrifice, sadness, and bitterness, so it is difficult to think about giving hope to enemies or feel this is right. Even *Optimists* cannot escape this.

Peacemaking is powerful but not easy. The optimality challenge is for wartime leaders and institutions to overcome *Pessimist* views against positive inducements, so states may achieve better outcomes. The morality challenge is for leaders and institutions to channel natural human impulses away from punitive and reciprocal justice, so states enter and conduct more just wars. To help wrestle with these challenges, we must continue researching wartime peacemaking.

References

Axelrod, Robert.1984. *The Evolution of Cooperation*. United States: Harper Collins Publishers.

Bar-Joseph, Uri. 2000. "Israel's 1973 Intelligence Failure." *Revisiting the Yom Kippur War*. Editor P.R. Kumaraswamy. Portland, Oregon: Frank Cass & Co. Ltd.

Beevor, Antony. 1998. *Stalingrad*. New York: Penguin Books Ltd.

Biddle, Stephen. 2004. *Military Power: Explaining Victory and Defeat in Modern Battle*. Princeton, N.J.: Princeton University Press.

Clausewitz, Carl von. 1976. *On War*. Edited and translated by Michael Howard and Peter Paret. Princeton, New Jersey: Princeton University Press.

Clodfelter, Michael. 2008. *Warfare and Armed Conflicts: A Statistical Encyclopedia of Casualty and Other Figures, 1494-2007, 3rd Edition*. Jefferson, North Carolina: McFarland & Company, Inc.

Daniel Jones, Stuart Bremer, and J. David Singer. 1996. "Militarized Interstate Disputes, 1816-1992: Rationale, Coding Rules, and Empirical Patterns," *Conflict Management and Peace Science*, 15:2: 163-215.

De Waal, Alexander. 1991. *Evil Days: 30 Years of War and Famine in Ethiopia*. New York: Human Rights Watch.

De Zayas, Alfred M. 1989. *The Wehrmacht War Crimes Bureau, 1939-1945*. Lincoln, Nebraska: University of Nebraska Press.

Downes, Alexander B. 2009. "How Smart and Tough Are Democracies? Reassessing Theories of Democratic Victory in War." *International Security*, 33:4 (Spring): 9–51.

Downes, Alexander B. 2012. *Targeting Civilians in War*. Cornell University Press: Ithaca.

Esherick, Joseph W. 1987. *The Origins of the Boxer Uprising*. Los Angeles, California: University of California Press.

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization*, 49:3 (Summer): 379-414.

Frank, Richard B. 1999. *Downfall: The End of the Imperial Japanese Empire*. New York, Random House.

Ghosn, Faten, Glenn Palmer, and Stuart Bremer. 2004. "The MID3 Data Set, 1993–2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21: 133-154.

Ghosn, Faten, and Scott Bennett. 2004. "Codebook for the Dyadic Militarized Interstate Incident Data, Version 3.10."

Grotius, Hugo. *On the Law of War and Peace*. Edited by Stephen C. Neff. Cambridge: Cambridge University Press, 2012.

Goemans, Hein E. 2000. *War and Punishment: The Causes of War Termination and the First World War*. Princeton, New Jersey: Princeton University Press.

Harris, Robert and Jeremy Paxman. 2002. *A Higher Form of Killing: The Secret History of Chemical and Biological Warfare*. New York: Random House Trade Paperback.

Hastings, Max. 2011. *Inferno*: The *World at War 1939-1945*. New York: Alfred A. Knopf.

Hastings, Max. 2013. *Catastrophe 1914: Europe Goes to War*. New York: Alfred A. Knopf.

Jahan, Rounaq. 2009. "Genocide in Bangladesh." *Century of Genocide: Critical Essays and Eyewitness Accounts Third Edition*. Edited by Samuel Totten and William S. Parsons. New York: Routledge.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Karnow, Stanley. 1991. *Vietnam: A History, Revised Edition*. New York: Penguin Books.

Kant, Immanuel. 2005. *Perpetual Peace*. New York: Cosimo Classics.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry*. Princeton, New Jersey: Princeton University Press.

Lake, David A. 1992 "Powerful Pacifists: Democratic States and War." *American Political Science Review*, 86:1 (March): 24-37.

Lo, Nigel, Dan Reiter, and Barry Hashimoto. 2008. "Ensuring Peace: Foreign-Imposed Regime Change and Postwar Peace Duration, 1914–2001." *International Organization* 62 (Fall 2008): 717-36.

Machiavelli, Niccolo. 1950. *The* Prince and The Discourses. Translated by Max Lerner. New York: The Modern Library.

Mearsheimer, John. 2001. *The Tragedy of Great Power Politics*. New York: W.W. Norton & Co.

Mesquita, Bruce Bueno, and Randolph M. Siverson. 1995. "War and the Survival of Political Leaders: A Comparative Study of Regime Types and Political Accountability." *The American Political Science Review* 89:4 (December): 841-855.

Mesquita, Bruce Bueno, James D. Morrow, Randolph M. Siverson, and Alastair Smith. 1999. "An Institutional Explanation of the Democratic Peace." *American Political Science Review*. 93:4 (December): 791-807.

Monty G. Marshall and Keith Jaggers. 2007. “Polity IV Project: Political Regime Characteristics and Transitions, 1800-2007.”

Morgenthau, Hans. 1956. *Politics Among Nations, Second Edition*. New York: McGraw-Hill.

Morrow, James D., and Hyeran Jo. 2006. “Compliance with the Laws of War: Dataset and Coding Rules.” *Conflict Management and Peace Science* 23 (Spring): 91–113.

Morrow, James D. 2007. “When do States Follow the Laws of War?” *American Political Science Review*, 101:3 (August): 559-572.

Nye, Joseph Jr. 1990. “The Changing Nature of World Power.” *Political Science Quarterly* 105:2: 177-192.

Overy, Richard. 1997. *Why the Allies Won*. New York, W. W. Norton & Company.

Pape, Robert. 1996. *Bombing to Win: Air Power and Coercion in War*. Ithaca, New York: Cornell University Press.

Preston, Diana. 2002. *Lusitania: An Epic Tragedy.* New York: Walker Publishing Company.

Reiter, Dan and Allan C. Stam. 2002. *Democracies at War*. Princeton: Princeton University Press.

Reiter, Dan, and Allan Stam. 1998. “Democracy, War Initiation, and Victory.” *American Political Science Review*, Vol. 92, No. 2 (June), pp. 377-389.

Roberts, Andrew. 2012. *The Storm of War: A New History of the Second World War*. New York: Harper Perennial.

Rosen, Stephen Peter. 1991. “Net Assessment as an Analytical Concept.” *On Not Confusing Ourselves*, Boulder: Westview Press.

Sarkees, Meredith Reid and Frank Whelon Wayman. 2010. *Resort to War: A Data Guide to Inter-State, Extra-State, Intra-State, and Non-State Wars, 1816-2007*. Washington, DC: CQ Press.

Sarkees, Meredith Reid. 2000. “The Correlates of War Data on War: An Update to 1997.” *Conflict Management and Peace Science* 18 (Fall): 123–44.

Schelling, Thomas. 1960. *The Strategy of Conflict.* Cambridge, Massachusetts: Harvard University Press.

Schelling, Thomas. 1966. *Arms and Influence*. New Haven: Yale University Press.

Schelling, Thomas. 1989. “Promises.” *Negotiation Journal* 5:2: 113-118.

Small, Melvin, and J. David Singer. 1982. *Resort to Arms: International and Civil Wars, 1816–1980*. Beverly Hills, CA: Sage.

Stam, Allan C. 1996. *Win, Lose, or Draw: Domestic Politics and the Crucible of War*. Ann Arbor, Michigan, University of Michigan Press.

Singer, J. David, Stuart Bremer, and John Stuckey. 1972. “Capability Distribution, Uncertainty, and Major Power War, 1820-1965.” in Bruce Russett (ed) *Peace, War, and Numbers*, Beverly Hills: Sage, 19-48.

Singer, J. David. 1987. “Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985.” *International Interactions*, 14: 115-32.

Summers, Jr., Harry G. 1982. *On Strategy: A Critical Analysis of the Vietnam War*. Novato, Calif.: Presidio Press.

Sun-Tzu. 1994. *The Art of War*. 1994. Translated by Ralph Sawyer and Mei-chun Sawyer. Boulder, Colorado: Westview Press.

Waltz, Kenneth N. 1979. *Theory of International Politics*. New York, New York: McGraw-Hill.

Walzer, Michael. 1977. *Just and Unjust Wars: A Moral Argument With Historical Illustrations*. United States of America: Harper Collins.