# Known Unknowns:

# Power Shifts, Uncertainty, and War

Alexandre Debs and Nuno P. Monteiro

*forthcoming in International Organization*

**Acknowledgments**

The authors contributed equally to this article. We thank Jonathan Caverley, Martha Crenshaw, James Fearon, Michael Horowitz, Jonathan Kirshner, Matthew Kocher, Andrew Kydd, John Mearsheimer, David Palkki, Benoit Pelopidas, Barry Posen, Robert Powell, Andrew Ross, Joshua Rovner, Bruce Russett, Scott Sagan, Anne Sartori, Kenneth Schultz, Duncan Snidal, Matthias Staisch, Michael Tomz, workshop participants at Chicago (PIPES), Cornell, Madison, MIT, and Yale, audiences at the ISSS/ISAC and PSS annual meetings, and the editors and three anonymous reviewers at *International Organization* for their comments and suggestions; and Will Bruno, Kelsey Ditto, Bonny Lin, and Hayden Stein for their research assistance. Alexandre Debs thanks CISAC at Stanford and the BCEP at UC-Berkeley for their hospitality.

**Authors' affiliations and emails:**

Alexandre Debs

Yale University

alexandre.debs@yale.edu

Nuno P. Monteiro

Yale University

nuno.monteiro@yale.edu

**Known Unknowns: Power Shifts, Uncertainty, and War**

Alexandre Debs and Nuno P. Monteiro

**Abstract**

Large and rapid power shifts resulting from exogenous economic growth are considered sufficient to cause preventive wars. Such power shifts are rare, however. Most large and rapid shifts result from endogenous military investments. In this case, preventive war requires uncertainty about a state's investment decision. When this decision is perfectly transparent, peace always prevails. A state's investment that would produce a large and rapid power shift would prompt its adversaries to launch a preventive war. Internalizing this, the state is deterred from investing. When investments may remain undetected, however, states may be tempted to introduce large and rapid shifts in military power as a *fait accompli*. Knowing this, their adversaries may strike preventively even without unambiguous evidence about militarization. In fact, the more effective preventive wars are, the more likely they will be launched against states that are not militarizing. Our argument restricts the role of commitment problems and emphasizes the role of imperfect information as causes of war. It also provides an account of why powerful states may attack weaker targets suspected of military investments even in the absence of conclusive information. We illustrate our theory through an account of the 2003 U.S.-led invasion of Iraq.

Word count: 12,271

# 1. Introduction

Power shifts are a long-standing object of study in international politics. For Thucydides, "increasing Athenian greatness and the resulting fear among the Lacedaemonians made going to war inevitable."[1] Recently, power transition theory focused on power shifts as a cause of systemic wars.[2] Likewise, rationalist explanations for war point to large and rapid power shifts as a sufficient cause for war.[3]

Still, two problems remain in the study of power shifts. First, most existing scholarship conceptualizes them as exogenous. Yet, most large and rapid shifts in the balance of power are endogenous to state interaction, resulting from militarization efforts. Second, the few pieces that endogenize power shifts fail to acknowledge the time lag between the moment in which a state decides to invest in military capabilities and the moment these become available, ignoring the possibility of a preventive attack. Yet, this possibility informs states' decisions to develop new military capabilities that would produce large and rapid power shifts, such as nuclear weapons. The prospect of prevention makes it possible to deter large and rapid power shifts peacefully. In sum, no existing scholarship fully incorporates key properties of most large and rapid power shifts.

We conceptualize power shifts as the result of a state's investment with delayed return in military capabilities and then determine when they trigger preventive wars.[4] When investments in military capabilities are transparent, other states can resort to preventive war

[1]Thucydides (1998, 15).

[2]Gilpin (1981).

[3]Powell (2006).

[4]"Preventive war" includes a range of military actions, from surgical strikes to full-scale war. See Levy (2008).

to preclude a large and rapid power shift. The specter of preventive war in turn deters states from making investments that would produce such shifts, as they would be attacked before the new capabilities become available. This means that preventive wars do not occur when military investments are transparent; peace always prevails regardless of the magnitude of possible power shifts. Small shifts would not be enough to trigger preventive wars. Large shifts would, but they are deterred through the possibility of war. Endogenous power shifts only produce war under (realistic) conditions of uncertainty about military investment decisions. Uncertainty about a state's military investment decisions prevents other states from trusting that investments will be detected. This uncertainty has two effects. First, it opens the possibility of a state secretly investing in significant military capabilities hoping to produce a large and rapid power shift as a *fait accompli*. Second, and consequently, it gives other states an incentive to strike preventively even in the absence of unambiguous evidence that an investment leading to a large and rapid power shift is underway. In other words, uncertainty about the timely detection of military investments makes preventive wars rational even against targets that are "innocent" – *i.e.*, have been deterred from undertaking military investments. The likelihood of preventive war in turn depends on its effectiveness: how long it guarantees that the target will not militarize. Greater effectiveness makes a preventive war more tempting, undermining peace. At the same time, greater effectiveness makes the threat of preventive war more credible, making military investments less likely. Taken together, these two effects increase the likelihood that preventive wars are mistaken.

Our argument makes two main contributions. First, we specify a new causal mechanism for preventive wars. In exogenous power shifts, preventive war can happen with perfect information. Not so when power shifts are endogenous, in which case preventive war requires uncertainty. Using the language of rationalist explanations for war, our argument shows that, when power shifts are endogenous, commitment problems only lead to war in the presence of

information problems. Second, we show how preventive wars can happen within a rationalist framework even against a target that is not investing in military power. These contributions have important practical implications in an era of U.S. military preponderance. An investment in military capabilities – *e.g.*, a nuclear-weapons program – is more likely to effect a large and rapid power shift when the prior relative power of the militarizing state is low. But if states with low relative power are more likely to produce large and rapid power shifts, they are also more likely to be targeted by preventive attacks launched by more powerful states. We show that such preventive attacks may occur in the absence of unambiguous evidence that the target is militarizing. This means that preventive military action launched by powerful states such as the United States will sometimes target states that are not conducting suspected military investments.

To illustrate our theory, we turn to the case of the 2003 U.S.-launched invasion of Iraq. We shed light on four important and hitherto underplayed aspects of the run-up to that war. First, we argue that an important factor triggering the war was the U.S. administration's fear that an Iraqi weapons of mass destruction (WMD) program would produce a large and rapid endogenous power shift. Second, we highlight how the war's timing was conditioned by the September 11, 2011, terrorist attacks, which shaped U.S. perceptions of the consequences of Iraqi nuclearization. Third, we provide an account of how uncertainty about whether Iraq was indeed pursuing a WMD program led to an ultimately mistaken preventive war. Finally, we account for why the United States attacked Iraq rather than other plausible but relatively more powerful targets such as North Korea. In sum, determined to prevent Saddam Hussein from acquiring nuclear weapons, the U.S. administration was unable to prove that Iraq was not, in fact, developing them. Faced with the possibility of a large and rapid power shift in favor of Iraq and operating with imperfect information about Iraq's militarization decision, the Bush Administration opted for a preventive war, which was mistaken as there was no active Iraqi

nuclear program.

We articulate our argument as follows. Section 2 reviews the relevant literature and introduces our theory. Section 3 formalizes the theory. Section 4 applies our theory to the case of the 2003 U.S.-led invasion of Iraq. Section 5 presents our conclusions. Proofs of the formal results are in the Appendix.

## 2. Theory and Literature

Power shifts have been the focus of a long research tradition: power transition theory.[5] For power transition theorists, shifts result from differential rates of economic growth, likely to set "the whole system sliding almost irretrievably toward war."[6] As a rising state increases its power, it is likely to become dissatisfied with the international *status quo*. At the same time, the declining state has incentives to preserve that *status quo*. This tension, power-transition theorists argue, is likely to produce war.

Power transition theory, however, has left several important questions unanswered. First, the causes of a rising state's dissatisfaction remain undetermined, with no agreement over whether it stems from risk acceptance,[7] status dissonance,[8] or territorial disputes.[9] Second, there is no agreement on how to operationalize dissatisfaction to make it empirically testable.[10] Finally, the literature is split on whether war will be launched by the rising

[5]Organski (1958); Gilpin (1981); Organski and Kugler (1980); Kugler and Lemke (1996).

[6]Organski and Kugler (1980, 61).

[7]Kim and Morrow (1992).

[8]Tessman and Chan (2009, 15-20).

[9]Vasquez (2009, 133).

[10]Werner and Kugler (1996, 188).

challenger[11] or the declining state.[12] In sum, power transition theory has reached a stalemate, with competing logics, different operationalization strategies, and inconsistent empirical predictions.

More recently, rationalist explanations for war identified commitment problems as the causal link between power shifts and war.[13] In the context of a large and rapid power shift, the rising state cannot credibly commit to refrain from exploiting its future advantage. This may lead the declining state to strike preventively. For Powell, "[t]he crucial issue in commitment problems is that in the anarchy of international politics, states may be unable to commit themselves to following through on an agreement and may also have incentives to renege on it. If these incentives undermine the outcomes that are Pareto-superior to fighting, the states may find themselves in a situation in which at least one of them prefers war to peace."[14]

Both these lines of scholarship conceptualize power shifts as the product of exogenous changes in state power, typically through differential rates of economic growth.[15] But exogenous power shifts are rarely large *and* rapid. In fact, most large and rapid power shifts result from changes in relative military power, which are the endogenous product of state decisions. A look at standard measures of state power, such as the Composite Index of National Capabilities (CINC) included in the Correlates of War (COW) dataset, corroborates this point.[16] The CINC index is the average of each state's share of total and urban

[11]Organski and Kugler (1980).

[12]Gilpin (1981); Bueno de Mesquita (1996); Copeland (2000).

[13]Fearon (1995).

[14]Powell (2006, 170).

[15]Fearon (1995); Powell (1999, 2004, 2006, 2011b).

[16]We use data from the National Military Capabilities dataset, v4.0, available at: http://www.correlatesofwar.org/COW2%20Data/Capabilities/nmc4.htm. See: Singer (1987).

population, military personnel and spending, energy consumption, and iron and steel production. Over the period 1816-2007 there are only three instances of an annual power shift greater than 10 percent between any two major powers during peacetime, representing 0.15 percent of all dyads. If we disaggregate the CINC index into its exogenous and endogenous components, however, the greater prevalence of endogenous power shifts becomes clear. To show this, we construct two new indices. CINCexo averages the four exogenous components of the CINC index, total and urban population, energy consumption, and iron and steel production. CINCend averages the two components of the CINC index that are more explicitly the product of endogenous decisions: military personnel and spending. As Powell noted, most shifts resulting from economic growth are in practice quite small.[17] Over the 1816-2007 period, we find five annual shifts greater than 10 percent in exogenous power (CINCexo) between major powers during peacetime, representing 0.25 percent of all dyads. In contrast, we find 111 annual shifts greater than 10 percent in endogenous power (CINCend), or 5.45 percent of all dyads. Figure 1 below shows the frequency of power shifts according to each of the three indices.

— Figure 1 about here. —

Given that most large and rapid power shifts – those severe enough to produce war – are endogenous, their conceptualization as exogenous appears inadequate. Instead, we conceptualize power shifts as the result of decisions to invest in *military* capabilities and focus on the strategic context in which these decisions are made.[18] Focusing on this strategic context highlights the possibility that a state will be the target of a preventive war between

[17]Powell (1999, 163).

[18]We define investment in military capabilities broadly to include any investments (in, *e.g.*, transportation or communications infrastructures) that allow the state to perform additional military missions.

the moment it is suspected of making an investment in military capabilities and the moment these would come to fruition. It also highlights the possibility that a state will be deterred from acquiring additional military capabilities by the prospect of preventive war.

Some recent work has looked at the effects of endogenous power shifts by analyzing strategic interactions between states over resources that affect the future distribution of power.[19] Although these works are a step in the right direction, they provide no guidance on what seem to be the most frequent power shifts, namely, those resulting from investments in military capabilities.[20] Addressing this final step, a literature has emerged on militarization, analyzing state decisions on which fraction of their resources to allocate toward military assets.[21] But these models assume that militarization decisions produce an immediate effect on the balance of power, missing the fact that military capabilities always result from investment decisions *with delayed returns – i.e.*, there is a non-negligible period of time between the decision to invest and the moment these capabilities become available. During this period, an attack against the militarizing state has the potential to destroy the investment in new weapons without having to face the additional capability that would come with their possession, making it possible to deter militarization by the threat of preventive war.

Certainly, there are situations where militarization *appears* to produce an immediate power

[19]Fearon (1996); Chadefaux (2011); Powell (2011a)..

[20]In Fearon (1996), there is an exogenous mapping from resources into military power. Commitment problems may cause war if incremental changes in the distribution of resources lead to large and rapid shifts in the distribution of power. We innovate by endogenizing the mapping from resources to military power, so that additional resources shift the balance of power only if they are invested in military capabilities.

[21]Powell (1993); Slantchev (2005); Baliga and Sjostrom (2008); Meirowitz and Sartori (2008); Jackson and Morelli (2009); Fearon (2010); Meirowitz and Ramsay (2010); Slantchev (2011); Bas and Coe (2012).

shift. Yet this perception results from information problems: it follows from a failure to observe the prior decision to invest in military capabilities. We therefore subsume this particular case into a general theory of militarization and evaluate the effect of the information environment on the likelihood of militarization and of preventive war.

Our theory focuses on the strategic interaction between a state deciding whether to invest in military capabilities (using deterrence terminology, the target) and another considering whether to prevent that investiment from yielding results (the deterrer), capturing the calculations of both. Militarization shifts the balance of power in favor of the target. Whenever this shift is greater than the cost of the investment in armaments, the target should invest. Yet, the target is aware of the risk that its investment, since it has delayed returns, be destroyed by a preventive strike before yielding fruit. It must thus anticipate the deterrer's incentive to prevent the power shift from occurring by striking preventively. The deterrer, for its part, must weigh the cost of preventive action against that of inaction. The cost of action is the cost of preventive war – *i.e.,* the value of the resources destroyed in preventing militarization.[22] The cost of inaction is the value of the concessions the target will demand once it has acquired additional military capabilities, reflecting the target's increased ability to convert policy preferences into outcomes once it possesses more military power.

Our central argument is that, in the context of endogenous power shifts, preventive war only takes place when there is uncertainty about a states' decision to invest in military capabilities. When one state's investment is not guaranteed to be detected by its adversaries, there is an incentive for attempting to present militarization as a *fait accompli*, precluding a preventive

[22]Per Fearon (1995), the cost the deterrer factors into its calculation is the aggregate cost of war, *i.e.*, the total resources destroyed by both sides during the fighting. The rationale for using this aggregate cost measure is that the resources destroyed by the target state in a preventive war could have been (at least partially) transferred to the deterrer in the context of a peaceful bargain had war been avoided.

war. Knowing this, states that fear large and rapid adverse power shifts may launch preventive attacks even when information about the target's investment decision is ambiguous. Overall, the likelihood that endogenous power shifts lead to war hinges on the net effect of militarization, defined as the effect of militarization on the balance of power relative to the cost of preventive war. The higher the net effect of militarization, the more likely preventive wars based on ambiguous evidence, increasing the incidence of mistaken wars. Put differently, powerful states, for which preventive action is less costly, are more likely to launch wars against weaker targets that are wrongly suspected of having made vast military investments.

When an investment decision is perfectly observable, peace always prevails regardless of the magnitude of the power shift it would produce. When the net effect of militarization is small, preventive war is not rationalizable. Since threats of preventive attack by the deterrer are not credible, the target makes its investment. When, however, the net effect of militarization is large, threats of preventive war are credible. The target, understanding that attempted militarization would invite a preventive strike, refrains from investing. Perfect information about the investment allows militarization to be peacefully deterred. In the case of endogenous large and rapid power shifts, therefore, a rising state's inability to commit not to exploit its future power advantage does not *per se* generate war.

In reality, though, the deterrer may be unable to ascertain the target's decision whether to invest in added military capabilities. The target will thus be tempted to develop additional capabilities trying to avoid detection. Understanding this, the deterrer cannot commit to refrain from preventively attacking the target. As a result, the target may develop additional capabilities and, independently, the deterrer may strike preventively.[23] The greater the

[23]In game-theoretic terms, there is a unique equilibrium in mixed strategies. This is not to say that 'anything goes'. Our unique equilibrium provides specific probabilities for each event. The fact that

quality of the information available to the deterrer about the target's investments in military power, the more confident it can be about deterring militarization peacefully. As the quality of that information increases, then, the lower is the likelihood of both conflict and militarization, and thus the lower is the likelihood of *mistaken* preventive wars – wars launched against an "innocent" target.

Moreover, the likelihood of preventive war depends on its effectiveness, defined as the period of time during which it ensures the target will not militarize. As it becomes more effective, preventive war is more tempting to the deterrer, making peace more difficult to enforce. At the same time, the threat of a preventive war becomes more credible, making militarization less likely and therefore preventive wars more likely to be mistaken.

Before we move on, it is important to acknowledge that there are at least two information problems at play in the context of endogenous power shifts, one related to whether a state has decided to invest in military capabilities and another related to the configuration of the particular weapons-program in which the state may have invested. Three reasons lead us to restrict our attention to the former. First, although information problems stemming from hidden actions are the appropriate lens to understand how militarization may cause war, they have received relatively little attention within the rationalist framework.[24] Military capabilities are acquired only after they are pursued, so that the key information problem is

---

the deterrer and target are uncertain about each other's action is realistic. Furthermore, a mixed-strategy Nash equilibrium can be understood as the pure strategy Nash equilibrium of close-by game of incomplete-information, where the players' payoffs are subject to private shocks, Harsanyi (1973).

[24]So far the literature has devoted more attention to problems stemming from incomplete information about actors' types. Examples of games where war is due to imperfect information about hidden actions include Meirowitz and Sartori (2008); Schultz (2010).

whether the state has taken the action of investing in such capabilities. Second, our treatment allows us to address uncertainty about the configuration of the particular weapons-program in which the state has invested by incorporating it into the cost of preventive war. As this second type of uncertainty increases, the deterrer needs to augment the target-set of a putative preventive strike in order to reach a particular level of effectiveness. Third, there is no necessary logical link between these two types of information problem.[25] In sum, we agree that the ability of the target state to generate uncertainty about the configuration or location of its weapons-development sites may be linked to the effectiveness of preventive war but opt for bracketing this type of information problem in order to focus on the, in our view, more basic problem of uncertainty about whether an investment in military capabilities has been made.[26]

We now formalize the theory through a game-theoretic model.

---

[25]In some cases, there is certainty that an investment has been made but uncertainty about the particular configuration and location of the weapons program. Such was the case with the Soviet nuclear-weapons program. The United States had no doubt the Soviet Union was developing nuclear weapons but did not possess sufficient information to build a target-set that would make a surgical preventive strike possible (Holloway, 1994; Gordin, 2009; Rhodes, 1995). In other cases, there is certainty about the configuration and location of a particular dual-use investment program but uncertainty that its aim is the production of new military capabilities. Such is, in the opinion of many, the case of the current Iranian nuclear program. U.S. decisionmakers have a good sense of the target-set related to the program, but uncertainty about whether the program aims at the production of nuclear weapons remains.

[26]It might be useful to extend our analysis to think about the set of sites that a deterrer includes in a preventive attack, opting for a low cost/low effectiveness strike or a high cost/high effectiveness strike, but that is a different question than what we are asking.

# 3. The Model

## *3.1. Basic Framework*

We model a strategic interaction between two states, $T$ (the 'target') and $D$ (the 'deterrer'). $T$ decides whether to acquire additional military capabilities, *i.e.* 'militarize', and $D$ decides whether to launch a preventive war. $T$'s militarization decision is a costly investment with delayed returns, *i.e.*, it costs $k > 0$ in period $t$ and comes to fruition in period $t + 1$ if $D$ does not strike preventively.[27]

In keeping with the literature, states face two 'problems.' The first is a *commitment problem*, in that they are unable to commit to abiding by any agreements in the future.[28] The second is an *information problem*, meaning here that $D$ may be imperfectly informed about $T$'s investment decision in period $t$ when deciding whether to strike.[29] If $T$ does not invest ($I_t = 0$), $s_t = 0$. If $T$ invests ($I_t = 1$), $s_t = 1$ with probability $p_s \in [0,1]$ and $s_t = 0$ with probability $1 - p_s$. Intuitively, $s_t = 1$ represents an unambiguous signal of $T$'s militarization, while $s_t = 0$ is an ambiguous signal. We say that information problems are absent if the signal is perfectly informative ($p_s = 1$).

After receiving its signal, $D$ decides whether to declare war ($dw_t = 1$ if it declares war, $dw_t = 0$ if it does not). The alternative to a preventive war is a peaceful division of the pie, where $D$ offers a share $z_t$ to $T$, keeping $1 - z_t$ for itself. $T$ then decides whether to

[27]We reserve for future work the intermediate case where preventive war is successful with probability strictly between 0 and 1.

[28]Fearon (1995).

[29]More generally, states may be partially informed about the resolve, intent, type, or actions of other states (Fearon, 1995).

accept $D$'s offer $z_t$ ($a_t = 1$ if it accepts the offer, $a_t = 0$ if it does not). If $T$ accepts the offer $z_t$, it is implemented at $t$. If $T$ rejects $D$'s offer, war ensues. In any war, country $i$ gets a payoff $w_i(M_t) \geq 0$ in period $t$, where $M_t$ represents $T$'s current military capabilities ($M_t \in \{0,1\}$, where $M_t = 1$ if and only if $T$ has acquired additional military capabilities). We assume that war is inefficient, *i.e.*, for any $M_t \in \{0,1\}$,

$$w_T(M_t) + w_D(M_t) < 1 \quad (1)$$

and call $1 - w_T(0) - w_D(0)$ the *cost of a preventive war*. We define a preventive war to be *mistaken* if it is made against a state that is not investing in military capabilities.

Furthermore, we assume that countries discount the future (by factor $\delta \in (0,1)$). We call $\delta[w_T(1) - w_T(0)]$ the *effect of militarization*, i.e. the improvement in $T$'s discounted war payoff resulting from its investment in military capabilities.

Consider first a simple, two-period version of the game.

## *3.2. The Two-Period Game*

### 3.2.1. Timing and Solution Concept

In this version, $T$ has a single opportunity to invest in military capabilities, in period 1 (for the game tree, see Figures 2.1 and 2.2).[30] After $T$'s decision to invest in period 1, Nature sends a signal $s_1$, then $D$ decides wheher to declare war or offer a share $z_1$, which $T$ may accept or reject. In period 2, $T$ cannot invest in military capabilities[31] and its prior

[30]Terminal nodes give the stage game payoff of each state, with the target's payoff on top and the deterrer's payoff below.

[31]This is without loss of generality, since militarization is a costly investment with delayed returns and period 2 is the last period of the game.

investment decision is revealed ($s_2 = I_1$), through the occurrence, or not, of a military exercise or test. After this signal, $D$ declares war or offers a share $z_2$, which $T$ may accept or reject.

— Figures 2.1 and 2.2 about here. —

We solve for Perfect Bayesian Equilibria (PBE) of this dynamic game of imperfect information. This requires that at each information set, play is sequentially rational given beliefs and beliefs are updated using Bayes' rule whenever possible.[32]

### 3.2.2. Solving the Game

The solution in period 2 is straightforward. First, peace prevails since war is inefficient and there is no commitment problem or information problem. Second, the terms that $T$ can extract increase with its military capabilities:

***Proposition 1.*** *In period 2, there is always peace, where D offers* $z_2^* = w_T(M_2)$ *and T accepts any* $z_2 \geq w_T(M_2)$.

***Proof.*** *Straightforward.*

Thus militarization is a costly attempt to extract resources from another state.

Moving up, we reach general conclusions about the outcome of period 1. First, if the effect of militarization is smaller than the cost of preventive war, preventive war is not rationalizable and peace must prevail. Indeed, if the maximum concessions that $D$ would have to make to $T$ in period 2 are smaller than the cost of preventive war, i.e. the bargaining surplus that $D$ does not extract by going to war, $D$ cannot rationalize the decision to go to war.

[32]Fudenberg and Tirole (1991, 325-326). If there is no investment in equilibrium, we assume that off-the-equilibrium-path beliefs after $s_1 = 1$ are that $T$ invested, the only decision that could generate this signal.

Second, if the effect of militarization is smaller than the cost of investment, militarization is not rationalizable and peace must prevail. Indeed, if the maximum benefit that $T$ can extract from militarization is smaller than the cost of investment, $T$ cannot rationalize the decision to invest in military capabilities. Thus, $D$ need not fear the target's militarization when facing an ambiguous signal and peace prevails. In sum:

***Proposition 2.*** *In period 1, there is always peace if the effect of militarization is smaller than the cost of a preventive war or smaller than the cost of the investment.*

***Proof.*** *See the appendix.*

The situation is richer if preventive war and the investment in military capabilities are both rationalizable.

***Proposition 3.*** *Consider period 1 and assume that the effect of militarization is greater than the cost of a preventive war and greater than the cost of the investment.*

*(i) If the signal is sufficiently informative, i.e.,*

$$(1 - p_s)\delta[w_T(1) - w_T(0)] \leq k \tag{2}$$

*then peace prevails.*

*(ii) If the signal is not sufficiently informative, i.e., (2) fails, then* $T$ *militarizes with probability* $q^* = \frac{1}{p_s + (1-p_s)\frac{\delta[w_T(1)-w_T(0)]}{1-w_T(0)-w_D(0)}}$. *After* $s_1 = 1$, *preventive war occurs. After* $s_1 = 0$, *peace prevails with probability* $r^* = \frac{k}{(1-p_s)\delta[w_T(1)-w_T(0)]}$ *and preventive war occurs with probability* $1 - r^*$.

***Proof.*** *See the appendix.*

This proposition shows that if the informativeness of the signal is sufficiently high (case i), the threat of preventive war is sufficient to deter any investment. $D$ is confident that it can detect militarization attempts, and need not attack preventively when it receives an ambiguous signal. If the informativeness of the signal is not sufficiently high (case ii),

strategic uncertainty remains and war occurs with positive probability. In sum, we conclude:

***Theorem 1.*** *When shifts in the balance of power are endogenous, commitment problems may cause war only if information problems are present.*

***Proof.*** *Follows from propositions 2 and 3.*

This stands in contrast with the traditional rationalist framework, where commitment problems may cause war even in the absence of information problems.[33] The difference comes from that fact that, when shifts in the balance of power are the result of a costly investment with delayed returns, the declining state can use the threat of a preventive strike to deter the target from militarizing. Therefore, large and rapid endogenous shifts in the balance of power do not occur and peace prevails. This theorem thus highlights a particular type of commitment problem that may cause conflict. Whereas when shifts in the balance of power are exogenous, commitment problems may cause war because a state cannot commit to refrain from exploiting its future power advantage, when shifts in the balance of power are endogenous, commitment problems may cause war because a state cannot commit to refrain from investing in military capabilities.

Information problems are thus necessary for conflict. The greater is their severity, the greater is the likelihood of conflict. Indeed, the more informative is the signal, the more effective are threats of preventive war and the more confident is the deterrer that it need not worry about an ambiguous signal. As a result, we conclude:

***Corollary 1.*** *The greater is the informativeness of the signal*

*(i) The more stringent become the conditions under which preventive wars happen with positive probability, and if such conditions are met,*

*(ii) The smaller is the probability of preventive war;*

*(iii) The smaller is the share of preventive wars that are mistaken.*

[33]Powell (2006, 182).

***Proof.*** *See the appendix.*

## *3.3. The Infinite-Horizon Game*

We now build an infinite-horizon game to assess the robustness of the previous results and investigate two additional questions. First, is war avoidable when states are engaged in an ongoing relationship? A finite game produces a unique solution, where the deterrer extracts the most favorable peaceful terms. We may worry that by precluding 'concessions' this set-up is responsible for war. Second, how does the likelihood of war depend on its effectiveness, *i.e.* the period of time during which it effectively demilitarizes the target?[34]

In the infinite-horizon version of the previous game, the target first decides whether to invest in military capabilities. If it makes the investment and peace prevails, the target possesses these additional military capabilities, and keeps them from then on. If a preventive war occurs, it demilitarizes the target for $N$ periods, where $N$ is the *effectiveness* of a preventive war.

First, we note that a necessary condition for war is that both militarization and preventive war are rationalizable. Yet even in those circumstances, we may believe that war can be avoided. The target should be willing to accept some concessions, short of the terms it would extract after militarization, to save the cost of investment. The deterrer should want to offer some concessions, knowing that if it later reneged on them, the target would militarize, thus extracting greater concessions in the future. Generally, peace can be enforced if the concessions needed to prevent militarization are smaller than the cost of a preventive war, averaged over the effectiveness of the preventive war, *i.e.* if the cost of peace is smaller than

[34]A closely-related question would be to vary the likelihood that a preventive war is successful in the two-period game, but the current question is closer to the relevant policy question.

the cost of war.[35]

The standard logic suggests that efficiency could be enforced if players are sufficiently patient.[36] If a deviation entails a short-term benefit, *e.g.* reneging on a payment, and a long-term punishment, *i.e.* reverting to the inefficient equilibrium, then as players become more patient, the threat of future punishment looms larger and efficiency can be enforced.

Yet this does not fully characterize the strategic interaction between the target and the deterrer. For the target, the best deviation is to militarize. This entails a short-term cost, *i.e.* the cost of investment, and a long-term benefit, *i.e.* greater concessions resulting from militarization. As the target becomes more patient, it demands greater – not smaller – concessions. For the deterrer, preventive wars produce a sustained benefit, since they obviate the need for concessions while the target is effectively demilitarized. Thus, war may occur even in an infinite-horizon game where countries are very patient.

Understanding that preventive war is not an artefact of the finite-horizon game, we can first confirm that an increase in the quality of the signal has the same effect as in the two-period game.[37] One additional mechanism, germane to the infinite-horizon setting, is that a better signal reduces the cost of peace. Indeed, as the signal becomes more informative, the target is less tempted to militarize and the deterrer is less concerned about ambiguous signals. Thus, the deterrer can make smaller concessions to prevent militarization: peace and efficiency are achieved more easily.

Next, we can analyze the role that the effectiveness of a preventive war plays on its likelihood. As a preventive war becomes more effective, it becomes more attractive to the deterrer, since it lifts the need to make concessions for a longer period of time. The first,

[35]Coe (2011).

[36]As stated in the 'Folk Theorem'.

[37]Corollary 3 in the Appendix.

direct implication is that it is more difficult to sustain efficiency. The second effect is that threats of preventive war are more credible. As a result, the target is less likely to militarize, so that preventive wars are less likely, and more likely to be mistaken. Formally, this means:

***Corollary 2.*** *The greater is the effectiveness of a preventive war*

*(i) The less stringent become the conditions under which preventive wars happen with positive probability, and if such conditions are met,*

*(ii) The smaller is the probability of preventive war;*

*(iii) The greater is the share of preventive wars that are mistaken.*

***Proof.*** *See the appendix.*

To summarize the results of our formalization, under conditions of perfect information about a state's decision to invest in additional military capabilities, the specter of a potentially large and rapid power shift never leads to preventive war. Rather, such large and rapid power shifts can be deterred. When uncertainty about the investment is introduced, peace prevails whenever the effect of militarization is smaller than the cost of war. If this effect is also smaller than the cost of the investment, militarization is not rationalizable. If, however, the effect of militarization is greater than both the cost of the investment and the cost of a preventive war, peace may break down. Both the set of circumstances under which war becomes possible and the probability that it will break out increase as the quality of the information decreases. Furthermore, as the likelihood of preventive war increases, so does the probability that it will be mistaken, *i.e.*, waged against an unarming target. Finally, the more effective is a preventive war in ensuring the absence of future investments in additional military capabilities, the broader the set of conditions under which war becomes possible, the lower the probability that it will break out, and the greater the likelihood that it will be mistaken. As with exogenous power shifts, investments in military capabilities may generate commitment problems that lead to produce war. Unlike exogenous power shifts, however,

such commitment problems will only lead to war when information is imperfect.

# 4. The U.S.-led Invasion of Iraq

In principle, the theory presented in the previous sections applies to any interaction between two actors where one can take a costly and potentially hidden action to affect the outcome of conflict.[38] Here, we focus on the model's lessons for an important topic: preventive counter-proliferation wars. More precisely, we offer an account of the 2003 invasion of Iraq, shedding light on important mechanisms that appear to have been at work.

The 2003 U.S. invasion of Iraq is a controversial case, with scholars debating the proper framework to analyze the crisis and also its lessons for international-relations theory. Our purpose here is not to claim that our theory offers a definitive, or complete explanation for the case. Rather, it is to show how our theory illuminates aspects of the Iraq case that have been heretofore underexplored. First, we show how the war was triggered by U.S. fear of an endogenous power shift resulting from a suspected Iraqi WMD program, the core dynamic in our theory. Second, we highlight how the terrorist attacks of September 11, 2011, led U.S. decisiomakers to attribute more worrisome consequences to suspected Iraqi nuclearization, accounting for the timing of the war in terms consistent with our theory. Third, we detail how, in agreement with our model, the war was the result of the imperfect information U.S. leaders possessed about the Iraqi WMD program. Finally, we use our theory to provide an account of why Washington launched an attack against Iraq rather than North Korea, a state that in fact possessed a nuclear-weapons program. We end by contrasting our interpretation

[38]For example, the model could apply to the context of civil wars in weakly-institutionalized settings, where a rebel group attempts to acquire weapons or build up capabilities in order to challenge a national government.

of the case with another prominent account in the literature[39] and discuss the war's lessons for international-relations theory.

On March 20, 2003, a U.S.-led coalition invaded Iraq. The United States' main motivation for the war was to prevent suspected Iraqi nuclearization, which Washington thought would bring about a large and rapid shift in the balance of power in favor of Iraq. The invasion came after a protracted United Nations inspection process aimed at guaranteeing the end of the Iraqi WMD program, including its nuclear component. Arguing that Saddam Hussein's regime was not fully cooperating with the inspectors, the U.S. administration justified the invasion as the only way to obviate the threat of a nuclear-armed Iraq. Coalition forces quickly prevailed, deposing Saddam's regime with relative ease. Baghdad fell a mere twenty one days after the invasion was launched and the major-operations phase of the war ended six days later.

During the run up to the invasion, the U.S. government's *casus belli* rested on suspicion that Saddam was developing WMD – including nuclear weapons – thus presenting an imminent threat.[40] Avoidance of a possible large and rapid power shift was therefore at the center of the case for war. Granted, a multiplicity of other arguments was introduced to justify the invasion, including Iraqi human-rights abuses and the need to democratize the Middle East.[41] But WMD played the central role in the U.S. administration's case for forcible regime change. At his 2002 State of the Union Address, President Bush had pledged not to "permit the world's most dangerous regimes to threaten us with the world's most destructive weapons."[42] He soon thereafter reiterated the argument: "Saddam Hussein must understand

[39]Lake (2010/11).

[40]Freedman (2004).

[41]Tenet (2007, 301).

[42]Bush (2002).

that if he does not disarm, for the sake of peace, we, along with others, will go disarm Saddam Hussein."[43] On February 5, 2003, in a highly publicized attempt to legitimize the invasion, Secretary of State Colin Powell made a presentation to U.N. Security Council titled 'Iraq: Failing to Disarm.' Powell argued that "possession of the world's most deadly weapons was the ultimate trump card [and] the United States would not – could not – run the risk to the American people that Saddam Hussein would one day use his weapons of mass destruction."[44] The day after Baghdad fell, in the absence of any immediate WMD findings on the ground, White House press secretary Ari Fleischer repeated it: "[W]e have high confidence that they have weapons of mass destruction. *That* is what this war was about."[45] In short, the case presented by the U.S. administration had at its core concerns about a large and rapid shift in the balance of power in favor of Iraq as a result of Baghdad's WMD investments.

Iraq's nuclear acquisition would represent a large and rapid power shift that would make Saddam immune to any externally-driven regime-change efforts, ending his vulnerability to U.S. military action. The cost of war against a non-nuclear Iraq, in contrast, was expected to be relatively low, as U.S. forces would, given the precedent of the 1991 Gulf War, no doubt prevail. In the past, Iraq had possessed impressive military forces. But already in 1991 Saddam found himself in a weaker position. During the Gulf War, Iraqi forces were

[43]CNN (2003).

[44]UN (2003).

[45]House (2003), emphasis added. Later, in the absence of WMD findings, Paul Wolfowitz, one of the architects of the war, claimed that there were four motivations behind the invasion: Iraq's WMD, its support for terrorism, the nexus between these two, and the regime's treatment of the Iraqi population. According to Wolfowitz, "for reasons that have a lot to do with the U.S. government bureaucracy we settled on the one issue [WMD] that everyone could agree on" (Wolfowitz, 2003).

decisively expelled from Kuwait by a U.S.-led coalition.[46] Significantly, Iraq lost much of its military materiel in that conflict.[47] Iraq's relative weakness also gave the U.S. free rein to impose a severe sanctions regime that further crippled Iraqi capabilities. By 2003, Iraq's military was poorly trained and equipped. The balance of power clearly favored the United States. In this context, U.S.'s ability to depose Saddam promptly was never in doubt. Although fighting a defensive war in its own country against an expeditionary force, the Iraqi army lost all engagements with coalition forces and ended up suffering 9,200 fatalities – or more than fifty times the 172 lives lost by coalition forces.[48] This cost was several orders of magnitude smaller than the expected cost of deposing a nuclear-armed Saddam. This difference accounts for U.S. insistence in guaranteeing Iraqi non-nuclear status, if necessary by force.

The second point our theory highlights in the case of the Iraq War is the role of the terrorist attacks of September 11, 2001, in accounting for the timing of the war. Iraq had been suspected of not abiding by the terms of the post-Gulf War cease fire, but an invasion did not happen until the events of 9/11 undermined the U.S. administration's trust in the intelligence community's ability to detect security threats in a timely manner.[49] In the aftermath of the

[46]Gordon and Trainor (1995).

[47]On February 22, 1991, well into the air campaign against Iraqi forces but before the ground war started, Saddam accepted a plan to withdraw from Kuwait in 21 days. The U.S. administration, however, rejected it, demanding a seven-day withdrawal, which would force Saddam to leave behind his materiel, in particular armored vehicles. The ground war resulted from Saddam's rejection of this demand, which would leave him with little more power than a military defeat. Most Iraqi materiel deployed in Kuwait ended up being destroyed, however (Pape, 1996).

[48]A U.S. Army study attributes the outcome of the war to "a synergistic interaction between advanced coalition technology and a major skill differential" (Biddle et al., 2004, v).

[49]In the language of our model, there can be two interpretations: the perceived effect of militarization

attacks, the Bush Administration became particularly worried, based on flimsy intelligence, with the possibility of an Iraqi nuclear handoff to a terrorist group for use against the U.S. targets.[50] In any case, a nuclear Iraq would pose a particularly thorny problem for U.S. security in that it would not only be able to provide a terrorist group with a nuclear weapon – it would also make any U.S. retaliation exceedingly costly. Specifically, a U.S. attempt to depose a nuclear-armed Saddam by force would place him in a "use them or lose them" situation, and was therefore unfeasible.[51] As then-National Security Advisor [NSA] Condoleeza Rice put it in her memoirs:

> A policy maker confronted with one assessment that says that Baghdad "could make a nuclear weapon within several months to a year" should it "acquire sufficient fissile material from abroad" and the INR alternative view that could not speak to timing is not likely to take the risks of accepting the latter, particularly after 9/11 and the specter of WMD terrorism. ... We'd failed to connect the dots on September 10 and had never imagined the use of civilian airliners as missiles against the World Trade Center and the Pentagon; that an unconstrained Saddam might aid a terrorist in an attack on the United States did not seem far-fetched.[52]

This higher sensitivity to a low probability event made peace harder to sustain. In this sense, the invasion of Iraq represented the paradigmatic application of the 'one percent doctrine,' attributed to Vice President Dick Cheney. This doctrine suggests that, in the post-9/11 security environment, the United States must deal with 'low-probability, high-impact' events

---

increased and the perceived informativeness of the signal decreased.

[50]Rhodes (2010, 269).

[51]Posen (1997, 22-26).

[52]Rice (2011, 169-170).

as if they were certain.[53] Thus the United States acted as if Iraqi nuclearization were all but certain and launched a preventive war, which subsequently proved mistaken.

Next, our model shows how the U.S. administration's inability to eradicate uncertainty about the status of Iraq's nuclear program was essential to the breakdown of peace. Enjoying a preponderance of power and determined to avoid the repetition of the September 11, 2011, terrorist attacks, the United States set a high standard of evidence for proof of Iraq's disarmament. Since the United States could not be certain of detecting a Iraqi nuclear program, it feared Iraq would be tempted to build a nuclear weapon and place Washington before a *fait accompli*. Unable to prove this negative point and unwilling to run the risk of facing a nuclear Iraq, the United States decided to launch a preventive war that ultimately proved to rest on mistaken grounds.[54]

---

[53]Suskind (2006). It might be argued that such extreme form of risk aversion can hardly be called "rational." But in terms of our model, the 'one percent doctrine' represents an increase in the expected effect of Iraqi nuclearization. In the model, a rational, risk-neutral decision-maker weighs the expected cost of intervention and the expected effect of nuclearization. This decision-maker may decide to go to war, even if the expected likelihood of proliferation is low, if the effect of nuclearization is expected to be sufficiently high. Risk aversion would simply exacerbate the cost-benefit analysis of a rational decision-maker. In other words, additional assumptions on the origin of the perceived cost and on risk aversion can strengthen the incentives to go to war, but they are not necessary.

[54]In the language of our model, when (i) the quality of the information available to the deterrer about the target's decision to nuclearize is not perfect and (ii) the effect of nuclearization on the balance of power is sufficiently large relative to the costs of preventive war, equilibrium is reached in mixed strategies. This might be interpreted as meaning that, under these circumstances, our model makes indeterminate predictions and is thus unfalsifiable. This is not the case, however. In this area of the parameter space, our model rules out open nuclearization. Openness about an investment in nuclear

Retrospectively, much has been made of intelligence failures, but the problem was that in order to change U.S. policy, intelligence reports would have to prove that Iraq *did not have and would not develop* WMD. Unfortunately, it is well nigh impossible for intelligence services to prove a negative. In fact, Western intelligence communities were determined not to miss the next threat and were therefore taking a more sanguine line regarding the Iraq's WMD program. As Jervis notes, "the belief that Iraq had active WMD programs was held by *all* intelligence services, even those of countries that opposed the war."[55] Based on the imperfect information the United States possessed about Iraq's WMD capabilities, "[a] responsible judgment could not have been that the programs had ceased."[56] Jervis concludes:

> At best, intelligence could have said that there was no firm evidence that Saddam had stockpiles of chemical and biological weapons or was actively pursuing nuclear bombs. It could not have said that he had ceased his efforts. ... Furthermore, intelligence could not have said that Saddam would not pursuit of WMD at some

---

weapons would (by providing the deterrer with clear information about the target's decision to nuclearize) attract a preventive attack with certainty. Instead, our model predicts that as the effect of nuclearization increases relative to the costs of preventive war (and maintaining the quality of the information fixed), the deterrer will be more tempted to launch a preventive strike. This rationale will be internalized by the target, which can implement one of two strategies. It will either abstain from developing nuclear weapons for fear of attracting a preventive attack or it will nuclearize for fear that a preventive attack against it will be launched anyway given the imperfect quality of the information it can make available about its decision *not* to nuclearize, which will make it impossible for the deterrer to be certain that it is not in fact facing a nuclearizing state. Iraq opted for the former, but nonetheless attracted a preventive attack.

[55]Jervis (2010, 134, Jervis's emphasis).

[56]Jervis (2010, 155).

point in the future.[57]

In part, uncertainty about the Iraqi WMD program was magnified by the strategic situation, in which the United States benefited from great relative power. Were Saddam ever to announce or acknowledge publicly his investment in a nuclear-weapons program – as Stalin did in 1945 – he would invite likely preventive action from the United States.[58] The U.S. administration, for its part, facing relatively low costs of war as a result of its power preponderance, demanded a high degree of certainty that Iraq had given up developing nuclear weapons in order to refrain from launching a preventive attack.

According to the terms of the 1991 ceasefire, Iraq was forbidden from developing WMD and long-range missiles. To verify Iraqi compliance, UNSCOM was created and its inspectors deployed in the country. After repeated violations of its disarmament obligations, growing tensions, and U.S. airstrikes, Iraq evicted UNSCOM in December 1998. Scott Ritter, the Commission's chief inspector at the time, noted that "without effective monitoring, Iraq can in a very short period of time measured in months, reconstitute chemical biological weapons, long-range ballistic missiles to deliver these weapons, and even certain aspects of their nuclear weaponization program."[59] Consequently, the U.N. was unable to certify Iraq's complete disarmament.[60] Under mounting pressure, Saddam finally agreed to let U.N. inspectors back in on September 26, 2002. Earlier that month, the British International

[57]Jervis (2010, 126). Jervis convincingly dismisses the thesis that politicization was responsible for the intelligence failure on Iraq's WMD: "while alternatives [to the picture intelligence reports painted of Saddam's putative WMD programs] should have been considered, doing so probably would not have changed the estimates" (Jervis, 2010, 128). For a dissenting opinion, see Rovner (2010, chapter 7).

[58]On Soviet openness about its nuclear program, see: Craig and Radchenko (2008); Holloway (1994).

[59]Ritter (1998).

[60]Thompson (2009).

Institute for Strategic Studies concluded that Iraq possessed the scientific apparatus to "assemble nuclear weapons within months if fissile material from foreign sources were obtained."[61] In fact, Hans Blix, the chief inspector for UNMOVIC (which had replaced UNSCOM), was unable to ascertain whether Iraq possessed any WMD. UNMOVIC's work was plagued by discrepancies between Iraqi reports of WMD quantities produced and destroyed. According to Blix, UNMOVIC's reports "do not contend that weapons of mass destruction remain in Iraq, but nor do they exclude that possibility. They point to lack of evidence and inconsistencies, which raise question marks, which must be straightened out, if weapons dossiers are to be closed and confidence is to arise."[62] In his final pre-war presentation to the U.N. Security Council, Blix was remarkably ambiguous:

> It is obvious that, while the numerous initiatives, which are now taken by the Iraqi side with a view to resolving some long-standing open disarmament issues, can be seen as 'active,' or even 'proactive,' these initiatives 3-4 months into the new resolution cannot be said to constitute 'immediate' cooperation. Nor do they necessarily cover all areas of relevance.[63]

Indeed, there was a broad political consensus in Washington that Saddam possessed, or intended to acquire, WMD.[64] Senator John Kerry from Massachusetts, soon to become the Democratic presidential nominee, stated: "According to the CIA's report, all U.S. intelligence experts agree that Iraq is seeking nuclear weapons. There is little question that

[61]Chapman (2002, 4).

[62]Blix (2003b).

[63]Blix (2003a).

[64]For a detailed counterfactual analysis of what a Gore Administration would have done that conclusively dispels the notion that war was caused by the Bush Administration's idiosyncratic preferences, see Harvey (2012).

Saddam Hussein wants to develop nuclear weapons."[65]

As became clear after the war, Iraq possessed no WMD and had no consistent WMD programs. Furthermore, Saddam seems to have abandoned his nuclear program – though not his intention of resuming it – years before the war. The United States had been able to terminate the Iraqi nuclear-weapons program through the sanctions regime in place since the 1991 war.[66] All the Iraq Survey Group could find after the invasion was evidence that Saddam intended to revive such programs if and when sanctions were lifted. The Group's Final Report states:

> Saddam wanted to recreate Iraq's WMD capability – which was essentially destroyed in 1991 – after sanctions were removed. ... Saddam aspired to develop a nuclear capability – in an incremental fashion, irrespective of international pressure and the resulting economic risks – but he intended to focus on ballistic missile and tactical chemical warfare (CW) capabilities.[67]

Finally, our theory helps account for why the United States set its crosshairs on Iraq rather than another plausible target: North Korea. The United States had long suspected of North Korea of intending to develop nuclear weapons and had gone to the brink of war over this issue in the 1994 crisis.[68] Furthermore, a mere two weeks after the fall of Baghdad, North Korean officials announced to their U.S. counterparts that they possessed nuclear weapons, a claim validated by the C.I.A. in August 2003.[69] Why, then, did Washington pursue a

[65]Kerry (2002, S10172-10173).

[66]CIA (2004).

[67]CIA (2004, 1). For a previous report on WMD findings in Iraq after the invasion, see Kay (2004).

[68]Lee and Moon (2003, 141

[69]Pollack and Reiss (2004, 278).

mistaken preventive strike against Iraq, whose nuclear program had in practice ended, rather than target North Korea, whose nuclear program was fast approaching completion?

The strategic interaction the United States had with North Korea at the time differs in a crucial aspect. Given Pyongyang's ability to impose heavy costs on the United States and its allies in case of a military conflagration, the cost of a preventive war against North Korea was expected to be much greater than that of a war against Iraq. By the same token, the effect of North Korean nuclearization would be relatively small, given the limited range of policy options available to the United States even vis-à-vis a non-nuclear North Korea. Indeed, Pyongyang could turn Seoul into a "sea of fire" using its conventional artillery.[70] At the peak of the 1994 crisis between the United States and North Korea over the latter's nuclear program, on June 17, then-President Kim Youngsam of South Korea reportedly told President Clinton over the phone:

> I can't make the Korean peninsula a battlefield. Once war breaks out, numerous citizens and soldiers would be killed, the economy would be ruined, and foreign capital would leave the country. Even if the United States carries out surgical air strikes, North Korea will immediately fire back to the major South Korean cities. Many lives were gone during the Korean War, and now we have stronger weapons. No war again.[71]

Decisionmakers in Washington consistently made a similar assessment. U.S. military estimates dating from 1994 expected that in case war broke out in the Korean peninsula "the US would sustain 52,000 casualties, and South Korea up to 490,000, in the first 90 days of a

[70]Perry (2002, 121).

[71]Kim (2001, 316-317).

full-blown conflict."[72] By 1999, then-former Secretary of Defense William Perry, reporting to Congress on the evolution of the situation, argued that "the intensity of combat in another war on the Peninsula would be unparalleled in U.S. experience since the Korean War of 1950-53."[73] Along similar lines, then-NSA Rice recalls in the context of President George W. Bush's misgivings about the North Korean regime's reactivation of its nuclear facilities in 2002 that "given the consequences of conflict on the Korean peninsula, there didn't seem to be many alternatives ... [T]he Pentagon wanted no part of armed conflict on the Korean peninsula. We were without a workable strategy."[74]

Against this background of a potentially costly preventive war, the U.S. government reached a peaceful agreement with Pyongyang – the 1994 Agreed Framework. In this agreement, North Korea committed to freezing its nuclear program. In exchange, the United States committed to building two light water nuclear reactors (from which weapons-grade fissile material cannot be derived) in North Korea, providing the country with energy supplies while these reactors were being built, issuing formal security assurances precluding the use of U.S. nuclear weapons against North Korea, and working towards the end of sanctions against the Pyongyang regime. This agreement helped ensure North Korea's non-nuclear status for almost another decade. Unfortunately, this type of grand bargain could not be reached with Iraq, where the net effect of nuclear acquisition was perceived to be greater.

A rationalist approach can thus account for important features of the Iraq war. Others disagree. In an important article, Lake argues that the rationalist framework cannot account for the war because of two shortcomings.[75] First, the two main causes of conflict in the

[72]Ayson and Taylor (2004, 264).

[73]Perry (October 12, 1999, 3).

[74]Rice (2011, 159).

[75]Lake (2010/11).

rationalist framework – information and commitment problems – are constant features of international relations.[76] Thus, according to Lake, the rationalist framework cannot explain why bargaining broke down in this particular instance but not others, such as the run up to North Korean nuclearization. Second, the main mechanism that produced the Iraq war lies outside the rationalist framework. For Lake, the war was caused by cognitive biases and irrational self-delusion.[77] While Washington failed to obtain better information about the cost of war and to incorporate mounting evidence that Iraq had disbanded its WMD program, Saddam failed to ascertain the level of U.S. resolve about fighting the war and continued to send mixed signals in an attempt to placate both its regional enemies and the United States. For Lake, these actions counter the basic assumption of rationality. As a solution, he calls for "a behavioral theory of war."[78]

In our view, however, neither criticism stands. First, the fact that the rationalist framework builds on fundamental problems of weakly institutionalized settings, *i.e.,* information and commitment problems, is a strength not a weakness. The framework's predictive power lies in its ability to highlight how particular circumstances exacerbate such problems, thus

[76]See also Gartzke (1999); Fearon (1995). In fact, Lake misrepresents the role of commitment problems in extant rationalist explanations for war as applied to Iraq. According to him, "Bargaining theory suggests that a fundamental cause of the war, and a key bargaining failure, was Iraq's inability to commit credibly not to develop WMD or share the resulting technologies with others, including terrorists." We believe this is correct as an account of the causes of the Iraq war – and indeed is the central dynamic of our model – but this is not what extant rationalist treatments of the causes of war would argue. In the existing literature, power shifts are exogenous, so a rising power has problems committing not to exploit its future advantage. The key contribution of our model is precisely to bring the type of commitment problem Lake mentions into focus.

[77]Lake (2010/11, 9).

[78]Lake (2010/11, 10). See also Lake and McKoy (2011/12).

increasing the likelihood that bargaining breaks down and war occurs. As we argued above, we believe that the cost of inaction against possible threats was perceived to be greater after 9/11 and that the cost of a preventive war, relative to the effect of nuclearization, was perceived to be greater against Iraq than against North Korea. The model thus highlights some factors that may help account for the timing of the war against Iraq and for the absence of war on the Korean peninsula in the same time period.

Second, Lake's call for a behavioral revolution does not appear warranted. Granted, the standard rationalist framework has important shortcomings. In our minds, the key commitment problem was that Saddam Hussein could not refrain from investing in nuclear weapons, and the key information problem was that Washington could not be certain of detecting an Iraqi nuclear investment before it yielded fruits. The standard rationalist framework cannot account for such problems because it assumes that shifts in the balance of power occur exogenously. Yet it is possible to capture such problems and remain within the rationalist framework by endogenizing such power shifts, as we do above, without launching a behavioral revolution.

The other aspects that Lake emphasizes – misperceptions and multiple audiences – may have played a role in the run-up to the war. Yet it is not clear that including misperceptions or multiple audiences is necessary for an account of the war, or that their inclusion requires a behavioral revolution.

Examining Iraqi behavior, Lake finds two dimensions that negate a rational treatment of the case. First, he concludes in retrospect that Saddam strongly misperceived U.S. resolve to invade. Yet, the rationalist framework highlights how the United States had an incentive to misrepresent its resolve to invade in order to achieve its preferred outcome peacefully. If it was clear to both sides that Saddam would back down if confronted with a U.N. security resolution or with a mere troop mobilization, then a (counterfactual) low-resolve U.S.

administration would have had an incentive to bluff by sending signals of high resolve. This, however, undermined the credibility of U.S. signals of resolve. The fact that the U.S. administration was highly resolved to invade Iraq does not therefore mean that a rational Iraqi government should have believed it. Furthermore, Saddam possessed a – granted, ultimately flawed – strategy to placate U.S. aggression. He hoped first that Russia or France would intercede on Iraq's behalf to prevent a U.S. invasion and that, in case these efforts failed, Iraqi forces would be capable of increasing U.S. military costs to the point at which American public opinion would force Washington to back down.[79] Given Iraq's international isolation and U.S. power preponderance, such hopes were soon dashed. But this does not mean his behavior was irrational or that an archetypal rational state could not reach the same conclusion.

Second, Lake points out the role of multiple audiences in accounting for the seemingly puzzling behavior Saddam displayed throughout the nineties and in the run-up to the crisis, never revealing that the Iraqi nuclear program had been disabled by the sanctions regime despite rising U.S. pressure to prove this point. Lake is right that, as first UNSCOM and then UNMOVIC reports repeatedly emphasized, Iraq did not act like it had nothing to hide. Rather, it often engaged international inspectors in a non-cooperative fashion, heightening suspicion. After the war, the Iraq Survey Group concluded that Saddam, while trying to persuade the United States that it had no active WMD, was trying to create uncertainty in order to impress upon Iran and possible also Israel and his own people that he might have some WMD capabilities.[80] While multiple audiences were in play, our theory shows that they were not necessary to produce war. A theory allowing for multiple audiences and multiple ambiguous signals could produce a richer account of Hussein's behavior. In

[79]Woods et al. (2006, 28-32).

[80]See: See: CIA (2004).

particular, understanding the United States' incentive to misrepresent its preferences, and gambling that Washington was not resolved, Hussein could have chosen one of the more ambiguous signals, in an attempt to shore up his credentials against other audiences and facilitate the eventual resumption of the WMD programs. Yet in our opinion a key friction that contributed to the occurrence of war was that the United States was uncertain about its ability to detect current and future attempts to nuclearize. This argument need not assume the existence of multiple audiences, nor does it warrant a behavioral revolution. Morever, our argument offers a more plausible account of the war's timing. After all, Saddam had been playing to multiple audiences for more than a decade before the war.[81]

Likewise, Lake finds two features in U.S. behavior that lie outside the rationalist framework. First, he argues that the Bush Administration's underestimation of the costs of war and post-war governance in Iraq played an important role in eliminating the bargaining range and producing war.[82] It is true that *ex ante* U.S. assessments of the cost of the Iraq War subsequently proved to have significantly underestimated it. It is also true that, in terms of our model, the lower the expected cost of war, the lower the anticipated effect of Iraqi nuclearization that could still justify war. But U.S. estimates of this effect, particularly given putative links between Saddam's regime and terrorist groups, were particularly dire, so that war would have been justified even if its estimated cost would have been higher than predicted at the time. In fact, war was attractive not because it was cheap, but because it

[81]More generally, in a way consistent with our theory, multiple audiences may generate additional opportunities for mistaken preventive wars, by providing incentives for leaders to act as if their states invested in additional military capabilities when in fact they did not. Our point is not that Lake is wrong in arguing that multiple audiences may have contributed to the Iraq War. Rather, we argue that they were not necessary to cause the war, and are therefore also unnecessary when accounting for it.

[82]Lake (2010/11, 16).

could lead to Saddam's ouster from power and thus offer an effective method of preventing his acquisition of nuclear weapons.

This brings us to the second supposedly irrational feature Lake points out in U.S. behavior in the Iraq case: the standard the Bush Administration demanded Iraq to satisfy in order to prove the absence of a WMD program was too stringent, indeed irrational. But what kind of evidence could Washington have "rationally" demanded in order to alleviate any concern about Saddam's intentions to nuclearize? WMD programs are relatively easy to dissimulate and future intentions are remarkably difficult to signal in a credible fashion. Indeed, the Iraq Survey Group took two years of unfettered access to Iraqi facilities to conclude that Saddam's WMD program had in fact been dismantled.[83] Furthermore, the C.I.A. concluded after the invasion that Saddam was intent on resuming his nuclear program.[84] Given intelligence reports estimating that Iraqi nuclearization could happen "within several months to a year," further delay could bring Iraq closer to nuclearization, leaving little time for the United States to react and strike preventively. How could inspections have proven the negative within a reasonable timeline?[85] Rationality does not impose restrictions on preferences, but simply on the expected behavior given particular preferences.

To conclude, by 2002-2003, the United States possessed an overwhelming power advantage over Iraq. Reeling from the September 11, 2001, terrorist attacks, the U.S. administration heightened its perception of the threat posed by Saddam's putative WMD capability. Unable to obtain definitive information about Iraq's decision to forfeit nuclear acquisition, and fearing a large and rapid adverse power shift, the United States launched a preventive war – a mistaken one, as it turned out, since Saddam had long ago given up his WMD programs.

[83]Harvey (2012, 28).

[84]CIA (2004).

[85]Rice (2011, 169).

# 5. Conclusion

This article introduced a theory of how power shifts may produce preventive war. Because militarization efforts can go undetected, some states may pursue such efforts in the hope of presenting their adversaries with a *fait accompli*. Aware of this possibility, other states may attack an adversary for fear of the latter's militarization even if they do not possess unambiguous evidence that such militarization is taking place. Thus, information problems play a crucial role in providing a rationalist explanation for war. Ever since Fearon's seminal article,[86] commitment problems are seen as a sufficient cause of war in the presence of large and rapid power shifts.[87] Thus far, however, the rationalist literature has assumed that shifts are produced exogenously, freely, or immediately, before another state can strike preventively. In our view, large and rapid power shifts typically result from a state's decision to militarize, which is best understood as a costly investment with delayed returns in military capability. Conceptualizing power shifts this way, we show that they do not in and of themselves lead to conflict. Only when informational problems are present do power shifts lead to war.

We illustrated the causal logic of our theory with an account of the U.S.-led invasion of Iraq in 2003. Iraq's inability to commit not to develop nuclear weapons, paired with the U.S.'s inability to verify the absence of a Iraqi nuclear program, played a key role in the decision to go to war. More broadly, our theory indicates that mistaken preventive wars are more likely under conditions of power preponderance, such as unipolar international systems. This

[86]Fearon (1995).

[87]Powell (2006).

qualifies generally accepted claims about the peacefulness of a unipolar world,[88] which have only recently been disputed.[89] Furthermore, our theory can help solve a long-standing contradiction in studies of preventive war.[90] Arguing about democracy's pacifying effects, Schweller claims that "only nondemocratic regimes wage preventive wars against rising opponents."[91] But he restricts this argument to "power shifts between states of roughly equal strength."[92] We show that a state – democratic or not – is more likely to launch a preventive war when the net effect of militarization is high – *i.e.*, the cost of preventive war is lower relative to the expected power shift it is meant to forestall. When the net effect of militarization is low, "appeasement" (*i.e.*, peaceful acceptance of another state's rise) becomes the rational policy choice.[93] Consistent with this evidence, democratic support for a preventive strike is greater if the target state is 'half as strong' rather than 'as strong as' the home country.[94]

Turning to current policy debates, our theory has implications for U.S. policy towards the Iranian nuclear program. Some scholars have argued in favor of a preventive strike. For Kroenig, a war is the 'least bad option,' because its cost is not as frightening as the consequences of a nuclear-armed Iran.[95] In other words, the net effect of Iranian

[88]Wohlforth (1999).

[89]Monteiro (2011/12).

[90]Powell (2006); Lemke (2003); Reiter (2006).

[91]Schweller (1992, 238).

[92]Schweller (1992, 248).

[93]Treisman (2004).

[94]Tomz and Weeks (2010).

[95]Kroenig (2012).

nuclearization would be high. Our analysis of the Iranian case differs in two points.[96] First, we believe Kroenig underestimates the cost of war. A preventive strike on Iran's nuclear program would be quite costly.[97] The Iranian program is extensive, with key installations close to population centers or buried underground. Furthermore, Iran possesses the ability to inflict damage on U.S. interests in the region, including disrupting the flow of oil through the Strait of Hormuz;[98] firing missiles at, and encouraging Hezbollah attacks on, Israel; and undermining U.S. goals in Iraq and Afghanistan. Second, we think Kroenig overestimates the effectiveness of a preventive strike. Short of an all-out invasion – which most proponents of a strike rightly rule out on the grounds that it would be too costly to be rational – a preventive strike will leave Tehran's regime in place. As a civilian, oil-rich, parliamentary dictatorship, the Iranian regime is relatively stable[99] and unlikely to be unseated by a preventive strike. Even if it were toppled, the nuclear program enjoys widespread support in Iran and could be resumed quickly. This means that a preventive strike would guarantee Iranian non-nuclearization for a relatively short period of time.[100] Therefore, in our view, Iranian nuclearization does not configure a shift of such magnitude that would justify a U.S. preventive strike, given its limited effectiveness.

Generalizing, our theory offers implications for the likelihood of preventive attacks against nuclear programs. In our view, the cost of a preventive war, relative to the effect of nuclearization, increases with the size of the target set, the proximity of such targets to population centers, and the relative capabilities of the target. In some cases, this cost may be

[96]Debs and Monteiro (2012a).

[97]Kahl (2012).

[98]Talmadge (2008).

[99]Debs (2011).

[100]For surveys of preventive strikes, see Fuhrmann and Kreps (2010); Gavin and Rapp-Hooper (2011).

so high that preventive war is not a rational option and proliferation occurs unimpeded. For example, we believe that the ability of the Soviet Union to end the U.S. nuclear monopoly in 1949 resulted from the U.S.'s inability to launch a surgical counter-proliferation preventive strike. Given the absence of intelligence that enabled the construction of a target set for the Soviet nuclear program, a preventive strike needed to cripple the Soviet Union. But since the U.S. nuclear arsenal was far from possessing this capability, any strike in practice meant a massive conventional attack. But given the conventional balance of power, such an attack was expected to lead to significant Soviet territorial gains. A preventive strike was simply too costly to be a viable option.[101]

Looking ahead, we believe that the strategic framework presented in this article is a useful building block in a strategic theory of nuclear proliferation. Our framework may help bridge the gap between existing theories that focus on those states considering nuclear acquisition (the 'demand side') and those that focus on states interested in preventing proliferation (the 'supply side'). One important next step in constructing this theory would be to supplement the argument laid out in this article with an analysis of the strategic interaction between a potential proliferator and a nuclear ally, which would allow a weaker state to deter preventive military action while developing its own nuclear capability. We reserve this question for future research.[102]

[101] Buhite and Hamel (1990, 382-383), and Williamson Jr. and Rearden (1993, 140-141).

[102] Debs and Monteiro (2012b).

# References


Ayson, Robert and Brendan Taylor. 2004. "Attacking North Korea: Why War Might be Preferred." Comparative Strategy 23(3):263-279.

Baliga, Sandeep and Tomas Sjostrom. 2008. "Strategic Ambiguity and Arms Proliferation." Journal of Political Economy 116(6):1023-1057.

Bas, Muhamet A. and Andrew J. Coe. 2012. "Arms Diffusion and War." Journal of Conflict Resolution 56(4):651-674.

Biddle, Stephen, James Embrey, Edward Filiberti, Stephen Kidder, Steven Metz, Ivan C. Oerlich and Richard Shelton. 2004. Toppling Saddam: Iraq and American Military Transformation. Carlisle Barracks, PA: US Army War College, Strategic Studies Institute.

Blix, Hans. 2003a. "Briefing of the Security Council: Oral introduction of the 12th Quarterly Report of UNMOVIC. March 7, 2003." presentation to the UNSC on Feb. 14, 2003.

Blix, Hans. 2003b. "An Update on Inspection." presentation to the UNSC on Jan. 27, 2003.

Bueno de Mesquita, Bruce. 1996. Beliefs About Power and the Risks of War: A Power Transition Game. In Parity and War, ed. Jacek Kugler and Douglas Lemke. Ann Arbor, MI: University of Michigan Press pp. 271-286.

Buhite, Russell and Wm. Christopher Hamel. 1990. "War for Peace: The Question of an American Preventive War Against the Soviet Union, 1945-1955." Diplomatic History 14(3):367-384.

Bush, George W. 2002. "State of the Union Address. Jan. 29, 2002." http://georgewbushwhitehouse.archives.gov/news/releases/2002/01/20020129-11.htm

l (Accessed Apr. 9, 2010) .

Chadefaux, Thomas. 2011. “Bargaining Over Power: When Do Rapid Shifts in Power Lead to War?” International Theory 3(2):228-253.

Chapman, John. 2002. “Iraq's Weapons of Mass Destruction: A Net Assessment.” IISS Strategic Dossier.

CIA. 2004. “Comprehensive Report of the Special Advisor to the DCI on Iraq's Weapons of Mass Destruction.”

CNN. 2003. “Bush, Blair: Time running out for Saddam.” cnn.com/2003/US/01/31/sprj.irq.bush.blair.topics/. Jan. 31, 2003 (Accessed Apr. 9, 2010).

Coe, Andrew. 2011. “Costly Peace: A New Rationalist Explanation for War.” Harvard University Mimeo.

Copeland, Dale C. 2000. The Origins of Major War. Ithaca, NY: Cornell University Press.

Craig, Campbell and Sergey Radchenko. 2008. The Atomic Bomb and the Origins of the Cold War. New Haven, CT: Yale University Press.

Debs, Alexandre. 2011. “Living by the Sword and Dying by the Sword? Leadership Transitions in and Out of Dictatorships.” Yale University Mimeo.

Debs, Alexandre and Nuno P. Monteiro. 2012a. “The Flawed Logic of Striking Iran: Instead, Handle it Like North Korea.”

Debs, Alexandre and Nuno P. Monteiro. 2012b. “The Strategic Logic of Nuclear Proliferation.” Yale University Mimeo.

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

Fearon, James D. 1996. “Bargaining over Objects that Influence Future Bargaining Power.” University of Chicago Mimeo.

Fearon, James D. 2010. “Arming and Arms Races.” Paper prepared for presentation at the Annual Meetings of the American Political Science Association, Washington, D.C.

Freedman, Lawrence. 2004. “War in Iraq: Selling the Threat.” Survival 46(2):7-49.

Fudenberg, Drew and Jean Tirole. 1991. Game Theory. Cambridge, MA: MIT Press.

Fuhrmann, Matthew and Sarah E. Kreps. 2010. “Targeting Nuclear Programs in War and Peace: A Quantitative Empirical Analysis, 1941-2000.” Journal of Conflict Resolution 54(6):831-859.

Gartzke, Erik. 1999. “War is in the Error Term.” International Organization 53(3):567-587.

Gavin, Francis J. and Mira Rapp-Hooper. 2011. “The Copenhagen Temptation: Rethinking Prevention and Proliferation in the Age of Deterrence Dominance.” Paper prepared for presentation at the annual meeting of the American Political Science Association, Seattle, September.

Gilpin, Robert. 1981. War & Change in World Politics. Cambridge, U.K.: Cambridge University Press.

Gordin, Michael. 2009. Red Cloud at Dawn: Truman, Stalin, and the End of the Atomic Monopoly. New York: Farrar, Straus, and Giroux.

Gordon, Michael R. and Bernard E. Trainor. 1995. The Generals' War: The Inside Story of the Conflict in the Gulf. Boston: Little, Brown.

Harsanyi, John c. 1973. “Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points.” International Journal of Game Theory 2:1-23.

Harvey, Frank P. 2012. “President Al Gore and the 2003 Iraq War: A Counterfactual Test of Conventional “W”isdom.” Canadian Journal of Political Science 45(1):1-32.

Holloway, David. 1994. Stalin and the Bomb: The Soviet Union and Atomic Energy, 1939-1956. New Haven: Yale University Press.

House, White. 2003. “Press Briefing with Ari Fleisher.”

<http://georgewbushwhitehouse.archives.gov/news/releases/2003/04/20030410-6.html>, April 10, 2003 (Accessed April 9, 2010).

Jackson, Matthew O. and Massimo Morelli. 2009. "Strategic Militarization, Deterrence, and Wars." Quarterly Journal of Political Science 4(4):279-313.

Jervis, Robert. 2010. Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War. Ithaca, NY: Cornell University Press.

Kahl, Colin H. 2012. "Not Time to Attack Iran. Why War Should Be a Last Resort."

Kay, David. 2004. "Iraq's Weapons of Mass Destruction." Miller center Report 20.

Kerry, John. 2002. "Congressional Record. 10/9/02."

Kim, Woosang and James D. Morrow. 1992. "When Do Power Shifts Lead to War?" American Journal of Political Science 36(4):896-922.

Kim, Youngsam. 2001. President Kim Youngsam Memoir. Seoul: Chosun Ilbosa.

Kroenig, Matthew. 2012. "Time to Attack Iran. Why a Strike is the Least Bad Option." Foreign Affairs 91(1):76-86.

Kugler, Jacek and Douglas Lemke. 1996. Parity and War. Ann Arbor, MI: University of Michigan Press.

Lake, David A. 2010/11. "Two Cheers for Bargaining Theory Assessing Rationalist Explanations of the Iraq War." International Security 35(3):7-52.

Lake, David A. and Michael K. McKoy. 2011/12. "Correspondence: Bargaining Theory and Rationalist Explanations for the Iraq War." International Security 36(3):172-178.

Lee, Jung-Hoon and Chung-in Moon. 2003. "The North Korean Nuclear Crisis." Security Dialogue 34(2):135-152.

Lemke, Douglas. 2003. "Investigating the Preventive Motive for War." International Interactions 29(4):273-292.

Levy, Jack S. 2008. "Preventive War and Democratic Politics." International Studies

Quarterly 52(1):1-24.

Meirowitz, Adam and Anne E. Sartori. 2008. “Strategic Uncertainty as a Cause of War.” Quarterly Journal of Political Science 3(4):327-352.

Meirowitz, Adam and Kristopher W. Ramsay. 2010. “Investment and Bargaining.” Princeton University Mimeo.

Monteiro, Nuno P. 2011/12. “Unrest Assured: Why Unipolarity Is Not Peaceful.” International Security 36(3):9-40.

Organski, A.F.K. 1958. World Politics. New York, N.Y.: Knopf.

Organski, A.F.K. and Jacek Kugler. 1980. The War Ledger. Chicago, IL: University of Chicago Press.

Pape, Robert A. 1996. Bombing to Win: Air Power and Coercion in War. Ithaca, N.Y.: Cornell University Press.

Perry, William J. 2002. The United States and the Future of East Asian Security: Korea Quo Vadis? In Building Common Peace and Prosperity in Northeast Asia, ed. Keunmin Woo. Seoul: Yonsei University Press.

Perry, William J. October 12, 1999. Review of United States Policy Toward North Korea. Washington D.C.: U.S. Congress.

Pollack, Jonathan D. and Mitchell B. Reiss. 2004. South Korea: The Tyranny of Geography and the Vexations of History. In The Nuclear Tipping Point: Why States Reconsider Their Nuclear Choices, ed. Einhorn Robert J. Campbell, Kurt M. and Mitchell B. Reiss. Washington D.C.: Brookings Institution Press pp. 254-292.

Posen, Barry R. 1997. “U.S. Security Policy in a Nuclear-Armed World or: What if Iraq had had nuclear weapons?” Security Studies 6(3):1-31.

Powell, Robert. 1993. “Guns, Butter, and Anarchy.” American Political Science Review 87(1):115-132.

Powell, Robert. 1999. In the Shadow of Power: States and Strategies in International Politics. Princeton, New Jersey: Princeton University Press.

Powell, Robert. 2004. "The Inefficient Use of Power: Costly Conflict with Complete Information." American Political Science Review 98(2):231-241.

Powell, Robert. 2006. "War as a Commitment Problem." International Organization 60(1):169-203.

Powell, Robert. 2011a. "Monopolizing Violence and Consolidating Power." UC Berkeley Mimeo.

Powell, Robert. 2011b. "Persistent Fighting and Shifting Power." UC Berkeley Mimeo.

Reiter, Dan. 2006. Preventive Attacks against Nuclear, Biological, and ChemicalWeapons Programs: The Track Record. In Hitting First: Preventive Force in U.S. Security Strategy, ed. William Keller and Gordon Mitchell. Pittsburgh: University of Pittsburgh Press pp. 27-44.

Rhodes, Richard. 1995. Dark Sun: The Making of the Hydrogen Bomb. New York: Simon & Schuster.

Rhodes, Richard. 2010. The Twilight of the Bombs: Recent Challenges, New Dangers, and the Prospects for a World without Nuclear Weapons. New York, N.Y.: Alfred Knopf.

Rice, Condoleezza. 2011. No Higher Honor. A Memoir of My Years in Washington. New York, N.Y.: Random House.

Ritter, Scott. 1998. "Interview with Elizabeth Farnsword." PBS NewsHour. August 31, 1998 .

Rovner, Joshua. 2010. Fixing the Facts: National Security and the Politics of Intelligence. Ithaca, NY: Cornell University Press.

Schultz, Kenneth A. 2010. "The Enforcement Problem in Coercive Bargaining: Interstate Conflict over Rebel Support in Civil Wars." International Organization

64(2):281-312.

Schweller, Randall L. 1992. “Domestic Structure and Preventive War: Are Democracies More Pacific?” World Politics 44(2):235-269.

Singer, J. David. 1987. “Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985.” International Interactions 14(2):115-132.

Slantchev, Branislav L. 2005. “Military Coercion in Interstate Crises.” The American Political Science Review 99(4):533-547.

Slantchev, Branislav L. 2011. Military Threats: The Costs of Coercion and the Price of Peace. New York, N.Y.: Cambridge University Press.

Suskind, Ron. 2006. The One Percent Doctrine: Deep Inside America's Pursuit of its Enemies since 9/11. New York, N.Y.: Simon and Schuster.

Talmadge, Caitlin. 2008. “Closing Time: Assessing the Iranian Threat to the Strait of Hormuz.” International Security 33(1):82-117.

Tenet, George. 2007. At the Center of the Storm: My Years at the CIA. New York: HarperCollins. with Bill Harlow.

Tessman, Brock and Steve Chan. 2009. Relative Decline: Why Does it War or Sustain Peace? In Systemic Transitions: Past, Present, and Future, ed. William R. Thompson. New York, N.Y.: Palgrave Macmillan pp. 9-30.

Thompson, Alexander. 2009. Channels of Power: The U.N. Security Council and U.S. Statecraft in Iraq. Ithaca, N.Y.: Cornell University Press.

Thucydides. 1998. History of the Peloponnesian War. Indianapolis, IN: Hackett. Translated by Steven Lattimore.

Tomz, Michael and Jessica L. Weeks. 2010. “An Experimental Investigation of the Democratic Peace.” Paper presented at the Annual Meeting of the American Political Science Association. Washington, D.C.

Treisman, Daniel. 2004. “Rational Appeasement.” International Organization 58(2):345-373.

UN. 2003. “Press Release SC/7658.” http://www.un.org/News/Press/docs/2003/sc7658.doc.htm. Feb. 2, 2003 (Accessed Apr. 9, 2010).

Vasquez, John A. 2009. Whether and How Global Leadership Transitions Will Result in War: Some Long-Term Predictions From the Steps to War - Explanation. In Systemic Transitions: Past, Present, and Future, ed. William R. Thompson. New York, N.Y.: Palgrave Macmillan pp. 131-160.

Werner, Suzanne and Jacek Kugler. 1996. Power Transitions and Military Buildups: Resolving the Relationship Between Arms Buildups and War. In Parity and War: Evaluations and Extensions of The War Ledger, ed. Jacek Kugler and Douglas Lemke. Ann Arbor, MI: University of Michigan pp. 187-208.

Williamson Jr., Samuel R. and Steven L. Rearden. 1993. The Origins of U.S. Nuclear Strategy, 1945-1953. New York: St. Martin's Press.

Wohlforth, William C. 1999. “The Stability of a Unipolar World.” International Security 24(1):5-41. 59

Wolfowitz, Paul. 2003. “Interview with Sam Tannenhaus.” Vanity Fair. http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2594. May 9, 2003 (Accessed Apr. 9, 2010) .

Woods, Kevin M., Michael R. Pease, Mark E. Stout, Williamson Murray and James G. Lacey. 2006. Iraqi Perspectives Report: A View of Operation Iraqi Freedom from Saddam's Senior Leadership. Washington, D.C.: Joint Center for Operational Analysis, United States Joint Forces Command.

# 6. Appendix

The Appendix is structured as follows. Section 1 offers proofs of the formal results in the two-period game discussed in the main text. Section 2 offers an expanded presentation of the infinite-horizon game set-up and results. Section 3 offers proofs of the formal results of the infinite-horizon game.[103]

## *6.1. Two-Period Game: Proofs*

We say that the effect of militarization is smaller than the cost of investment if

$$\delta[w_T(1) - w_T(0)] \leq k \tag{3}$$

We say that the effect of militarization is smaller than the cost of preventive war if

$$\delta[w_T(1) - w_T(0)] \leq 1 - w_T(0) - w_D(0) \tag{4}$$

We can show:

**Proof.** (Proof of Proposition 2).

*(i) If (3) holds strictly or if (4) holds strictly, then peace prevails.*

*(ii) If (3) holds, the PBE is:* $I_1^* = 0$. *D offers* $z_1^* = w_T(0)$ *unless* $s_1 = 1$ *and (4) fails, in which case* $D$ *declares war.* $T$ *accepts any* $z_1 \geq w_T(0)$.

*(iii) If (3) fails and (4) holds, the PBE is:* $I_1^* = 1$. *D offers* $z_1^* = w_T(0)$ *after any* $s_1$. $T$ *accepts any* $z_1 \geq w_T(0)$.

First, it is clear that in any equilibrium $T$ accepts any offer $z_1 \geq w_T(0)$.

Consider $D$'s choice between declaring war and offering $z_1 = w_T(0)$. The expected utility

[103] In non-generic regions of the parameter space, such as (3) and (4) below holding with equality, there could be multiple equilibria. In these cases we assume that any player breaks indifference in favor of the efficient action, investing for $T$ and making a peaceful offer for $D$.

for $D$ of a preventive war is $w_D(0) + \delta\big(1 - w_T(0)\big)$. The expected utility for $D$ of offering $z_1 = w_T(0)$ is at least equal to $1 - w_T(0) + \delta\big(1 - w_T(1)\big)$. Thus, if (4) holds strictly, $D$ strictly prefers to offer $z_1 = w_T(0)$ and peace prevails.

Consider $T$'s decision to invest in military capabilities. The expected utility for $T$ of investing in military capabilities is at most equal to $-k + w_T(0) + \delta w_T(1)$. The expected utility for $T$ of not investing in military capabilities is equal to $w_T(0) + \delta w_T(0)$. Thus, if (3) holds strictly, $T$ strictly prefers not to invest in military capabilities and as a result $D$ strictly prefers to offer $z_1 = w_T(0)$, since war is inefficient $(1 - w_T(0) + \delta\big(1 - w_T(0)\big) > w_D(0) + \delta\big(1 - w_T(0)\big))$. This completes the proof of (i).

Now assume that (3) holds. If it holds strictly, then $T$ strictly prefers $I_1 = 0$. In the non-generic case where (3) holds with equality, there may be multiple equilibria, yet by assumption $T$ breaks indifference in non-generic regions of the parameter space in favor of the efficient action, i.e. not investing. Thus, if (3) holds, $T$ chooses $I_1^* = 0$. As a result, $D$ prefers to declare war if and only if $w_D(0) + \delta\big(1 - w_T(0)\big) > 1 - w_T(0) + \delta\big(1 - w_T(s_1)\big)$ or if and only if $s_1 = 1$ and (4) fails (where we use the fact that in the non-generic case where (4) holds with equality, $D$ breaks indifference in favor of the efficient action, i.e. offering $z_1 = w_T(0)$). This completes the proof of (ii).

Now assume that (3) fails and (4) holds. If (4) holds strictly, then $D$ strictly prefers to offer $z_1 = w_T(0)$. In the non-generic case where (4) holds with equality, there may be multiple equilibria, yet by assumption $D$ breaks indifference in non-generic regions of the parameter space in favor of the efficient action, i.e. offering $z_1 = w_T(0)$. Thus, if (4) holds, $D$ offers $z_1^* = w_T(0)$. As a result, $T$ chooses $I_1^* = 1$ since (3) fails. This completes the proof of (iii).

**Proof.** (Proof of proposition 3). *Assume that (3) and (4) fail.*

*(i) If (2) holds, the PBE is:* $I_1^* = 0$. *D offers* $z_1^* = w_T(0)$ *after* $s_1 = 0$ *and declares war*

*after* $s_1 = 1$. $T$ *accepts any* $z_1 \geq w_T(0)$.

*(ii) If (2) fails, the PBE is:* $T$ *invests with probability* $q^* = \frac{1}{p_s + (1-p_s)\frac{\delta[w_T(1) - w_T(0)]}{1 - w_T(0) - w_D(0)}}$. *After* $s_1 = 0$, $D$ *offers* $z_1^* = w_T(0)$ *with probability* $r^* = \frac{k}{(1-p_s)\delta[w_T(1) - w_T(0)]}$ *and declares war with probability* $1 - r^*$. *After* $s_1 = 1$, *D declares war.* $T$ *accepts* $z_1 \geq w_T(0)$.

First, it is clear that $T$ accepts any $z_1 \geq w_T(0)$. Also, it is clear that $D$ declares war after $s_1 = 1$ since (4) fails.

After signal $s_1 = 0$, $D$ must choose between the best response to either pure strategy by $T$. Let $D$ offer $z_1^* = w_T(0)$ with probability $r^*$ and declare war with probability $1 - r^*$ after signal $s_1 = 0$. Let $T$ invest with probability $q^*$.

Next we show that $q^* < 1$. Indeed, if $q^* = 1$, then $D$ prefers $r^* = 0$, so that $T$ prefers $q^* = 0$, a contradiction. Second, if $q^* = 0$, then $D$ prefers $r^* = 1$, so that $T$ has no incentive to deviate if and only if

$$(1 + \delta)w_T(0) \geq -k + (1 + \delta)w_T(0) + (1 - p_s)\delta\big(w_T(1) - w_T(0)\big)$$

or (2) holds. If (2) fails, we must have $q^* \in (0,1)$. This in turn requires that $r^* \in (0,1)$. $q^* \in (0,1)$, $r^* \in (0,1)$ are given by the indifference conditions:

$$(1 + \delta)w_T(0) = -k + (1 + \delta)w_T(0) + r^*(1 - p_s)\delta\big(w_T(1) - w_T(0)\big)$$

**Proof.** (Proof of corollary 1). $(i)$ War occurs with positive probability if and only if (2) and (4) fail. Clearly, (2) is less stringent as $p_s$ increases.

$(ii)$ If (2) and (4) fail, the probability of preventive war is

$$1 - (1 - q^* p_s)r^* = 1 - \frac{\frac{k}{1 - w_T(0) - w_D(0)}}{p_s + (1 - p_s)\frac{\delta[w_T(1) - w_T(0)]}{1 - w_T(0) - w_D(0)}}$$

which is decreasing in $p_s$.

$(iii)$ If (2) and (4) fail, the share of preventive wars that are mistaken is

$$1-\frac{\big(1-r^*(1-p_s)\big)q^*}{1-(1-q^*p_s)r^*}=1-\frac{1-\dfrac{k}{\delta[w_T(1)-w_T(0)]}}{p_s+(1-p_s)\dfrac{\delta[w_T(1)-w_T(0ht)]}{1-w_T(0)-w_D(0)}-\dfrac{k}{1-w_T(0)-w_D(0)}}$$

which is decreasing in $p_s$.

## *6.2. Infinite-Horizon Game: Set-Up and Results*

Consider the following extension of the baseline model. The game starts as in period 1 above (section 3.2.1.) and each subsequent period repeats the same timing, as long as the target does not invest in military capabilities and peace prevails. If $T$ invests at $t$, then it obtains additional military capabilities at $t+1$ only if $D$ does not strike preventively. Once $T$ obtains additional military capabilities, it conserves them and each period from then on proceeds as in period 2 above (section 3.2.1.). If $T$ has its baseline level of military capabilities and suffers a preventive war at $t$, it loses the opportunity to militarize for $N$ periods, *i.e.* the timing in periods $t+1$ to $t+N$ follows period 2 above (with $M_{t+s}=0$ $\forall 1\leq s\leq N$), and the timing in period $t+N+1$ follows period 1 above (and continues according to the same rules). We call $N$ the *effectiveness* of a preventive war. Finally, if $T$ rejects an offer from $D$ at $t$, then $t+1$ follows the same timing as $t$.

We analyze this game and impose that, at any information set, play is sequentially rational given beliefs, and beliefs are obtained using Bayes' rule whenever possible. There can be multiple equilibria in the infinite-horizon game. We first restrict attention to the set of *Markov Perfect Equilibria* (MPEs) of the game. An MPE requires that players play *Markovian* strategies, *i.e.,* strategies that they depend on history only through payoff-relevant state variables, here the military capabilities of the target ($M_t$), whether the target has the

option to militarize ($O_t$) and if it does not, the number of periods ($n_t$) of the effectiveness of the preventive war that have passed already.[104] An MPE is a vector of Markovian strategies that are mutual best-responses, beginning at any date $t$ for any value of the payoff-relevant state variables. For the full of solution of MPEs, see Proposition 7 in section 6.3.

We then consider the set of *Perfect Public Equilibria* (PPEs) of the game. A PPE requires that players play *public* strategies, *i.e.,* strategies that depend not just on the payoff-relevant state variables listed above, but on the full public history of the game, here the signals about $T$'s militarization decision, the offers from $D$ and the decisions by $T$ to accept or reject $D$'s offers. A PPE is a vector of public strategies that are mutual best-responses, beginning at any date $t$ for any public history.

We ask whether countries can sustain the efficient outcome in a PPE, where militarization does not occur and peace prevails, if any public history revealing a deviation triggers the MPE. For simplicity, we restrict attention to the set of efficient PPEs in *stationary* strategies, where along the equilibrium path, $T$ does not militarize, $D$ makes a fixed offer $z_t = z^*$ after $s_t = 0$, which $T$ accepts. If there is no efficient PPE in stationary strategies, countries play the MPE.

In parallel with the two-period game, we characterize the threshold values for the effect of militarization. We say that the effect of militarization is smaller than the cost of investment, averaged over all periods, if

$$\delta\big(w_T(1) - w_T(0)\big) \leq (1-\delta)k \tag{5}$$

We say that the signal is sufficiently informative, or that the effect of militarization is smaller than the cost of investment, averaged over all periods, assuming it goes undetected, if

[104] $O_t \in (0,1)$, where $O_t = 1$ if and only if the target has the option to militarize and $n_t \in \{0,\dots,N-1\}$, where we set $n_t = 0$ if $O_t = 1$.

$$(1-p_s)\delta[w_T(1)-w_T(0)] \leq (1-\delta)k \tag{6}$$

We say that the effect of militarization is smaller than the cost of preventive war, averaged over all periods, if

$$\delta\big(w_T(1)-w_T(0)\big) \leq (1-\delta)\big(1-w_T(0)-w_D(0)\big) \tag{7}$$

We say that the effect of militarization is smaller than the cost of preventive war, averaged over the effectiveness of a preventive war, if

$$\delta\big(w_T(1)-w_T(0)\big) \leq \frac{1-\delta}{1-\delta^{N+1}}\big(1-w_T(0)-w_D(0)\big) \tag{8}$$

First, we can show that an efficient MPE exists if militarization is not rationalizable or if it can be deterred. If the effect of militarization is smaller than the cost of investment, averaged over all periods, then militarization is not rationalizable: $D$ need not fear $T$'s militarization and peace prevails. If the effect of militarization is greater than the cost of investment, averaged over all periods, but the signal is sufficiently informative, then $D$ may use the threat of preventive war to deter militarization. Such a threat of preventive war is credible if and only if the effect of militarization is greater than the cost of preventive war, averaged over all periods. In either case, $D$ achieves its prefered outcome, peace without concessions ($z^* = w_T(0)$).

**Proposition 4.** *(i) An efficient MPE exists if either (i.1) the effect of militarization is smaller than the cost of investment, averaged over all periods ((5) holds), or (i.2) the effect of militarization is greater than the cost of investment, averaged over all periods ((5) fails), the signal is sufficiently informative ((6) holds), and the effect of militarization is greater than the cost of preventive war, averaged over all periods ((7) fails).*

*(ii) In such circumstances, an efficient PPE in stationary strategies always exists, where $D$ offers $z^* = w_T(0)$ after signal $s_t = 0$.*

***Proof.*** *See section 6.3.*

Second, we note that there is a unique MPE where peace prevails and militarization occurs if

an investment in military capabilities is rationalizable but a preventive war is not. Moreover, the same outcome is produced if the signal is not sufficiently informative, so that $T$ would be willing to militarize, hoping to avoid detection, and $D$ is not willing to launch a preventive war, given its limited effectiveness. In these cases, we note that efficiency may not necessarily be ensured in a PPE in stationary strategies. As $T$ becomes more patient, it demands greater - not smaller - concessions, which may be too high for $D$.[105] Formally:

**Proposition 5.** *(i) There is a unique MPE where peace prevails and $T$ militarizes if either (i.1) the effect of militarization is greater than the cost of investment, averaged over all periods ((5) fails), and smaller than the cost of preventive war, averaged over all periods ((7) holds), or (i.2) the signal is not sufficiently informative ((6) fails), the effect of militarization is greater than the cost of preventive war, averaged over all periods ((7) fails), but smaller than the cost of preventive war, averaged over the effectiveness of a preventive war ((8) holds).*

*(ii) In such circumstances, an efficient PPE in stationary strategies, where $D$ offers $z^* > w_T(0)$ after signal $s_t = 0$, exists if and only if*

$$[1 - \delta(p_s + (1 - p_s)\delta)]\delta\big(w_T(1) - w_T(0)\big) \leq (1 - \delta)k \tag{9}$$

*This condition may fail even if there is no restriction on how patient countries are.*

***Proof.*** *See section 6.3.*

Finally, if the signal is not sufficiently informative, so that $T$ would be willing to militarize, hoping to avoid detection, and $D$ is willing to launch a preventive war, despite its limited effectiveness, then strategic uncertainty remains and war occurs with positive probability in the MPE.

[105] This result hinges on the fact that the effectiveness of preventive wars is finite ($N < \infty$). If $D$ was unwilling to launch an infinitely effective preventive war, it would be willing to accede to $T$'s demand as it becomes more patient.

We can hope that efficiency, and peace, can be sustained in a stationary PPE. In order to be dissuaded from militarization, $T$ must receive concessions ($z^* > w_T(0)$). Yet concessions make preventive war attractive, as it ensures $T$'s demilitarization without any concession. An efficient PPE in stationary strategies exists if the cost of peace is smaller than the cost of war, *i.e.,* if the concessions needed to prevent militarization are smaller than the cost of a preventive war, averaged over the effectiveness of the preventive war (condition (12), below, holds).[106] This condition may fail even if countries are very patient, as explained in the text. Formally:

**Proposition 6.** *(i) There is a unique MPE where war happens with positive probability if the signal is not sufficiently informative ((6) fails) and the effect of militarization is greater than the cost of preventive war, averaged over the effectiveness of a preventive war ((8) fails). In this MPE, $T$ militarizes with probability*

$$q^* = \frac{1}{1 + \frac{1-p_s}{1-\delta}\left[\frac{\delta\left(w_T(1) - w_T(0)\right)}{\frac{1-\delta}{1-\delta^{N+1}}\left(1 - w_T(0) - w_D(0)\right)} - 1\right]} \tag{10}$$

*After a signal $s_t = 1$, preventive war happens with probability one. After a signal $s_t = 0$, peace prevails with probability*

$$r^* = \frac{(1-\delta)k}{(1-p_s)\delta\left(w_T(1) - w_T(0)\right)} \tag{11}$$

*and preventive war happens with probability $1 - r^*$.*

*(ii) In such circumstances, an efficient PPE in stationary strategies, where $D$ offers $z^* > w_T(0)$ after signal $s_t = 0$, exists if and only if*

[106] Coe (2011). The cost of war is discounted by $\delta$ since a deviation at $t$ triggers war only at $t + 1$.

$$\frac{(1-p_s)\delta\big(w_T(1)-w_T(0)\big)-(1-\delta)k}{p_s+(1-p_s)\delta} \leq \delta\frac{1-\delta}{1-\delta^{N+1}}\big(1-w_T(0)-w_D(0)\big) \tag{12}$$

*This condition may fail even if there is no restriction on how patient countries are.*

***Proof.*** *See section 6.3.*

We now investigate the role of information problems and the effectiveness of preventive war on the likelihood of conflict. We conclude:

**Corollary 3.** *The greater is the informativeness of the signal*

*(i) The more stringent become the conditions under which preventive wars happen with positive probability, and if such conditions are met,*

*(ii) The smaller is the probability of preventive war;*

*(iii) The smaller is the share of preventive wars that are mistaken.*

***Proof.*** *See section 6.3.*

Next, we investigate the effect of the effectiveness of preventive war on the likelihood of conflict, summarized in corollary 2 in the main text (see the proof in section 6.3).

## *6.3. Infinite-Horizon Game: Proofs*

**Proposition 7.** *(A) In a period t where T does not have the option to militarize, there is a unique MPE: D offers* $z_t^* = w_T(M_t)$ *and T accepts* $z_t \geq w_T(M_t)$.

*(B) Consider a period t where T has the option to militarize.*

*(B.i) In an efficient MPE: T chooses* $q^* = 0$. *After* $s_t = 0$, *D offers* $z^* = w_T(0)$. *After* $s_t = 1$, *D offers* $z^* = w_T(0)$ *if (7) holds and declares war otherwise. T accepts* $z_t \geq w_T(0)$.

*An efficient MPE exists if and only if either*

*(B.i.1) (7) and (5) hold,*

*(B.i.2) (7) fails and (6) holds.*

*(B.ii) In an inefficient MPE where peace prevails: T chooses* $q^* = 1$. *After any* $s_t$, *D offers* $z^* = w_T(0)$. *T accepts* $z_t \geq w_T(0)$. *This MPE exists if (5) fails, (8) holds.*

*(B.iii) In an inefficient MPE where war happens with positive probability, T chooses* $q^*$ *given by (10). After* $s_t = 1$, *D declares war. After* $s_t = 0$, *D offers* $z_t^* = w_T(0)$ *with probability* $r^*$ *given by (11) and declares war with probability* $1 - r^*$. *T accepts* $z_t \geq w_T(0)$. *This MPE exists if neither (6) nor (8) hold.*

***Proof.*** *First, T accepts any* $z_t \geq w_T(M_t)$, *since strategies are not history-dependent.*

*Now consider (A). If* $M_t = 1$, *D offers* $z^* = w_T(1)$, *since war is inefficient and* $1 - w_T(1)$ *is from then on D's maximum per-period payoff. Likewise, if* $M_t = 0$, *D offers* $z^* = w_T(0)$, *since war is inefficient and* $1 - w_T(0)$ *is D's maximum per-period payoff.*

*Now consider (B). After any signal, D chooses between declaring war and offering* $z_t = w_T(0)$. *After* $s_t = 1$, *D declares war if and only if*

$$w_D(0) + \delta(1-\delta^N)\frac{1-w_T(0)}{1-\delta} + \delta^{N+1}V_D^{MPE}(0,1,0) > 1 - w_T(0) + \delta\frac{1-w_T(1)}{1-\delta} \tag{13}$$

$$\Leftrightarrow \delta\big(w_T(1) - w_T(0)\big) - (1-\delta)\big(1 - w_T(0) - w_D(0)\big) > \delta^{N+1}[1 - w_T(0) - (1-\delta)V_D^{MPE}(0,1,0)] \tag{14}$$

*where* $V_i^{MPE}(M_t, O_t, n_t)$ *is country i's continuation value in the MPE.*

*After* $s_t = 0$, *D declares war if and only if*

$$w_D(0) + \delta(1-\delta^N)\frac{1-w_T(0)}{1-\delta} + \delta^{N+1}V_D^{MPE}(0,1,0) > 1 - w_T(0) + \delta\frac{q^*(1-p_s)}{1-q^*p_s}\frac{1-w_T(1)}{1-\delta} + \delta\frac{1-q^*}{1-q^*p_s}V_D^{MPE}(0,1,0) \tag{15}$$

$$\Leftrightarrow \delta\big(w_T(1)-w_T(0)\big)-(1-\delta)\big(1-w_T(0)-w_D(0)\big)>$$

$$\delta^{N+1}[1-w_T(0)-(1-\delta)V_D^{MPE}(0,1,0)]$$
$$+\delta\frac{1-q^*}{1-q^*p_s}\Big((1-\delta)V_D^{MPE}(0,1,0)-\big(1-w_T(1)\big)\Big) \tag{16}$$

*Clearly, if (16) holds, then so does (14) (since* $V_D^{MPE}(0,1,0)\geq\frac{1-w_T(1)}{1-\delta}$*). Thus either (i)* $D$ *offers* $z^*=w_T(0)$ *after any* $s_t$*, or (ii)* $D$ *declares war after* $s_t=1$ *and, after* $s_t=0$*,* $D$ *offers* $z_t=w_T(0)$ *with probability* $r^*\in[0,1]$ *and declares war with probability* $1-r^*$*. In (i),* $T$ *prefers* $I_t=1$ *if* $-k+w_T(0)+\delta\frac{w_T(1)}{1-\delta}>w_T(0)+\delta V_T^{MPE}(0,1,0)$*, or*

$$k<\delta\left[\frac{w_T(1)}{1-\delta}-V_T^{MPE}(0,1,0)\right] \tag{17}$$

*In (ii),* $T$ *prefers* $I_t=1$ *if*

$$-k+w_T(0)+\delta(1-(1-p_s)r^*)V_T^{MPE}(0,0,0)+\delta(1-p_s)r^*\frac{w_T(1)}{1-\delta}>$$

$$\Leftrightarrow k<\delta r^*\left[(1-p_s)\left(\frac{w_T(1)}{1-\delta}-V_T^{MPE}(0,1,0)\right)-p_s(1-\delta^N)\left(V_T^{MPE}(0,1,0)-\frac{w_T(0)}{1-\delta}\right)\right] \tag{18}$$

*using the fact that the value of the game for* $T$ *after a preventive strike at* $t$ *is*

$$V_T^{MPE}(0,0,0)=\frac{1-\delta^N}{1-\delta}w_T(0)+\delta^N V_T^{MPE}(0,1,0) \tag{19}$$

*Let us characterize the conditions under which an efficient MPE exists. In an efficient MPE,* $V_D^{MPE}(0,1,0)=\frac{1-w_T(0)}{1-\delta}$*,* $V_T^{MPE}(0,1,0)=\frac{w_T(0)}{1-\delta}$*. From (16),* $D$ *offers* $z^*=w_T(0)$ *after* $s_t=0$ *since war is inefficient. From (14),* $D$ *offers* $z^*=w_T(0)$ *after* $s_t=1$ *if (7) holds and declares war otherwise. From (17) and (18),* $T$ *prefers not to invest if and only if either (i) (7) and (5) hold or (ii) (7) fails and (6) holds.*

*Now assume that the efficient equilibrium does not exist and peace prevails. Thus* $q^*>0$

*and* $D$ *offers* $z^* = w_T(0)$ *after any* $s_t$. *Generically, we cannot have* $q^* \in (0,1)$.[107] *Therefore,*

$$V_D^{MPE}(0,1,0) = 1 - w_T(0) + \frac{\delta}{1-\delta}\big(1 - w_T(1)\big) \tag{20}$$

$$V_T^{MPE}(0,1,0) = w_T(0) - k + \frac{\delta w_T(1)}{1-\delta} \tag{21}$$

*Using (14),* $D$ *offers* $z^* = w_T(0)$ *after any* $s_t$ *if and only if (8) holds. Using (17),* $T$ *chooses* $q^* = 1$ *if and only if holds (5) fails.*

*Now assume that the efficient equilibrium does not exist and war occurs with positive probability.* $D$ *must declare war after* $s_t = 1$ *and, after* $s_t = 0$, *offer* $z_t = w_T(0)$ *with probability* $r^* \in [0,1]$ *and declare war with probability* $1 - r^*$. *Now note that* $q^* > 0$. *Indeed, if* $q^* = 0$ *then* $r^* = 1$ *((15) fails, since war is inefficient), so that war does not occur, a contradiction. Second, note that* $r^* > 0$. *Indeed,* $r^* = 0$ *implies* $q^* = 0$ *((18) fails), which we just ruled out. Third, note that* $q^* < 1$. *Indeed if* $q^* = 1$, *then (14) and (16) are equivalent, so that* $r^* = 0$, *which we just ruled out. Next,* $q^* \in (0,1)$ *implies that* $T$ *is indifferent about investing, i.e.* $V_T^{MPE}(0,1,0) = w_T(0) + \delta(1 - r^*)V_T^{MPE}(0,0,0) + \delta r^* V_T^{MPE}(0,1,0)$, *or*

$$V_T^{MPE}(0,1,0) = \frac{w_T(0) - \delta(1 - r^*)\big(V_T^{MPE}(0,1,0) - V_T^{MPE}(0,0,0)\big)}{1-\delta} \tag{22}$$

*Rearranging (19), we get*

$$V_T^{MPE}(0,1,0) - V_T^{MPE}(0,0,0) = (1 - \delta^N)\left[V_T^{MPE}(0,1,0) - \frac{w_T(0)}{1-\delta}\right] \tag{23}$$

*(22) and (23) imply* $V_T^{MPE}(0,1,0) = V_T^{MPE}(0,0,0) = \frac{w_T(0)}{1-\delta}$. *Replacing in (18), which must*

[107] In non-generic regions of the parameter space, any player breaks indifference in favor of the efficient action. Thus, if $D$ plays a pure strategy after any signal and $T$ is indifferent about investing, then $q^* = 0$.

*hold with equality, we solve for* $r^*$ *and obtain (11). Generically, we cannot have* $r^* = 1$.[108] *Moreover,* $r^* < 1$ *if and only if (6) fails. Now since* $D$ *declares war after* $s_t = 1$*, and is indifferent about declaring war after* $s_t = 0$*, we get* $V_D^{MPE}(0,1,0) = w_D(0) + \delta(1 - \delta^N)\frac{1-w_T(0)}{1-\delta} + \delta^{N+1}V_D^{MPE}(0,1,0)$ *or*

$$V_D^{MPE}(0,1,0) = \frac{1}{1-\delta}\left[1 - w_T(0) - \frac{1-\delta}{1-\delta^{N+1}}\big(1 - w_T(0) - w_D(0)\big)\right] \tag{24}$$

*Replacing in (16), which must hold with equality, we solve for* $q^*$ *and obtain (10).* $q^* \in (0,1)$ *if and only if (8) fails.*

**Proof.** (Proof of Proposition 4) Part (i) follows from proposition 7. Part (ii) is straightforward since Markovian strategies are stationary and depend trivially on history.

**Proof.** (Proof of Proposition 5) Part (i) follows from proposition 7. For part (ii), recall that $V_D^{MPE}(0,1,0)$ and $V_T^{MPE}(0,1,0)$ are given by (20) and (21) respectively. Let us construct an efficient PPE in stationary strategies.

First, $T$ must accept $z^*$, *i.e.* $w_T(0) + \delta V_T^{MPE}(0,1,0) \leq \frac{z^*}{1-\delta}$, or

$$z^* \geq w_T(0) + \delta\big[\delta\big(w_T(1) - w_T(0)\big) - (1-\delta)k\big] \tag{25}$$

Next, after any history revealing a deviation, $T$ accepts any $z_t \geq w_T(0)$, since from $t+1$ countries play the MPE.

Moving up, $D$ offers $z_t = w_T(0)$ after $s_t = 1$ ((14) fails since (8) holds).

Next, $T$ refrains from militarization if and only if

$$-k + p_s w_T(0) + (1-p_s)z^* + \delta\frac{w_T(1)}{1-\delta} \leq \frac{z^*}{1-\delta} \tag{26}$$

[108] In non-generic regions of the parameter space, any player breaks indifference in favor of the efficient action. If $D$ plays a pure strategy after any signal and $T$ is indifferent about investing, then $q^* = 0$.

$$\Leftrightarrow z^* \geq w_T(0) + \frac{\delta\big(w_T(1) - w_T(0)\big) - (1-\delta)k}{p_s + (1-p_s)\delta} \tag{27}$$

which is a tighter condition than (25).

Also, the best deviation for $D$, after $s_t = 0$, is to offer $z_t = w_T(0)$ ((16) fails since (8) holds). $D$ does not offer $z_t = w_T(0)$ if and only if $1 - w_T(0) + \delta V_D^{MPE}(0,1,0) \leq \frac{1-z^*}{1-\delta}$, or

$$z^* \leq w_T(0) + \delta^2\big(w_T(1) - w_T(0)\big) \tag{28}$$

(27) and (28) hold if and only if (9) holds. (8) implies (9) if and only if

$$\frac{1-\delta}{1-\delta^{N+1}}\big(1 - w_T(0) - w_D(0)\big) \leq \frac{(1-\delta)k}{1-\delta(p_s + (1-p_s)\delta)} \tag{29}$$

Taking the limit as $\delta$ approaches 1, and using l'Hopital's rule, this becomes

$$\frac{1}{N+1}\big(1 - w_T(0) - w_D(0)\big) \leq \frac{k}{2-p_s} \tag{30}$$

which may fail. If $1 - w_T(0) - w_D(0) = (N+1)\frac{k}{1-p_s} = (N+2)\big(w_T(1) - w_T(0)\big)$, then $\exists \delta' \in (0,1)$ such that $\forall \delta \in (\delta', 1)$, (5) fails, (8) holds and yet (9) fails.

**Proof.** (Proof of Proposition 6). Part (i) follows from proposition 7. For part (ii), recall that $V_D^{MPE}(0,1,0)$ is given by (24) and $V_T^{MPE}(0,1,0) = \frac{w_T(0)}{1-\delta}$. Let us construct an efficient PPE in stationary strategies.

First, $T$ must accept $z^*$, *i.e.* $w_T(0) + \delta V_T^{MPE}(0,1,0) \leq \frac{z^*}{1-\delta}$, or $z^* \geq w_T(0)$. Next, after any history revealing a deviation, $T$ accepts any $z_t \geq w_T(0)$, since from $t+1$ countries play the MPE.

Moving up, $D$ must declare war after $s_t = 1$ ((14) holds since (8) fails).

Next, $T$ does not want militarize if and only if

$$-k + p_s\frac{w_T(0)}{1-\delta} + (1-p_s)\left(z^* + \delta\frac{w_T(1)}{1-\delta}\right) \leq \frac{z^*}{1-\delta} \tag{31}$$

$$\Leftrightarrow z^* \geq w_T(0) + \frac{(1-p_s)\delta\big(w_T(1) - w_T(0)\big) - (1-\delta)k}{p_s + (1-p_s)\delta} \tag{32}$$

which ensures that $T$ accepts $z^* \geq w_T(0)$.

Third, the best deviation for $D$, after $s_t = 0$, is to offer $z_t = w_T(0)$, since it gives $1 - w_T(0) + \delta V_D^{MPE}(0,1,0)$, strictly greater than $V_D^{MPE}(0,1,0)$, the payoff of declaring war. $D$ does not offer $z_t = w_T(0)$ if and only if

$$1 - w_T(0) + \delta V_D^{MPE}(0,1,0) \leq \frac{1 - z^*}{1-\delta} \tag{33}$$

$$\Leftrightarrow z^* \leq w_T(0) + \delta \frac{1-\delta}{1-\delta^{N+1}}\big(1 - w_T(0) - w_D(0)\big) \tag{34}$$

(32) and (34) hold if and only if (12) holds. (12) may fail. For example, if $w_T(1) - w_T(0) = k = f\big(1 - w_T(0) - w_D(0)\big)$, for $f > \frac{1}{(N+1)(1-p_s)}$, then $\exists \delta'' \in (0,1)$ such that $\forall \delta \in (\delta'', 1)$, (6) and (8) fail and yet (12) fails.

**Proof.** (Proof of corollary 3). War occurs with positive probability if and only if (6), (8), and (12) fail.

(i) As $p_s$ increases, (6) and (12) are less stringent.

(ii) If (6), (8), and (12) fail, the probability of preventive war is

$$1 - (1 - q^* p_s) r^*$$

$$= 1 - \frac{(1-\delta)k}{\delta\big(w_T(1) - w_T(0)\big)} \frac{1 + \frac{1}{1-\delta}\left(\frac{\delta\big(w_T(1) - w_T(0)\big)}{\frac{1-\delta}{1-\delta^{N+1}}\big(1 - w_T(0) - w_D(0)\big)} - 1\right)}{1 + \frac{1-p_s}{1-\delta}\left(\frac{\delta\big(w_T(1) - w_T(0)\big)}{\frac{1-\delta}{1-\delta^{N+1}}\big(1 - w_T(0) - w_D(0)\big)} - 1\right)} \tag{35}$$

which is decreasing in $p_s$.

(iii) If (6), (8), and (12) fail, the share of preventive wars that are mistaken is

$$1-\frac{\big(1-r^*(1-p_s)\big)q^*}{1-(1-q^*p_s)r^*}=1-\frac{1-\frac{(1-\delta)k}{\delta[w_T(1)-w_T(0)]}}{\frac{1-(1-q^*p_s)r^*}{q^*}}$$

which is decreasing in $p_s$ if $\partial\frac{\frac{1-(1-q^*p_s)r^*}{q^*}}{\partial p_s}<0$. This is indeed the case since, using (35),

$$\frac{1-(1-q^*p_s)r^*}{q^*}=1+\frac{1-p_s}{1-\delta}\left(\frac{\delta\big(w_T(1)-w_T(0)\big)}{\frac{1-\delta}{1-\delta^{N+1}}\big(1-w_T(0)-w_D(0)\big)}-1\right)$$

**Proof.** (Proof of corollary 2). War occurs with positive probability if and only if (6), (8), and (12) fail.

(i) As $N$ increases, (8) and (12) are more stringent.

(ii) The probability of preventive war, $1-(1-q^*p_s)r^*$, decreases in $N$ since $q^*$ decreases in $N$.

(iii) The share of preventive wars that are mistaken increases in $N$ since $\frac{1-r^*}{q^*}+r^*p_s$ increases in $N$, given that $q^*$ decreases in $N$.

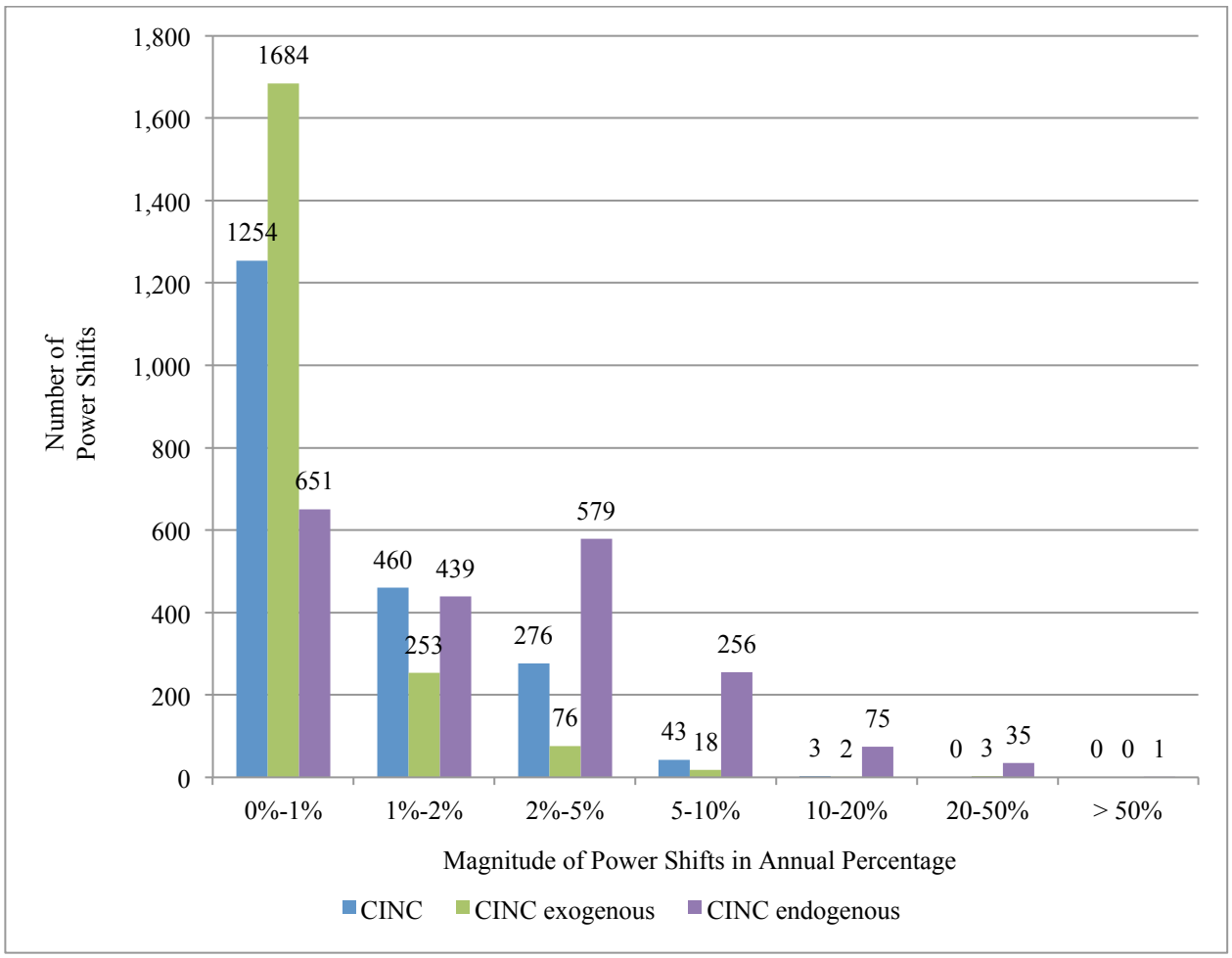
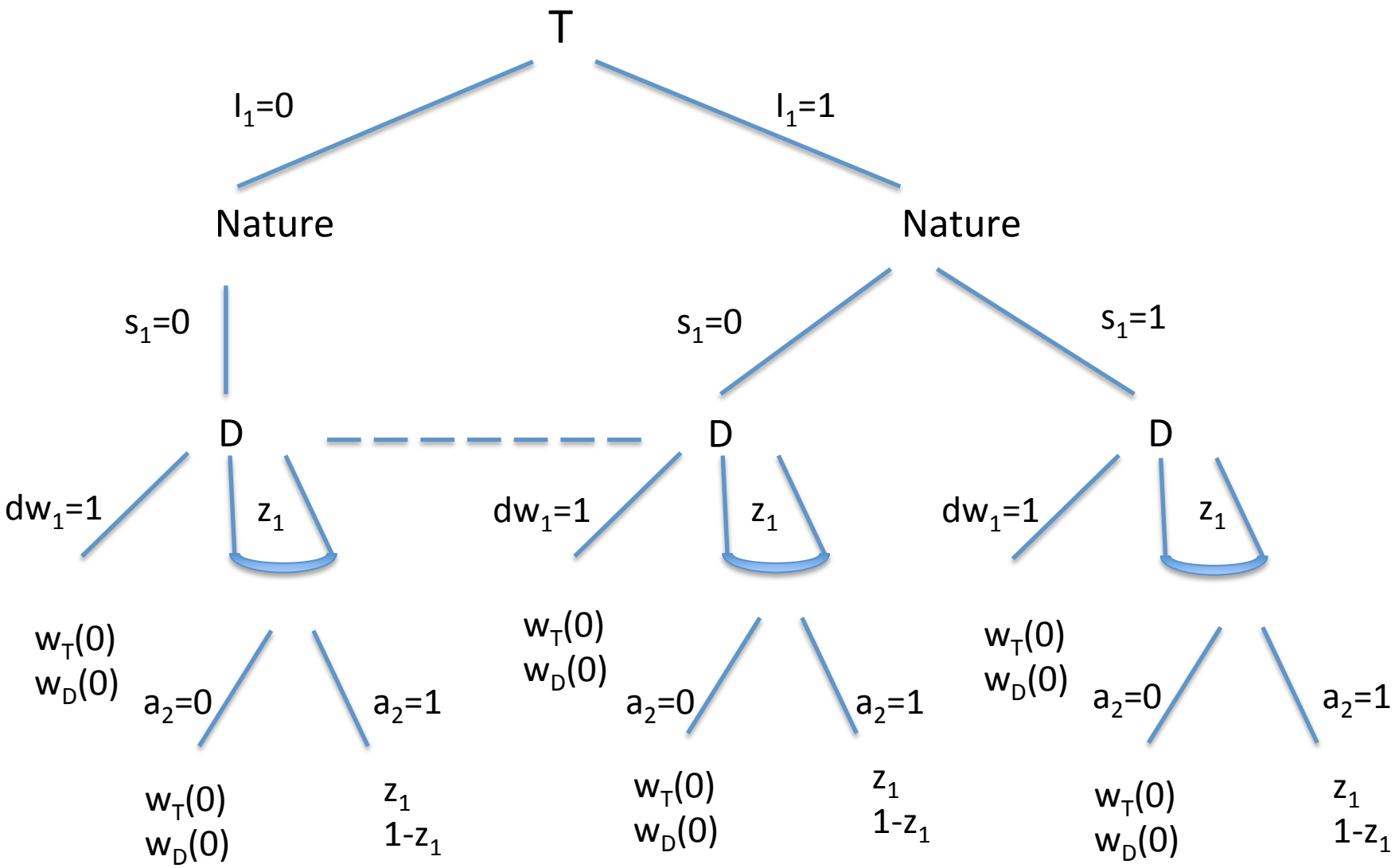



Figure 1: Frequency of exogenous and endogenous power shifts

# Figure 2.1: Game Tree, Period 1



# Figure 2.2: Game Tree, Period 2