Economic Origins of War and Peace

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

Citation	Coe, Andrew. 2012. Economic Origins of War and Peace. Doctoral dissertation, Harvard University.
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:9823976
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Thesis advisor   Author

Jeffry Frieden   **Andrew James Coe**

Economic Origins of War and Peace

Abstract

Why do wars happen, and what do societies fight over? Why are internatioal relations sometimes fearful and aggressive and other times harmonious? I show that these questions can be fruitfully explored by importing some basic economic theory into the existing bargaining theory of war. A separate essay analyzes the interactions between the United States and countries that may be pursuing nuclear weapons.

"Costly Peace: A New Rationalist Explanation for War" posits a new explanation for war: sometimes peace is more costly (in the sense of leaving both sides worse off in expectation) than war. This means that some wars improve overall welfare relative to peace. I develop models for three common sources of costly peace tailored to particular wars and analyze them to expose the common underlying logic for war. The costs of: arming explain the Iraq War; imposition explain the civil conflicts within Iraq after the earlier Gulf War; and predation explain the American War of Independence.

"The Modern Economic Peace" develops a theory of the origins of international disputes, in which the *economic* conflict of interests between two states is determined by the benefits and costs of transferring wealth from one state's economy to the other's. Whether such a transfer happens depends on the military situation be-

tween the two states and also the characteristics of their economies and governments. Nations with sensitive, integrated ("modern") economies of comparable size and representative governments have little to fight over. This might explain not only the puzzling comity of the West, but also long-run global patterns in organized violence, economic liberalization, and democratization.

"A Model of Arms Proliferation and Prevention" is co-authored with Muhammet Bas. We develop a formal model of bargaining between two states, where one can invest in developing nuclear weapons and the other imperfectly observes its efforts and progress over time, and use it to analyze the occurrence of proliferation and war, the viability of non-proliferation agreements, and the role of intelligence-gathering and estimates. The model explains some of the complex phenomena that occur in these interactions, such as mistaken wars, cyclical crises, and the failure of non-proliferation deals.

Contents

Acknowledgments

I offer thanks to my advisors Robert Powell, Dustin Tingley, and especially my chair, Jeffry Frieden, for their counsel on intellectual and professional matters. Thanks also to my collaborators, who were essential not only to the papers we authored together, but also to those on which only my name appears: Muhammet Bas, Adam Glynn, Victor Utgoff, and Jane Vaynman. Contributors to individual papers are thanked in each chapter, but I wish to single out two. Robert Powell's blunt criticism of my papers and wry commentary on the discipline I wished to join were invaluable. James Fearon gave me the most detailed and comprehensive set of comments I've ever received on a paper, which led to a far better version of my job market paper. I am also grateful to the Program on Global Society and Security and the Weatherhead Center for International Affairs at Harvard for their administrative and financial support, and for their provision of a stimulating environment in which to work.

Finally, I offer heartfelt gratitude to those classmates at Harvard who kindly endured the many long conversations and abstruse drafts that preceded this dissertation, and who contributed heavily to whatever eloquence or insight it possesses and whatever success it brings me. These include Amy Catalinac, Richard Nielsen, Iain Osgood, Jonathan Renshon, Jane Vaynman, and especially Jennifer Larson.

Chapter 1

Introduction

Why do wars happen, and what do societies fight over? Why are relations among some states in some eras fearful and aggressive, while other states enjoy harmonious relations, without any threat of violence? Until recently, the body of scientific knowledge about why wars occur consisted solely of a number of observed patterns in the historical record (e.g., violence has decreased over time) and an assortment of informal theories that don't really crystallize the reasons for war or for these patterns in its occurrence. A new line of theorizing asserts that, under certain conditions, negotiations between two political entities over their opposed interests will fail, and the two will instead resort to violence to settle their differences. Unlike previous theories, this "bargaining theory of war" is logically precise and yields clear, testable empirical predictions; perhaps as a result, it has swiftly risen to preeminence among political scientists and economists studying war. However, at present it has severe limitations: it simply takes for granted that any two relevant political entities will have opposed interests, and it treats war and peace as absolutes—there is either all-out savagery or

live and let live. Obviously, neither assumption is realistic.

At least two additions are essential to furthering our understanding of the origins of war. First, the bargaining theory must be broadened to include not just war, but also the many other responses to the failure of negotiation that political entities sometimes employ: arms races, sanctions, the extraction of tribute, proxy conflicts. Second, it must be joined by a theory of the origins of conflicts of interest among political entities: for some nations, the weight of common interests mean that war is unthinkable, while for others with severely opposed interests, the risk of war is ever-present. The first two of the papers presented below constitute a start on these tasks. "Costly Peace and War" offers a game-theoretic analysis of three alternatives to war for responding to failures of negotiation, thereby generating new insights into the origins of several historical wars. "The Modern Economic Peace" develops a model of the incentives for political entities to coerce one another and how these are affected by the characteristics of their governments and economies; in so doing, it offers a new explanation for the recent deep comity of international relations within "the West." The common thread between these papers is their reliance on essentially economic arguments, and it is from this that the dissertation's title arises. Respectively, the choice of war versus another response to negotiation failure is due to a certain kind of cost-effectiveness, while the presence or absence of opposed interests has to do with whether one political entity's prosperity, broadly defined, comes at the expense of another. The final, coauthored paper offered below analyzes the interactions between the United States and countries that may be pursuing nuclear weapons and explains some of the complex phenomena that occur in these interactions, such as mistaken

wars, cyclical crises, and the failure of non-proliferation deals.

An implicit conviction behind all three of the papers presented below is that the best way to advance the bargaining theory of war is to try to apply it to individual cases of war or peace. When the extant versions of the theory fail to explain certain cases, as they do for the cases studied in these papers, then the search for the simplest revision that does yield a plausible explanation also reveals promising avenues for improving the theory more generally. So, each of these papers begins with a certain set of mysterious cases, proceeds to develop a bargaining model that incorporates new features that are specific to the cases, and applies this model to explaining these cases. In so doing, each illuminates gaps in the existing theory that are potentially relevant to all cases.

A second animating conviction of this research is that there are gains to be had from taking economics seriously in the study of international conflict. The bargaining theory of war takes as its point of departure the economic notion of an "inefficiency puzzle" of war: why actors choose a costly means (war) of implementing a settlement, when the same settlement could have been agreed peacefully and implemented without the costs, leaving both sides better off. But it is otherwise economically contentless. The essays "Costly Peace" and "The Modern Economic Peace" incorporate the economic notions of opportunity costs, private goods, and taxation into otherwise standard bargaining models. This additional economic content enables the theory to speak to the possibility of multiple costly means of resolving disputes (e.g., arming, imposition, and predation, in addition to war), to the specific stakes over which disputes arise (e.g., the disposition of private goods), and to the means of altering the

status quo (e.g., taxation for the purpose of transferring wealth from one side to the other). It also renders a number of previously confusing empirical cases explicable within the rationalist framework: the Iraq War, the civil conflict within Iraq, the American Revolution, and the deep comity among certain nations in recent decades.

On a more personal level, these three papers analyze empirical phenomena that I have been trying to understand for a long time; almost a decade in the case of nuclear proliferation and prevention. These papers stem in no small part from the dissonance between the intuitive understanding of international affairs I acquired while working in the US defense community, and the academic body of knowledge on these affairs I have absorbed while a graduate student. I saw in the academic literature no clear explanation for why the Iraq War had occurred, at least not one that squared with the discussions I heard among policy advisors in Washington in the run-up to the war. The explanations the various rationalist theories suggested for why the US and its allies get on so well also struck me as wrong: bargaining efficacy and the costs of war, in my mind, had nothing to do with the way US policy-makers think about the other countries of the West. Finally, the available statistical and game-theoretic analyses of nuclear proliferation and preventive war seemed too simple to render comprehensible the drawn-out, roller-coaster character of the interactions between the US and Iraq, North Korea, and Iran. These essays are, at some level, an attempt to make these two bodies of knowledge, based on policy intuition and academic rigor, consonant.

The first essay, "Costly Peace: A New Rationalist Explanation for War," argues that scholars of war have neglected a theoretically and empirically important explanation for war, namely that sometimes peace is more costly (in the sense of leaving both

sides worse off in expectation) than war. This explanation is theoretically important because it implies that *some wars improve welfare relative to peace.* It is empirically important because it provides sounder accounts of some historical wars than other explanations. I identify three empirically common sources of costs in peace: arming (costly attempts to shift the balance of power), imposition (sanctions imposed or rewards offered for certain policies), and predation (the extraction of tribute, in whatever form). For each, I develop a simple model that is tailored to a particular empirical case of war and whose analysis exposes the common underlying logic of why war happens. The costs of arming explain the Iraq War; the costs of imposition explain the civil conflicts within Iraq after the earlier Gulf War; and the costs of predation explain the American War of Independence.

The second essay, "The Modern Economic Peace," argues that rationalist theories of interstate conflict do a poor job of explaining some international relationships, especially those in which peace seems most comprehensive, such as recent relations among the countries of the West. This is because these theories have little to say about a factor they nonetheless identify as central to war and peace: the presence and magnitude of underlying conflicts of interest. I begin to develop a theory of the origins of disputes among nations, based on the idea that the *economic* conflict of interests between two states is determined by the benefits and costs of transferring wealth (by whatever means) from one state's economy to the other's. Whether such a transfer happens in equilibrium depends not only on the military situation between the two states, but also on the characteristics of their economies and governments. Nations that have very sensitive, highly integrated economies—what are commonly

referred to as "modern" economies—of comparable size and that are ruled by highly representative governments have little to fight or coerce one another over. If this theory is correct, it could potentially explain not only the warmth of relations among such nations, but also the very-long-run decline in organized violence and the rapid global economic liberalization and political democratization of the past two centuries. It also suggests a previously unrecognized link between nations' material interests and their moral sentiments about the proper way for nations to behave.

These first two essays can be thought of as two sides of the same coin. On one side, there are a whole range of costly behaviors in international relations: not just war, but also arms races, proxy conflicts, tariffs, restrictions on foreign investment, and so on. All these things are the result of underlying commitment problems and asymmetric information of varying severity, and they are all partial substitutes for each other; war occurs only when it is the most cost-effective option. So the mechanisms for bargaining failure that rationalist scholars thought were explanations for war are only partially so—we also need to know why war was chosen over other costly behaviors. More deeply, many of the behaviors scholars of international relations think of as interesting are costly, and these are all explicable with the bargaining failures framework and an appreciation of the substitutability of each. This is a deep equivalence between international political economy (IPE) and international security (IS), two bodies of scholarship that until now have developed largely separately. On the other side of the coin, there are some dyads which seem to manifest almost none of these costly behaviors. But it's not that the underlying commitment problems and asymmetric information aren't there, it's that these don't matter because there is

nothing really to bargain over. This is peace, not in the sense of the absence of war, but peace in the sense of the absence of conflict altogether, at least at the level of national governments. The theory I offer builds another bridge between IPE and IS: now both sides have a common framework for exploring the origins of conflicts of interest.

Turning to the final essay, the spread of nuclear weapons has been one of the most powerful sources of international conflict since the second world war. Analysts of nuclear proliferation have long recognized and catalogued the complexity of the strategic interactions among states that might seek nuclear weapons and others that wish to prevent them from doing so. However, there is considerable confusion about how to parse these interactions. Suppose, for instance, that the United States acquired military capabilities that lessened the cost of a war to prevent proliferation. It is understood that this would lessen the incentives for other states to pursue nuclear weapons, thus decreasing the likelihood of proliferation or war, and also decrease the willingness of the US to offer inducements in exchange for a state's agreement to accept inspections and abandon its nuclear programs, thus increasing the chance of proliferation or war. But extant theories have no ability to determine which effect would dominate, and so cannot offer firm predictions for how any given interaction will turn out or prescriptions for how to change the likely outcome.

"A Model of Arms Proliferation and Prevention," which was written with Muhammet Bas, is the latest of a series of papers we've written that seeks to build a rigorous theory of these interactions and thereby dispel the confusion. We develop a formal model of bargaining between two states, where one can invest in developing nuclear

weapons and the other imperfectly observes its efforts and progress over time, and use it to analyze the occurrence of proliferation and war, the viability of non-proliferation agreements, and the role of intelligence-gathering and estimates. We show that "surprise" proliferation, sporadic crises over the uncertain progress of a proliferant's efforts, and "mistaken" preventive wars can all arise endogenously in the model. We find that much of the variation in behavior over time is driven, not by exogenous factors like the costs of war and the effects of proliferation, but by stochastic elements such as when the proliferant's program will make progress and when the other state will discover this. Moreover, while exogenous factors do influence the probabilities of war, proliferation, and non-proliferation, their effects are often counter-intuitive and non-monotonic. We also find that non-proliferation can be undermined by the possibility of a "better deal" once the proliferant's program has made progress, suggesting that some states invest in a program as much to secure a more favorable non-proliferation deal as to actually get nuclear weapons. The analytical discipline provided by our model enables us to assess competing effects like those in the above example, so that we can answer some questions that previous theories could not. However, it also reveals that some widely-agreed intuitive answers to other questions are in fact wrong, or at least only right under certain conditions. It is our hope that this work will spur further reexamination of the conventional wisdom about nuclear proliferation and how the United States and other nations should respond to it.

Note that the proofs for all propositions stated in the three essays, as well as a description of the algorithm employed in the final essay, are given in an appendix that is available, on request, from the author.

Chapter 2

Costly Peace: A New Rationalist Explanation for War

Abstract:[1] I argue that scholars of war have neglected an important rationalist explanation for war: sometimes peace is more costly than war. This explanation is important because it provides sounder accounts of some historical wars than other explanations. To demonstrate this, I identify three empirically common sources of costs in peace: arming, imposition, and predation. For each, I provide a simple model that demonstrates the conditions under which war can occur and show that these wars are not due to other rationalist explanations. I then offer analytic narratives of the Iraq War, the civil conflicts in Iraq after the Gulf War, and the American Revolution based on these models, and argue that these accounts fit the facts better than other

rationalist accounts.

2.1 Introduction

Why do wars happen? Since wars are destructive, it is not obvious why they would be chosen over peaceful bargaining as a means of resolving contested interests between rational actors. The bargaining theory of war therefore focuses on explaining why peaceful bargaining might fail and thereby lead to costly violence. This increasingly influential and wide-ranging body of scholarship focuses on two main causes of war: shifts in the balance of power between two disputants, which can lead a side that fears decline to attack the other in order to forestall its rise; and asymmetric information about a side's willingness or ability to resolve an issue through war, which may motivate one side to risk war in order to call the other's possible bluff. These two rationalist explanations have been applied to particular wars, whole classes of wars, and even specific aspects of war and international relations more generally.[2] They are becoming pervasive in the study of conflict by both political scientists and economists, to the exclusion of many alternative explanations for war (Jackson and Morelli, 2009*a*).

This essay is intended to convince you that a third rationalist explanation for war—costly peace—is of comparable importance to the other two, despite its relative neglect by scholars. The essence of this explanation is that the anticipated costs of peace may exceed those of war. Rational actors will then make demands of each other that cannot be mutually satisfied, because the total value of any feasible peace is less

---

[2]For recent reviews, see Jackson and Morelli (2009*a*); Powell (2002); Reiter (2003).

than that of war. Thus, bargaining will fail and war will occur, even in the absence of any shift in power or asymmetry in information.

This is not merely a theoretical possibility. There are several empirically common sources of costs in peace, and this essay will argue that there are historical wars that are more cogently explained by costly peace than by the other rationalist explanations. Thus, this explanation is important because it is relevant to understanding the origins of empirical wars.

This essay explores three sources of costly peace, termed arming, imposition, and predation. For each, a simple model is provided that exposes the way in which this source can make peace costly and lead to war. Each model is then used to construct an analytic narrative of an empirical war, and evidence is presented that these wars derive from the costs of peace introduced by each source.[3]

First, actors may take expensive measures, most commonly arming, in order to defend their claims to disputed stakes. If this costs enough over the longer term, they may prefer to fight immediately in order to avoid paying these costs. It will be argued that the US-Iraq War that began in 2003 arose from the costs of maintaining the balance of power between the belligerents. More generally, wars of consolidation, secession, and succession may derive substantially from the costs of arming.

Second, peace may mean the imposition of penalties on, or removal of rewards from, two potential adversaries by outside actors. If these are severe enough, war may result as actors fight to obtain these external gains. It will be argued that the civil conflicts in Iraq that followed the Gulf War of 1990–91 arose from the sanctions

---

[3]For more on the method of analytic narratives, see Bates et al. (1998).

imposed by the international community in the aftermath of the war. More generally, powerful countries often impose sanctions or offer rewards explicitly aimed at inciting regime change, violently if necessary.

Finally, peace sometimes entails the transfer of resources or the exaction of tribute from one side by another. This predation undermines incentives for productive activity, and so is costly. If the costs are high enough, war may occur as one side tries to lessen the other's predation and the other asserts control. It will be argued that the American War of Independence was fought because Britain had begun to prey on the thirteen colonies, and the colonists expected more. More generally, wars from ancient Rome through to 19th century America were fought to discourage or eliminate piracy and raiding.

To be clear, theorists of war have known of this third rationalist explanation for some time. Powell (2006) points out the possibility that the costs of arming might alone suffice to cause war. Powell (1993, 1999) analyze models of arming in which the costs of arming might cause war. McBride and Skaperdas (2007) uses this mechanism to explain why conflict happens in low-income countries, McBride and Skaperdas (2009) tests it in a laboratory experiment, and McBride, Milante and Skaperdas (2011) explores how good institutions might mitigate it. Fearon (2011*b*) uses this mechanism to provide a theory of the democratic peace. Slantchev (2010) models a related mechanism, whereby the need in peacetime to repay debt taken on to increase military strength can lead to war. Bas and Coe (2012*b*) analyzes nuclear proliferation, showing that wars over the spread of nuclear weapons are most likely to arise from the costs of containment. And finally, Powell (2011) shows that the

imposition of outside benefits for decisive outcomes can also cause war.[4]

What is missing from this body of research on costly peace is a *demonstration of its ability to explain real wars.* The primary contributions of this essay are to provide three models of war due to costly peace that are tailored to explaining why specific historical instances of war happened, and to show that these accounts fit the facts better than others. It also offers analyses of the three models that are designed to expose the underlying common logic of war due to costly peace, and identifies and analyzes predation, a source of costly peace that has not previously appeared in the literature. Finally, it offers suggestions intended to help researchers in applying this explanation to other wars.

The next section explains costly peace and its relationship to war, situating it within the bargaining theory of war and explaining some previously unnoticed implications of costly peace for the broader understanding of war. Section 3 presents a model of war due to the costs of arming that is tailored to the context of the interaction between the United States and Iraq after the first Gulf War. It then constructs an analytic narrative of the Iraq War based on the model, and evaluates the performance of this model against other explanations for the war. Section 4 does the same for imposition and the civil conflicts that took place in Iraq between the Gulf War and the Iraq War. Section 5 does the same for predation and the American War of Independence. Section 6 concludes with suggestions for how to determine if the costs of peace could have caused a particular war, and points out additional candidates for such wars.

---

[4]A line of work following Hirshleifer (1991) analyzes the distortionary effects of predation, but does not make the connection to war and the failure of peaceful bargaining explicit.

2.2   Costly Peace and War

It is helpful to begin by explicating the quintessential elements of the bargaining theory of war and its way of framing war as an "inefficiency puzzle." With these elements in hand, I will explain what costly peace is and how it can lead to war. I then examine its relation to the other rationalist explanations for war.

Cheap Peace and the Inefficiency Puzzle of War

The bargaining theory of war models a situation in which two rational actors have opposed interests over some stake, whatever it might be, and may attempt to revise its extant disposition.[5] They have just two means of doing so: they can bargain in an attempt to come to a revision that is implemented by mutual agreement, or they can fight to impose a revision unilaterally. Bargained revision (peace) is assumed to be free; violent revision (war) is assumed to be costly.

From this perspective, war is mysterious. Why would war ever occur, when the participants could simply peacefully implement the expected outcome of a war, thereby avoiding its costs and doing strictly better? When the actors are risk-neutral over the contested stake, as is usually assumed, any peacefully agreed outcome is efficient, while war is inefficient because it imposes costs. Thus, the "inefficiency puzzle" of war is why actors would choose an inefficient means of resolving a dispute when an efficient means is available. Even if actors are allowed to be risk-averse, then peace may not be efficient, but it is still always "cheaper" overall than war.

The crux of this framing is that peace is cheaper than war, not because of some

---

[5]The exposition of the bargaining theory of war given here relies heavily on Fearon (1995).

deep empirical law, but *by construction.* The seminal models of the theory are designed so that fighting is the only thing that imposes costs on the players: there is nothing else they can do, and nothing else that can happen to them, that will destroy value. This probably explains why costly peace was initially missed as a rationalist explanation for war. With a very few exceptions, even the many extensions of the seminal models, dealing with domestic politics, intra-war bargaining, diplomacy, and many other aspects, nonetheless retain this property.[6]

This construction has been very useful in advancing the theory, but it is importantly flawed. As subsequent sections will demonstrate, there are several empirically important sources of costs in peace. Each of the models presented there makes use of a simple property of costly peace: when the costs of peace are higher than those of war, then war must occur. To see why this is true, we first need to talk about what costly peace is.

When Costly Peace Leads to War

Peace is costly if it involves one or both actors taking measures other than war that nonetheless destroy some of the value—whether wealth or any other desirable thing—that could otherwise be realized by the two. Arming is one such measure: it destroys value because resources that could be consumed and enjoyed are instead devoted solely to protecting one actor's share of the total value or taking more of the other's. Another is that one actor might not agree to a settlement that would bring in external rewards; by doing so, he imposes the loss of these rewards on both. And

---

[6]The exceptions will be discussed in subsequent sections.

finally, if one actor produces less because the other will prey on the fruits of his labor, then there will be less goods and service to consume.

War might lead to a reduction in these various costs. For instance, if one actor is vanquished, the other will no longer have to arm against him to protect his share of the value. In the case of imposition, if the truculent actor is defeated, the other can freely implement a settlement that would reap the external rewards. And if the predator is eliminated, then there is no more discouragement to production and so no loss of consumption. Of course, war is itself costly in that it also destroys some of the value available to the players. So the costs of war can be thought of as including both the direct cost of fighting, and also any costs of peace that will remain after the war.

If the anticipated costs of peace are greater than the anticipated costs of war, then fighting would be expected to leave more value for the surviving actor(s) to enjoy than the two combined would take from peace. That is, war would increase the value available to the actors over what they could realize from peace. When this is true, there is no peaceful deal that both actors would prefer to war. Any deal that gave one actor as much value as he would expect to get from war would leave the other actor short of his own war value. Because of this, there is no way to divide up the value of peace so that both actors are satisfied.

If, by contrast, the anticipated costs of peace are less than those of war, then fighting would be expected to leave less value to the actors than settlement. That is, war would decrease the value available to the actors relative to peace. When this is true, there are settlements that would divide up the value of peace so that both

actors got at least their war values. This does not guarantee the actors would not fight, as costly peace is not the only rational cause of war, and if peace is costly, then it might exacerbate another cause of war. But it does mean that, when this condition holds, the costs of peace alone do not suffice to cause war.

Thus, costly peace is sufficient to cause war if and only if the costs of peace exceed the costs of war. This fact is true, but vacuous, for the seminal models of the bargaining theory of war. In those models, as in all the models to be presented here, the costs of war are always positive because war entails destruction. But with the typical assumption of risk-neutrality, the costs of peace are always zero by construction, because in these games there is nothing the players can do that destroys value other than fight. Even if risk-aversion is allowed, the costs of peace are always strictly less than those of war, because the latter then entails not only destruction but also costly uncertainty. Thus, the costs of war exceed those of peace, and war can never happen due to costly peace in these models.

It is enlightening to to restate this fact in terms of efficiency: costly peace leads to war because it means that peace is even more inefficient than war. From this perspective, the inefficiency puzzle of war also applies to costly peace.[7] If the actors could simply not take the measures that make peace inefficient, then peace would be efficient and thus yield more value than both costly peace and war. But then there is a way to divide the value of peace so that both actors would get more value from this division than from costly peace or war. If this is true, why would actors choose either to make peace costly or to fight?

---

[7]Slantchev (2010) makes this observation as well.

The answers are just the same as those used in the bargaining theory of war: commitment problems (CP) and asymmetric information (AI). At root, the rationalist explanations for war offered by Fearon (1995) are actually rationalist causes of *inefficiency*, whether it comes in the form of war or of costly peace.[8] If one of these causes is present, then the occurrence of costly peace or war simply depends on the availability and costs of each.

One might then conclude that, theoretically speaking, there is nothing new here. If CP and AI cause costly peace, just as they cause war, then what have we learned? One implication of taking costly peace into account is that, while CP or AI might be necessary causes for war to occur, their presence alone does not suffice. If there are costly measures other than war that the actors could take in response to CP and AI, then the presence of these causes might lead only to costly peace. War will happen only if these measures are ineffective or too costly. The seminal bargaining models of war missed this because they rule out such measures by assumption.[9]

Implications of Costly Peace

Empirically, there are in fact often costly measures other than war that actors can employ in response to deal with both CP and AI. In some cases, these are sufficiently cheap and effective so that war can be avoided; in others, war is cheaper. An immediate implication is that, to fully explain an actual case of war, one needs

---

[8]Fearon (1995) offered not only commitment problems and asymmetric information as rationalist explanations for war, but also issue indivisibility. However, Powell (2006) shows that this last explanation in fact relies on a commitment problem.

[9]Of course, these models were designed to pose the inefficiency puzzle of war in the starkest possible terms, and for that purpose assuming away costly peace is a valuable simplification.

not only to determine the ultimate source of inefficiency (whether CP or AI or both), but also why costly measures other than war were not employed instead. A second implication is that, to the extent that applications of the seminal models are used to make predictions, these will be biased toward predicting war, because the possibility of other costly measures to address the underlying CP and AI is ignored.

A third, and more unsettling, implication is that in some cases wars may actually improve the welfare of the belligerents relative to what would happen if they remained at peace. If, in the absence of war, the actors would take costly measures to address underlying issues of CP and AI, then peace will be costly. If it is more costly than war, then war would improve their social welfare, and so preventing war between these actors might actually make them worse off. This is radically different from the implications that follow from models of war where peace is always efficient, and so bears some explanation.

In the absence of costly peace, there is necessarily more value to go around before a war than after. Thus, there is always at least one actor who is made worse off by war in expectation and so regrets the war the moment it starts—because of this, war without costly peace can never be a Pareto improvement. To illustrate with the three causes of war from Fearon (1995), in the case of shifting power, the regretful player is the one who expects to gain advantage in the future. With asymmetric information, the regretful player is the one who is surprised by an unexpectedly strong adversary. With issue indivisibility, there is always one player who would receive more than his war value in peace, and this player has reason to regret a war.

By contrast, because a war due to costly peace increases the total value available

to the actors, at least one of them must be left strictly better off by war. It is even possible that, in expectation, *both* actors will do strictly better from war.

The crux of this difference is that wars in the absence of costly peace are entirely about one actor's willingness to pay to prevent a too-generous portion of the total value from being allocated to the other. Thus, regardless of who wins, these wars merely shift value between actors. The problem is that this shift comes at a cost: the total value is reduced because the war is costlier than peace. This is why at least one actor must be left worse off. Before the war, the expected allocation favors one player, who would like to prevent war but cannot because of budget, credibility, or divisibility constraints, and who thus regrets the war. In contrast, wars due to costly peace are not about altering the allocation of value, but about increasing the total value available. It is this increase in value that makes a Pareto improvement from war possible.

This property of wars due to costly peace calls for a fundamental re-examination of the policy advice that typically follows from bargaining models of war. It implies that there may be some wars—those due to costly peace—that are in the actors', and possibly the world's, best interest. Thus, unlike wars in the absence of costly peace, it may not always be true that wars due to costly peace should be prevented or stopped.

Developing this implication is beyond the scope of this paper, but I can give one example to substantiate the claim that a re-examination is needed. With wars in the absence of costly peace, it is always true that increasing the costs of war by enough will cause the actors to forego war and thus leave them better off overall. (Increasing

the costs of war by too little just means that war still happens but is more costly for the participants, making them worse off.) Thus, the application of sanctions, pressure for a ceasefire and mediation, and introduction of peace-keeping forces are all steps that, if they are strong enough to prevent or stop a war, should be taken. However, with wars due to costly peace, increasing the costs of war enough to take war out of equilibrium can actually leave the participants worse off if it does not also decrease the costs of peace enough. This implies that an ability to distinguish whether a particular ongoing war was primarily due to costly peace would be extremely valuable for policy-makers. I'll return to this problem in the final section of the essay, as it is also relevant for scholars looking to make use of costly peace in explaining historical wars.

The subsequent sections of this essay will substantiate the empirical relevance of these ideas. Three historical wars will be examined using models that explicitly incorporate costly measures that actors can take in response to CP and AI. I will argue that in all three cases, the available measures would have effectively eliminated the difficulties associated with CP and AI in these contexts. And I will endeavor to show that these measures were abandoned in favor of war because they were more expensive than war.

2.3   Arming and the Iraq War

The first, and perhaps most obvious, empirical source of costs in peace is *arming*: expensive measures taken by either actor to improve its prospects in war or undermine its opponent's. It is well-known to bargaining theorists that if the anticipated costs of future arming exceed those of war, then actors will choose war because it is

cheaper than sustained arming (Powell, 1999, 2006; McBride and Skaperdas, 2007, 2009; McBride, Milante and Skaperdas, 2011; Slantchev, 2010; Bas and Coe, 2012*b*; Fearon, 2011*b*). But it is not known whether this explanation for war is empirically relevant: is there a historical war that is plausibly due to the costs of arming?

In this section, I will resolve this question by giving an analytic narrative of the 2003 US-Iraq War (henceforth, the "Iraq War"). The narrative is based on a model of war and arming (in this case, containment) that is tailored specifically to the interaction between the United States and Iraq prior to the war. To apply this model, a set of assumptions must be made that calibrate the model to the empirical circumstances that pertained during the decade of US containment of Iraq. From these assumptions, a single relationship is derived that governs whether war occurs: if the costs of containment come to be perceived as exceeding the costs of war, then there will be war. The analysis of the model yields a new account of the Iraq War—based on costly peace—that is quite different from existing accounts and that more closely fits the empirical record of what happened. It thereby demonstrates that the costs of arming is an empirically important explanation for war.

Since I will argue that the model applies directly to an actual war, the empirical accuracy, or at least theoretical innocuousness, of its features must be carefully described and defended. After doing this, I will state the calibrating assumptions to be used, and defend their empirical validity. There follows a proposition that characterizes the equilibrium outcome under these assumptions, which is proved in the appendix. We will discuss the intuition for this result: why costs arise in peacetime, how war is expected to reduce them, and why the actors could not otherwise avoid

these costs. And finally, I will give a narrative of the Iraq War based on this analysis and compare it to existing accounts.[10]

A Model of the US-Iraq Interaction after the Gulf War

To begin constructing the model, suppose there are two players, the US and Iraq. Given the totalitarian rule Saddam Hussein exerted over Iraq, modeling the country as a unitary actor in this way is not a bad approximation. In the US case, there were of course many different influential actors involved in decision-making over Iraq, but in the end most executive officials, majorities of both houses of Congress, and a large majority of the American public supported the strategy the Bush administration pursued, so that the US can be taken as, in practice, acting unitarily.

The two players have conflicting interests over a set of issues, represented by the unit interval $[0, 1]$, which they must somehow divide between them over time. These issues include influence over other states in the region, control over regional oil reserves, relations with regional terrorist organizations, the treatment of the people of Iraq, and perhaps others. Any settlement of these issues is a point in $[0, 1]$; assume for convenience that the US favors settlements closer to one, while Iraq favors settlements closer to zero. Further, assume for simplicity's sake that both players are risk-neutral over these issues, and that both discount payoffs over time at a constant rate $\delta \in$

---

[10]There are a great many non-rationalist (or at least, not formally rationalist) accounts of the war, in terms of the personal characteristics of key decision-makers, the confluence of key ideas among important elites, dysfunctional bureaucracies, concerns over re-election, special interests, and misperceptions (Flibbert, 2006). While these accounts elucidate the many issues over which the US and Iraq had conflicting interests, they do not crystallize why the two states could not find a peaceful settlement that both would prefer to war. By contrast, the few existing rationalist accounts of the war, to be cited and described below, focus on explaining the failure of peaceful negotiation. Because this is also the focus of my narrative, I will compare it only to the other rationalist accounts.

$(0, 1).$[11] Thus, the utility of a sequence of payoffs from settlements, with q_t the implemented settlement at time $t \in \mathbb{N}$, is $U_{US} = \sum_{i=0}^{\infty} \delta^i q_i$ for the US and $U_{IR} = \sum_{i=0}^{\infty} \delta^i (1 - q_i)$ for Iraq.

The first period of the game can be thought of as the period immediately after the end of the Gulf War of 1990–1991. It begins with the US choosing whether to engage in containment, an action to be discussed momentarily. The US then chooses either to go to war against Iraq or to make a peaceful offer of a settlement for that period. Iraq can then accept the US offer, in which case it is implemented for that round, or reject it and go to war.[12] If Iraq accepts the US offer, then it can also choose whether or not to try to acquire nuclear weapons. If a peaceful settlement is agreed, then in every subsequent round, Nature moves first and determines whether Iraq's effort in the previous period to acquire nuclear weapons is successful, and whether the US detects this effort. If it is successful, then Iraq is assumed to have nuclear weapons for that and all future peaceful periods. If it is not, then Iraq must try again to have a chance of getting nuclear weapons in the next period. After Nature's move, the choices described above repeat. War is treated as a game-ending costly lottery, to be described shortly.

I am assuming here that the acquisition of nuclear weapons is an inherently un-

---

[11]These assumptions are standard in the bargaining theory of war, but it is worth noting the effects of relaxing them. None of the qualitative results derived here will change if I allow the players to be risk-averse, or to discount future payoffs at differing rates, so long as the discount is always positive and grows over time. Allowing the players to be risk-acceptant generates the possibility that war occurs purely due to the desirability of gambles, but the costs of arming that are the focus here would still contribute to the incentives for war.

[12]Here, for expositional simplicity, the US is modeled as making take-it-or-leave-it offers to Iraq. The choice of bargaining protocol will not affect the results, as they do not depend on the allocation of the bargaining surplus, but only on whether a surplus exists.

certain endeavor. Many technologically-sophisticated inputs are required to produce a nuclear weapon, and successful mastery of each of these inputs takes an uncertain amount of time. Additionally, both containment and the attempt to acquire nuclear weapons are represented as binary choices—the US cannot choose *how much* to contain Iraq, and Iraq cannot choose how hard to try to get nuclear weapons. Allowing for continuous choices here would complicate the analysis considerably, but would not change the conclusions presented below, because the problem that leads to war does not derive from the inability of the US or Iraq to choose just the right degree of weapons development or containment. I also assume that the US cannot react instantly to Iraq's efforts to acquire nuclear weapons—if Iraq tries to get them, there is some probability it will be successful before the US can launch a war. Even if the US instantly and accurately observed Iraq's efforts, it would take some time to mobilize itself for war, during which Iraq might be successful.

Most importantly, I will assume that the probability that governs Iraq's development efforts depends only on whether the US elects to contain or not—it cannot, for example, rise over time as Iraq keeps trying and masters various inputs to nuclear weapons. Allowing for this to occur when the US does not contain would strengthen the conclusions, because it would increase the incentives for both containment and war, but it is important for the results that the probability be bounded above when the US does engage in containment. This restriction can be defended on empirical grounds: even strident advocates in the US for war believed that Iraq's efforts could be held in check so long as the US did what was necessary to uphold the sanctions and remained willing to compel Iraq's periodic opening to inspectors (Pollack, 2002),

so that any progress it made could be detected and reversed. It can also be defended on theoretical grounds: Bas and Coe (2012*b*) models this interaction and shows that the US can effectively cap the probability that Iraq's efforts are successful over time. Since Iraq has the ability to credibly reveal its progress at any given time by easing access for inspectors, and the US becomes willing to go to war if its estimate of the progress (and thus, the likelihood of a subsequent shift in the balance of power) gets too high, it is always in Iraq's interest to avoid war by opening up to inspectors and reassuring the US that the probability remains low.[13] The model presented here abstracts away from these aspects in order to focus on the costs of containment.

War is assumed to have only two possible outcomes: a complete US victory over Iraq and a (much less probable) complete Iraqi victory over the US. Allowing for a larger range of possible outcomes would not qualitative alter the results, so long as there remains a substantial probability of a complete US victory that would be expected to eliminate the costs of containment. The expected outcome of a war started in a given period, as well as the probabilities that in the next period Iraq will obtain nuclear weapons and that its efforts will be detected, depend only on whether the US engaged in containment in the given period and whether Iraq had previously acquired nuclear weapons. First suppose that Iraq does not have nuclear weapons. If the US does not contain, then the probability of US victory in war is p, the costs of

---

[13]Some accounts of the war argue that Saddam's need to maintain ambiguity about his weapons programs, especially about his ability to deploy chemical and biological weapons, in order to deter foreign enemies such as Iran and domestic insurrection, impeded his ability to signal clearly to the US that he had abandoned these programs (Baliga and Sjöström, 2008; Lake, 2010). However, it seems clear that Saddam would prefer fighting Iran or an Iraqi opposition to fighting the US, given the overwhelming military advantages of the US. So, given a choice between certain war with the US in order to maintain ambiguity about his weapons programs, and possible revolt or war with Iran upon revealing the absence of such capabilities, he would choose the latter. As I will explain later, this is exactly what Saddam did once the US threat of war became credible in 2002.

war for the players are $d_{US}, d_{IR} > 0$[14], the probability that Iraq will obtain nuclear weapons in the next period if it tries to get them is $\lambda > 0$, and the probability that the US will detect Iraq's efforts is $\sigma > 0$.[15]

Containment, as the US and its allies practiced it after the first Gulf War, was a strategy designed to ensure that Iraq was unable to threaten or coerce its neighbors, that Saddam was unable to inflict massacres on the Shi'i in the south or the Kurds in the north, and that Saddam's regime was unable to reconstitute and expand its once-formidable pursuit of weapons of mass destruction, including nuclear weapons.[16] It consisted of a comprehensive package of measures to achieve these ends. General economic sanctions were placed on Iraq to starve the regime of hard currency to re-build its military, including restrictions on Iraq's ability to sell its oil, and US forces were stationed at bases near Iraq to ensure a rapid response to any new provocation. The net effect of these measures was to increase the chance that the US would be successful in a new war against Iraq to $p^c > p$ and to lower the cost of fighting such a war to $d_{US}^c < d_{US}$. They also lowered the costs of war for Iraq to $d_{IR}^c < d_{IR}$, since they ensured that the next war would be quick and entail minimal destruction of Iraqi resources in the form of rebuilt military forces. Additionally, the enforcement of no-fly zones in the south and north of Iraq to protect the minorities living there, including suppression of Iraq's air defenses, further increased the US advantage over

---

[14]For mathematical convenience, all war costs are taken to be per-period costs, so that, e.g., the total cost of war for the US is $d_{US}/(1 - \delta)$.

[15]For simplicity, I ignore the possibility of a "false positive": US detection of cheating when none is taking place. Allowing for a small probability of this occurring would not qualitatively change the results.

[16]I rely here on the discussion of the objectives, elements, and effects of containment in Pollack (2002), chapters 2 and 3, and the assessment of their effects in Cordesman (2002).

Iraq and lessened the costs should war recur. Finally, intrusive monitoring of Iraq's military, including repeated inspections of any sites suspected of WMD-relevant activity, and limited strikes as necessary to convince Iraq's regime of the seriousness of US concerns at particular junctions and compel its compliance with inspections, lessened the probability that Iraq would obtain nuclear weapons in the near future to $\lambda^c < \lambda$ and increased the chance that the US would detect such efforts to $\sigma^c > \sigma$. Finally, these measures were themselves costly, so that in each period containment imposes costs of $c_{US}, c_{IR} > 0$ on the players.

Now suppose instead that, at some period after the start of the game, Iraq obtains nuclear weapons. Clearly, this would increase Iraq's chance of prevailing in a new war, and also increase the anticipated costs of such a war for both players. With a nuclear-armed Iraq, the US chance of victory would be $p^n < p < p^c$, and the costs would be $d_{US}^n > d_{US}$ and $d_{IR}^n > d_{IR}$. Note here that I am abstracting away from the possibility of containment after Iraq had obtained nuclear weapons. It is certainly possible that the US would continue some of the pre-nuclear measures, such as the stationing of forces nearby, and perhaps even add some, such as bolstering the missile defenses of its allies in the region, but others might be ended, as Iraq would be unlikely to tolerate limited strikes intended to protect its minorities or compel access for inspectors. Regardless, so long as Iraq would continue to want nuclear weapons even if it realized the US would try to contain it after it acquired them, relaxing this assumption and allowing for costly containment after Iraq got nuclear weapons would only strengthen the conclusions presented below.

Overall, it is assumed that Iraq prefers war when it is nuclear-armed to war when

it is not and the US is not engaging in containment, and prefers both to war when the US is engaging in containment. That is, $1 - p^n - d_{IR}^n > 1 - p - d_{IR} > 1 - p^c - d_{IR}^c$. Analogously, the US is assumed to prefer war under containment to war without containment, and to prefer both to war when Iraq is nuclear-armed, or $p^c - d_{US}^c > p - d_{US} > p^n - d_{US}^n$. I will also assume that, if Iraq has not gotten nuclear weapons and the US intends to go to war, then it is worthwhile for the US to engage in containment prior to starting the war in order to gain the concomitant advantages: $\frac{p^c - d_{US}^c}{1 - \delta} - c_{US} \geq \frac{p - d_{US}}{1 - \delta}$.

Finally, all of the parameters of the game are assumed to be common knowledge. Before turning to the analysis of the game, it is worth remarking on three implicit features. First, I ignore the question of Iraq's *other* weapons of mass destruction (WMD): biological and chemical weapons. Whatever the uncertainties surrounding them, the shift in expectations about war outcomes due to these weapons was small, because the US military was, especially by the 2000s, perfectly capable of winning a war against Iraq under chemical and/or biological attack (Cordesman, 2002; Pollack, 2002, pp. 33–37 and ch. 11). Second, there is no incomplete information about the two sides' preferences. Both sides had had plenty of time to learn each other's interests by 2002; there was no uncertainty about Saddam's desire to obtain nuclear weapons, or the US's willingness to bear substantial costs to minimize the chance of his success (Director of Central Intelligence, 2002).

Third, there is no terrorism in the model. Iraq's support for regional terrorist organizations might be regarded as increasing the costs of peace between the US and Iraq, which would strengthen the results presented below. Iraq might also transfer

WMD or relevant expertise to terrorists, but despite the Bush administration's public assertions, the intelligence community placed low probability on this occurring, except perhaps if Iraq's survival was directly threatened (Director of Central Intelligence, 2002). Perhaps the most well-informed advocate of the war, Kenneth Pollack, supported this assessment and also regarded terrorism as the least of the threats posed by Iraq (2002, pp. 153–158, 178–180). Moreover, there were good reasons to believe Iraq would otherwise be deterred from doing so by the danger of US retaliation for any terrorist WMD attack (Mearsheimer and Walt, 2003; Pollack, 2002). We will return to these issues when we consider alternative accounts of the war.

Calibration and Analysis of the Model

To derive a precise prediction from the model, it is necessary to make four calibrating assumptions that narrow what can happen in equilibrium. Though they are stated verbally to ease understanding, each is equivalent to a certain relationship among the various parameters of the model. These assumptions imply that the equilibrium outcome of the game turns on a simple comparison between the costs of containment and the costs of war. After discussing what the assumptions mean and defending their empirical validity, I will state a proposition that characterizes equilibrium, and then discuss the intuition for this result. The proposition is proved in the appendix, and the translation of each assumption into a condition on the model parameters is performed therein.

Calibrating Assumptions:

 1. *If the US chose to tolerate Iraq's nuclear weapons efforts and gave Iraq nothing*

prior to its acquisition of them, then Iraq would try to get them and would not start a war.

2. *The US would be better off going to war than allowing Iraq to freely pursue nuclear weapons, even if it gave Iraq no concessions at all prior to its acquisition of nuclear weapons.*

3. *Absent containment, Iraq would covertly pursue nuclear weapons under any US-offered deal.*

4. *Containment would decrease the probability of Iraq getting nuclear weapons enough that the US would be able to hold Iraq to its war value.*

Intuitively, the first assumption means that Iraq would choose to wait until it acquired nuclear weapons to challenge the US, even if in the meantime the US conceded absolutely nothing to its interests, so long as the US did not engage in containment. This is crucial to the rest of the analysis: it ensures that Iraq anticipates substantial gains from trying to get the weapons. In the appendix it is shown that this assumption is equivalent to the following inequality:

$$1 - p - d_{IR} \leq \frac{\delta\lambda}{1 - \delta}(p - p^n + d_{IR} - d_{IR}^n) \tag{2.1}$$

The left side of the inequality is how Iraq would expect to do in a war, absent containment. The right side is the total future benefit for Iraq of having nuclear weapons, relative to what it would get without the weapons if the US offered it just enough to avoid war, weighted by the probability that its efforts to get them are successful.

Empirically, this condition surely held. Judging from the earlier Gulf War, the US military was far superior to Iraq's even prior to containment, and the costs of war for Iraq were substantial, so that the left side would be close to zero. Both the US and Iraq believed that Iraq's acquisition of nuclear weapons would substantially improve its ability to extract concessions from the US.[17] Moreover, Iraq was very close to mastering the technology by the time of the earlier Gulf War, so that the probability of success in developing them was high. Even after the war and the years of containment following it, the knowledge remained and the program could be reconstituted in the absence of containment (Duelfer, 2002*b*).

The second assumption means that the US would prefer war to just ignoring Iraq, even if that meant making no concessions to Iraq prior to its acquisition of weapons. It is equivalent to a complicated condition given in the appendix, but if the US is taken to be fairly patient ($\delta \approx 1$), as seems likely, then the condition simplifies to roughly $d^c_{US} + d^n_{IR} < p^c - p^n$. The left side of the inequality is the surplus from avoiding war that the US would gain by tolerating Iraq's nuclear program and conceding nothing to it, and the right side is the shift in the balance of power from Iraq getting nuclear weapons. I have already argued that containment greatly enhanced the already-large military superiority of the US over Iraq, and reduced the anticipated costs of war, so that p^c was close to 1 and d^c_{US} was small. Moreover, for Iraq to perceive substantial

---

[17]For evidence that Saddam saw nuclear weapons as giving him more than just a guarantee of survival, see Pollack (2002, pp. 175–178). Some have argued that such "offensive" gains cannot be acquired with nuclear weapons, even for an aggressive leader like Saddam (Mearsheimer and Walt, 2003). However, it is widely believed within the US policy community that the spread of nuclear weapons, especially to potential adversaries, erodes US power and influence—one need only look at any edition of the National Security Strategy of the United States from the Clinton or Bush (Jr.) administrations to confirm this fact.

gains from nuclear weapons, as assumed and justified above, it must be that p^n is substantially less than p^c (the balance of power shifts in Iraq's favor) and d_{IR}^n is not too large (Iraq can afford to assert the advantages deriving from its nuclear weapons). Thus it is plausible that the inequality held true empirically.

Some observers argued before the war that toleration (and deterring Iraq once it got nuclear weapons) would be better for the US than war (Mearsheimer and Walt, 2003). Having seen the *ex post* costs of the war, many more would agree with this assessment now. However, the US had previously proven willing to run at least the risk of war to stop North Korea from developing nuclear weapons, though it ultimately failed (Mazarr, 1995, Chapter 8). Thus it seems reasonable to assume that, given a stark choice between tolerating Iraq's pursuit of nuclear weapons and fighting a war to stop it, the US would decide on the latter.

The second assumption ensures that the US will be willing to bear at least some cost to try to prevent Iraq's program from being successful. This cost could come in the form of war or containment, which would make the equilibrium inefficient because each destroys value, or a concession offered to Iraq in exchange for abandoning its nuclear program, which would be efficient because it simply transfers value from one player to another.

The third assumption means that the last option won't work: there is no way to avoid Iraq pursuing nuclear weapons without containment or war. If the US offered the most generous concession to Iraq that the US could tolerate in exchange for Iraq abandoning its program, and threatened Iraq with war if it was caught cheating, then Iraq would agree to the deal but still cheat on it. Anticipating this, the US would be

unwilling to offer any such deal. From the appendix, this assumption is equivalent to:

$$(1 - \lambda)\sigma \left[\frac{d^c_{US} + d^c_{IR}}{1 - \delta} + c_{US} + c_{IR} \right] < \lambda \left[\frac{p^c - p^n - d^c_{US} - d^n_{IR}}{1 - \delta} - c_{US} \right] \qquad (2.2)$$

The left side is the surplus from avoiding containment and war that would be lost if Iraq was caught cheating on the deal, weighted by the probability that its covert effort was both unsuccessful and detected. The right side is the gain to Iraq from obtaining nuclear weapons, relative to the generous concession offered by the US to secure the deal, weighted by the probability that its covert effort succeeded. When the inequality holds, Iraq will discount the threat of US punishment, because it is unlikely to get caught before it gets the weapons, and even if it did, the punishment is not so bad.

Empirically, this condition very probably held. No serious analyst doubted that Iraq would reconstitute its nuclear program if containment lapsed, and throughout the era of containment there was little discussion of a deal of the kind considered here. Looking at the condition from the model, it is easy to see why. As already argued, the first and second assumptions imply that Iraq's chance of getting nuclear weapons absent containment (λ) was high, and that the bracketed term on the right side is positive. Containment's reduction of the costs of war (as described earlier) implies that the bracketed term on the left is not too large, but the most important factor there is σ, the chance that Iraq would get caught in the absence of containment. There was every reason to believe that this probability was quite small. Iraq had come close to acquiring nuclear weapons before the earlier Gulf War without the intelligence community or IAEA noticing. Even after years of containment, with its

forced and intrusive monitoring, inspectors were unaware of major elements of Iraq's WMD programs, and only discovered them upon the chance defection of the director of these programs, a member of Saddam's family (Duelfer, 2002*a*).[18]

Finally, the fourth assumption means that containment works well enough that Iraq has relatively little to gain from pursuing nuclear weapons under it, because this effort is so unlikely to succeed. Thus, to avoid Iraq starting a war, the US may have to make a (small) concession to Iraq's interests to make up for the fact that the advantages Iraq would get from acquiring nuclear weapons are unlikely to be realized any time soon. It also means that, under containment, the US is not interested in offering Iraq additional concessions in exchange for a promise to abandon its nuclear program. Any deal that was generous enough to win Iraq's compliance would be strictly preferred by Iraq because of the additional concession it entailed, but because containment works so well, it would not lessen the probability that Iraq would eventually get nuclear weapons enough to make it worthwhile for the US to offer this concession. In terms of the model, this assumption is equivalent to:

$$W_{IR}^c \geq \frac{-c_{IR} + \delta\lambda^c V_{IR}^n}{1 - \delta(1 - \lambda^c)} \tag{2.3}$$

The left side is Iraq's expected value from war under containment; the right side is the value to Iraq of trying to get nuclear weapons under containment while being given nothing by the US. The lower Iraq's chance of success under containment (λ^c), the smaller the right side will be relative to the left.

---

[18]As Pollack (2002, pp. 75–76) explains, this defection was not motivated by any foreign enticement, but rather by an improbable quarrel within Saddam's family that threatened the personal safety of the director. Thus, this highly unlikely revelation should not be considered as indicating good monitoring of Iraq's program.

The evidence available as of the early 2000s suggested that λ^c was very low. After all, starting from 1991, when it had been very close to getting the weapons, a decade had passed under containment and Iraq still did not have nuclear weapons. Moreover, many believed this was not for lack of trying—throughout the 1990s, the intelligence community had reporting from human sources that Iraq's program was continuing (Iraq Intelligence Commission, 2005, pp. 53–55). While this turned out to be false, it would nonetheless surely have contributed to a US perception that λ_c was quite low. Even in 2002, when Iraq had had four years without inspections to make progress, and the erosion of international sanctions had increased the resources available for pursuing nuclear weapons, the US intelligence community estimated that Iraq was still five to seven years away from getting a weapon (Director of Central Intelligence, 2002).

In reality, Iraq's ability to reconstitute its nuclear program decayed steadily after containment began, so that λ_c moved toward zero, in part because of Iraq's decision to temporarily abandon WMD in order to motivate the lifting of sanctions (Iraq Survey Group, 2004, Key Findings). But the bottom line is that the US had every reason to believe that containment had greatly reduced Iraq's ability to develop nuclear weapons, as indeed it did, and the fourth assumption thus held true.

Together, these assumptions imply that the equilibrium outcome must be either containment or war. The fourth assumption means that Iraq is indifferent between these; either way, it just gets its war value. Because of this, if war gives the higher total value for both players then both would prefer it to containment, and vice versa. Equivalently, if war imposes less costs on the players than containment, then it will

occur. More formally, we have:

Proposition 2.1. *Under the calibrating assumptions, war is the unique equilibrium outcome if and only if the costs of containment exceed the costs of war. If the costs of containment are less than the costs of war, the unique equilibrium outcome has the US engaging in containment and Iraq pursuing nuclear weapons.*

Remarkably, the proposition implies that the equilibrium outcome is costly (that is, inefficient), whether it is peaceful or not.[19] The US and Iraq would do better overall if they neither fought a war nor engaged in containment. It is important to understand what causes this inefficiency, and why war occurs under the stated condition. There are three features of the model that are necessary for inefficiency to occur. First, Iraq cannot commit not to take advantage of nuclear weapons to obtain a more favorable settlement from the US once it has them. Second, the US cannot perfectly and costlessly police Iraq's pursuit of nuclear weapons. And third, absent costly containment by the US, Iraq cannot commit not to pursue the weapons.

The first commitment problem is exactly the same as the problem that leads to war in the standard models of shifting power (Fearon, 1995; Powell, 2006). If Iraq could commit not to take advantage of newly-acquired nuclear weapons, and did so, then the US would not care whether Iraq pursued nuclear weapons or not, and there would be no reason to engage in containment or fight a war. And if Iraq could commit, then it would certainly do so in equilibrium, because the US would be able and willing to entice it to do so by offering it part of the surplus gained from avoiding the costs

---

[19]The statement of the proposition ignores the knife-edge case where the costs are equal, in which there is an equilibrium with war and an equilibrium with containment.

of containment or war. Play would thus be efficient.

Since Iraq cannot commit not to take advantage of nuclear weapons if it gets them, the US does care whether Iraq tries to develop them. If the US were able to perfectly and costlessly monitor Iraq's efforts, and able to react instantly to the initiation of such efforts, then it could threaten to react to Iraq's efforts with war or containment. Faced with this threat, Iraq would be deterred from pursuing nuclear weapons, because the instant it did so, the US would detect it and take action to stop it—there would be no chance for Iraq's efforts to bear fruit, and thus nothing to gain from them. So, with perfect monitoring and instantaneous reaction, the commitment problem described above could be avoided, because Iraq would not pursue nuclear weapons and thus no shift in power could occur. Then, neither war nor containment would be needed and play would be efficient. Unfortunately, this possibility is ruled out in the model because, without containment, it is too easy for Iraq to conceal its efforts and takes too long for the US to react to them once detected.

Even imperfect monitoring and delayed reaction on the part of the US would not matter were it not for the second commitment problem. If Iraq could commit not to pursue nuclear weapons, and did so, then US monitoring of and reaction to Iraq's efforts would be irrelevant, as would be Iraq's inability to commit to not taking advantage of the weapons once it had them, and there would again be no reason to engage in containment or go to war. Once again, Iraq would make this commitment in equilibrium, because the US would entice it do so with a share of the surplus from avoiding the costs of containment or war, and play would be efficient.

Since Iraq cannot commit not to use nuclear weapons to coerce the US once it has

them, the US cannot perfectly and costlessly police Iraq's nuclear efforts, and Iraq cannot commit not to pursue the weapons, the US is willing to employ costly measures to stop it from trying and being successful. This ensures that the equilibrium will be inefficient: the only issue is whether it is peaceful (containment) or not (war). Containment is costly, but it lessens the probability that Iraq's nuclear program will be successful to the point that the US no longer expects it to happen, and so renders moot both the US inability to perfectly police Iraq's nuclear efforts and Iraq's commitment problems. In effect, the availability of containment ensures that neither shifting power nor asymmetric information are sufficient to cause war in equilibrium.

Similarly, war is also costly, and also renders irrelevant the inability to police and the commitment problems. If the US wins, then it can govern Iraq itself or install leaders with preferences similar to its own; either way, the dispute is ended and there is no more need to worry about Iraq's pursuit of nuclear weapons. If Iraq wins, then the US is (presumably) rendered unable to interfere further in Iraq's affairs, so that Iraq gets its way on the disputed issues and freely develops nuclear weapons if it so desires.[20]

The only reason that war would be chosen over containment is that the latter would be so costly that going to war would give the players a higher overall value. From the US perspective, going to war and "solving" the Iraq problem once and for all might be preferred to the indefinite continuation of expensive containment. Thus, in the model, war happens due to costly peace. To summarize, equilibrium is costly

---

[20]Note that given the overwhelming superiority of US forces relative to Iraq's, the probability of Iraq's victory might be approximately zero, so that assumptions about what would happen if Iraq won are essentially without loss of generality. It was widely believed that the US would surely win the war, but for a detailed assessment, see Cordesman (2002).

because of the commitment problems and inability to perfectly police, but it is only violent if the best candidate for peace—containment—is too costly.

An Analytic Narrative of the Iraq War

The reason the Iraq War happened is that the costs of containment grew rapidly through the 1990s and into the early 2000s, and were set to escalate further, while the anticipated costs of war steadily declined through the same period. Eventually, the costs of containment came to be perceived as exceeding those of war, and so, as Proposition 2.1 predicts, the US went to war. To substantiate this explanation, I will first describe the components of the costs of containment, and how each changed over time. After doing the same for the anticipated costs of war, I will review the available quantitative estimates of the costs of war and of containment. I will then consider why the war happened in 2003, in particular, and how the war was sold domestically. Finally, I will discuss some other explanations for the Iraq War and the evidence for them.

The Increasing Costs of Containing Iraq

The many costs of the US strategy of containment are described in Table 2.1. The table is divided into six components of the total cost: military, economic, humanitarian, diplomatic, political, and security costs. For each component, the principal costs are listed. To the left of each component is a symbol that indicates whether it generally rose (+), fell (−), or was approximately constant (0) over the course of containment from 1991 to 2003. The table is an original but obvious compilation of

the costs of containment described across many sources, but relies heavily on Pollack
(2002).[21] However, the trends specified for each component of costs deserve further
explanation.

Trend	Component	Specific Costs
+	Military	committing forces to region; operations to enforce sanctions; strikes to compel compliance with inspections; enforcement of no-fly-zones; sporadic mobilizations to respond to Iraqi mobilization; exposure of troops to terrorist attack
+	Economic	trade and investment opportunities lost due to sanctions, including access to oil; loss of market share to defectors from sanctions; contraction of Iraqi economy
+	Humanitarian	impoverishment of Iraqi people; repression of Iraqi people; deterioration of Iraqi civil society; casualties inflicted by strikes
+	Diplomatic	recurring negotiations to maintain sanctions; coordinating responses to lack of compliance
+	Political	unpopularity and perceived illegitimacy of sanctions; appearing to persecute Muslims; corruption of UN bureaucracy by oil-for-food program; deterioration of relations with defectors from sanctions; domestic unpopularity of basing troops in hostile areas; inducement of terrorism by US military presence in holy lands
+	Security	of Iraq's support for terrorism; of foregone opportunity for catalyzing regional liberalization; of potential for terrorist use of Iraqi WMD; of induced vulnerability of Iraq to Iran

Table 2.1: The Costs of Containment of Iraq and Their Trends

First, the military component of the costs of containment had increased over time,
though only modestly. The principal rise was in the exposure of US forces stationed
in the region to terrorist attack: the growth of Al-Qaeda raised the perceived risk,

---

[21]This might seem suspect, given that Pollack was a strong advocate for the war. However, the
actual measures of which the US strategy of containment toward Iraq was composed, as discussed in
the model setup, were a matter of public record and uncontroversial. Given a list of these measures,
most of the specific costs listed in the table would be familiar to, for instance, anyone who regularly
read the international section of the New York Times during the years of containment.

especially after the Khobar Towers attack of 1996 and the bombing of the *USS Cole* in 1998.

Second, the economic costs rose substantially over time. The sanctions imposed on Iraq led to a long, deep depression that reduced its economy to a shadow of its pre-Gulf War strength (Nordhaus, 2002, p. 58). The costs of lost access to Iraq's economy, especially to its oil, increased dramatically from 1999 as the price of oil rose quickly. Moreover, the costs increased for the US because its firms were made to respect the sanctions, while their competitors in Russia, France, and China were given leave to violate the sanctions in exchange for bribes from Iraq under the Oil-for-Food Program.

Third, the root cause of much of the increase in the costs of containment lay in the steadily increasing humanitarian costs of the sanctions. As Iraq's economy contracted, much of the remaining surplus was appropriated by the regime to pay for the military and for the kickbacks necessary to undermine the sanctions through Oil-for-Food. As a result, the Iraqi people suffered increasing deprivation throughout the years of containment. This was universally viewed as undeserved and was widely attributed to the sanctions.

Fourth, the diplomatic costs of containment increased substantially over time as international support for containment declined. Especially as many countries began to respond to concerns over the impoverishment of Iraq caused by sanctions, and as some countries began to cheat on the sanctions, the diplomatic effort and political capital that had to be spent on the part of the US to police and preserve the sanctions and generate support for firm responses to Iraq's provocations rose.

Fifth, and in part as a result of rising humanitarian concerns, the political costs of containment rose dramatically over time. The sanctions, and especially US efforts to preserve them, became increasingly unpopular abroad, and fueled perceptions that the US was persecuting Muslims, especially after the Second Intifada and 9/11. Relations between the US and the defectors from sanctions suffered, especially in the cases of Russia, France, and China. The UN bureaucracy was increasingly undermined and corrupted by the dirty dealings of the Oil-for-Food program. Moreover, basing troops in areas where they were subject to terrorist attack became increasingly unpopular with the US public.

Sixth, and finally, the negative security externalities caused by containment also grew. Iraq increased its support to Palestinian terrorism as a means of bolstering its public image with the region's Muslims. In part because of this, the continued survival of Saddam's regime was viewed by some as a major impediment to Middle East peace. As the region's most domestically unpopular regime, its survival entailed foregoing the best available opportunity to catalyze democratization and liberalization and, some thought, a truly durable peace in the region. Last, after 9/11, prior beliefs about the risk of extreme terrorist attacks were revised upward, and the possibility that whatever WMD Iraq had might somehow fall into the hands of terrorists, whether inadvertently or intentionally, received greater weight.

Overall, the costs of containment rose quite substantially over the period 1991–2003.

The Decreasing Anticipated Costs of War with Iraq

Now consider the anticipated costs of war with Iraq, and how these changed over time. These are described in Table 2.2. The table is divided into the same six components as were used for organizing the costs of containment. For each component, the principal costs are listed; the general trend in each component from 1991 to 2003 appears on the left. The table condenses the highly detailed compilations of the actual and anticipated costs of war collected in Bilmes and Stiglitz (2006); Davis, Murphy and Topel (2006); Nordhaus (2002); Wallsten and Kosec (2005). As with the costs of containment, some explanation of the posited trends in the components of the anticipated costs of war is in order.

First, over the years of containment, the anticipated military costs of war had declined very substantially. Sanctions had drastically constrained the funds available to support Iraq's military, and its strength decayed rapidly. By contrast, the US military became much more capable over the 1990s—the development of high-precision bombing, dramatic improvements in inter-service coordination, and the introduction of modern information technology all radically increased the US military's ability to win wars. Moreover, the development of operational art featuring greater use of special operations forces and much-reduced manpower also promised to greatly reduce the cost of war, as it had in Afghanistan. Finally, the US military's ability to fight effectively under chemical and biological attack also improved substantially (Cordesman, 2002).

Second, the economic costs of war also fell. The same sanctions that strangled Iraq's military also devastated its economy. Oil production during the sanctions

Trend	Component	Specific Costs
−	Military	reduced availability of forces elsewhere; operations to support invasion and occupation; casualties and equipment destroyed in fighting; paid benefits for veterans; more difficult recruitment and retention; use of WMD
−	Economic	sharp rise in oil price; concomitant macroeconomic effects; destruction of civilian infrastructure in fighting; reconstruction of Iraq; increased risk imposed by war
−	Humanitarian	civilian suffering caused by war; potential for internal conflict after war
0	Diplomatic	marshalling international support for war and assistance with reconstruction; negotiating new US-Iraq relationship and status of forces in Iraq
+	Political	increased influence of Iran over Iraq; alienation of supporters of Iraq's regime; potential domestic unrest or punishment of incumbents if war goes badly; perception of attack on Muslims
−	Security	risk of expansion of war to Israel; potential new training ground for terrorists; externalities of post-invasion civil conflict, possibly including partition of Iraq, for surrounding countries

Table 2.2: The Anticipated Costs of War with Iraq and Their Trends

period, even after the initiation of Oil-for-Food, was only around half of production before the Gulf War. War would be less damaging to Iraq's economy because there was less infrastructure to blow up, and less room for economic activity to drop due to the disruption of war (Nordhaus, 2002).

Third, the humanitarian costs of war declined for the same reason as the economic costs. As the Iraqi people become more impoverished due to the sanctions, the room for further deprivation due to war was reduced. Additionally, the increasing military superiority of the US implied a shorter, more surgical invasion which would minimize civilian casualties. Finally, there is no obvious rationale for believing that the potential for conflict within Iraq after the war changed substantially over time.

Fourth, the diplomatic costs did not change much over the years of containment. Those states that bore a larger share of the costs of containment—the UK and some of Iraq's neighbors—became more sympathetic to the case for war and restoring Iraq to working order after a war. However, the states that were making gains by defecting from sanctions became increasingly accommodating toward Iraq and opposed war (Pollack, 2002, Chapter 6 and pp. 352–365).

Fifth, there were competing trends within the political costs of war. First, the growing superiority of US forces meant less chance of international and domestic backlash from a long, destructive invasion. But, after 9/11 and the war in Afghanistan, the risk that a war with Iraq would be seen as further attacks on Muslim society, and thereby bolster support for international terrorism, increased. Also, as Iran recovered from the long, terrible war with Iraq of 1980–88, its ability to exert influence in Iraq, especially if Saddam's regime fell, grew. And as the economic links between the

defectors from sanctions and Iraq grew stronger, the political costs of spurning these states' desire for the peaceful continuation of these links also rose. Overall, it seems plausible to assume that the political costs grew somewhat over time, but especially after 9/11.

Sixth, and finally, there were also competing trends within the security component. A better US military and weaker Iraqi military, especially with reduced missile capabilities, meant less risk of a war expanding to include Israel, or other neighboring countries. It also implied that the US would be more able to contain the externalities of any post-war civil conflict within Iraq. However, a war in Iraq, especially if it were followed by civil conflict, might provide a potential training ground for international terrorists, and this risk grew with the rise of Al-Qaeda. Still, as many observers have pointed out, this is potentially a net benefit, if it occupies terrorists in a location distant from the US and its allies and provides opportunities for their killing or capture by US forces.

Thus, most of the components of the anticipated costs of war with Iraq declined considerably over the period 1991–2003. It seems plausible to assume that, even at the end of this period, the combined fall of the military, economic, humanitarian, and security costs of war overwhelmed the rise in political costs, especially given the Bush (Jr.) administration's demonstrated insensitivity to international political costs.

Comparing the Costs of Containment and War

Immediately after the Gulf War, the Bush (Sr.) administration regarded war as more expensive than containment. That the costs of containment increased and those

of war decreased over the subsequent years of containment does not imply that they ever crossed, so that containment became more expensive than war. So, here I will review the available estimates of the *absolute* costs of containment and of war, to see how they compare.

There are several potential pitfalls in this exercise. First, if the goal is to understand the US decision to go to war, then estimates based on information that became available *after* the war began are essentially irrelevant. Whether the war and subsequent occupation were competently managed, and whether this competence or lack thereof, or the occurrence of sectarian conflict in Iraq after the war, *should* have been predicted beforehand, are important questions, but they are distinct from the question of what *was* predicted. All that matters for this purpose are estimates of the costs of war and containment based on information available *before* the war.

Second, and relatedly, it should be clear from the preceding tables that many of the components, and indeed some of the most important components, of the costs of war and containment are exceedingly hard to measure. Even those which are naturally quantitative (e.g., economic costs) cannot be estimated with any precision. Thus, any attempt to total up the costs of containment and war is fraught with "imponderables" and "uncertainties" (Krueger, 2006).

Presumably for this reason, there is, to my knowledge, no publicly available estimate of the costs of war and containment that addresses all of the components identified in the preceding discussion. The most comprehensive estimates, on which I focus here, study only the military and economic costs in detail, and provide only back-of-the-envelope estimates of the humanitarian and security costs. The diplo-

matic and political costs are simply intangible—though no less real.

The restriction to reasonably comprehensive estimates based on *ex ante* information leaves just two studies. First, Nordhaus (2002) combines existing estimates from government sources and other economists with original calculations, based on information available as of late 2002, to estimate the cost of war. This study has been praised for the soundness of its predictions and its frank treatment of uncertainty by the most comprehensive study of the war's actual (*ex post*) costs, and also by a post-war review of the pre-war estimates (Bilmes and Stiglitz, 2006; Krueger, 2007). It includes military and economic costs and, indirectly, some of the items listed as security costs in Table 2.2.

Second, Davis, Murphy and Topel (2006) (DMT) compiles a wide range of pre-war estimates of the costs of war and containment with original calculations, all based solely on information available before the war, to estimate the costs of a range of scenarios for both containment and war. It is the only study to estimate costs for both war and containment, and its figures include military, economic, humanitarian, and security costs.

The bottom line estimates are as follows. According to Nordhaus, the costs of war are predicted to range from approximately $100 billion to $2 trillion. According to DMT, the costs of war would be predicted to range from $100 to $870 billion, and the costs of containment would be from $300 to $700 billion. Moreover, the latter study also estimates that a war would improve the economic well-being of most Iraqis and their survival chances relative to containment.

There are several points to bear in mind when considering these estimates. First,

though by focusing on the information available just before the war, they take account of the decline in the costs of war over the years of containment, they do not account for the trend in the costs of containment. That is, DMT assumes that the costs of containment in the future will be as they were in the past, averaged over those years, though their estimated range does include some low probabiity, unprecedented events such as the internal overthrow of Saddam. Second, both studies ignore the components of costs that are the most difficult to measure, and yet still produce ranges that vary by an order of magnitude. Moreover, both freely admit that these other components might well be of comparable magnitude to those they do estimate, and hence would increase the uncertainty substantially.

In a review of these estimates and others, Krueger (2006) takes a nihilistic view, observing that "all costs and benefits can be contested as wildly inaccurate—in either direction" and dismissing the exercise as "little more than educated guessing by other means." But whether these studies should be taken seriously or not as estimates of the *true* costs of war and containment doesn't matter for our purposes. The fact that the costs of each option are so wildly uncertain could instead be taken as representative of an underlying, diffuse distribution of beliefs among the expert community about the costs of war and containment.

Interpreted this way, something *can* be learned from these estimates: there was a substantial overlap between the distribution of expert beliefs about the costs of containment and that of beliefs about the costs of war, even if the potential difference in either direction was great. Under these conditions, the beliefs of those in power about which option is better matter most. If these beliefs are strongly held, it is

rational for those in power to use what means are available to convince the public to support their favored option, even if this entails shouting down experts with differing views or exaggerating aspects of their case.[22]

To separate the issue of why the war happened from how the case was made for it, it is helpful to examine why the war happened when it did, which I turn to next.

Why did the War Happen in 2003?

Every administration since the Gulf War eventually concluded that Iraq's regime had to be changed.[23] So, why is it that this was not done until 2003? An easy answer is that 9/11 happened, making the US public receptive to policy-makers' case for a war with Iraq, and war occurred as soon thereafter as the US was militarily prepared to invade. There are good reasons to believe that this was a necessary condition for the war to occur, but I will argue here that it was not sufficient. Two other conditions also had to be met, and the models developed in this section and the next expose the importance of these conditions. First, a third, cheaper option—inciting internal regime change—had to have been exhausted. Second, a large enough portion of the US policy community had to perceive that the costs of containment exceeded the costs of war. I'll elaborate on each of these two conditions, before re-considering the role of 9/11 in light of them.[24]

Containment was not originally designed to be a long-term strategy. At the end

---

[22]To be clear, by "rational" I do not mean to imply "moral" or "responsible." This essay is not intended to address moral concerns with the preparation, conduct, or consequences of the Iraq War.

[23]This fact is documented in Section 4.

[24]The historical material here is drawn from Chapters 2, 3, and 7 of Pollack (2002).

of the Gulf War, the US expected that Saddam's weakened grip on power would lead to his ouster by Iraq's generals and replacement with someone the US could tolerate. When this didn't immediately happen, the US implemented a policy of containment, which was intended to minimize the threat Iraq presented while the US waited for a coup. As part of this strategy, the US continued most of the unprecedentedly harsh sanctions from during the Gulf War and instituted a program of covert action, both intended to encourage the Iraqi opposition to depose Saddam. Unfortunately, Saddam proved rather talented at escaping a litany of coup and assassination attempts and suppressing several popular revolts. By 1996, all of the CIA-supported Iraqi opposition groups had been either broken or ejected from Iraq, and by 1999, further attempts were regarded as highly unlikely to be successful. In Section 4, I will provide a model that explains why, despite its failure, this strategy was reasonable, and document how views on this third option evolved. But suffice it to say that, by 2002, waiting for someone to depose Saddam was seen as a recipe for the indefinite continuation of containment.

While containment was highly effective at keeping Saddam in check, it came with all of the costs described above. Most importantly, it enabled Saddam to paint the sanctions, and their principal supporter the US, as the principal cause of the destitution of the Iraqi people. When, in response to international concern about the humanitarian externalities of containment, the Oil-for-Food program was implemented, Iraq thereby gained an indirect source of hard currency, in the form of a fixed but large quantity of oil contracts that could be allocated to buyers at its discretion. It used these contracts to separate France, China, and Russia from the US and UK,

giving them privileged access to the Iraqi economy and threatening to end this access if they did not support its campaign to end containment.

Once this strategy of driving a wedge between members of the Security Council began to work, it became increasingly difficult for the US to enforce containment, as Iraq's provocations were now met with divided opinion and prevarication on the Council rather than firm consensus. With minor violations going unpunished, the floodgates were opened and many of Iraq's neighbors now began to violate the UN resolutions governing containment with abandon. The trend was clear by 1998, and in late 1999 the US consented to expanding Oil-for-Food to encompass a much broader array of goods that Iraq could now import, essentially recognizing what was already taking place illicitly.

The Bush (Jr.) administration's first major foreign policy review upon entering office centered on Iraq. It decided to pursue the revitalization of containment with "smart sanctions." This policy offered to end the remaining economic sanctions on Iraq in exchange for renewed agreement to enforce restrictions on what goods Iraq was allowed to import, and offered incentives to the various defectors from containment that were intended to restore international consensus on the matter. Two attempts to move this initiative through the Security Council failed, because the incentives the Bush administration offered were not nearly enough to make up for the highly lucrative illicit trade in which many of Iraq's neighbors and supporters were already engaged.[25]

---

[25]A form of smart sanctions was eventually adopted in May 2002, but it excluded the anti-smuggling provisions, which were essential to the original design. Speculatively, if it is true that by that point the Bush administration had decided on war, then there was little point in continuing to fight for a better set of sanctions.

By 2002, it was clear that to continue effective containment, the US would have to take on *all* of the associated costs. Iraq's neighbors would have to be compensated by the US for the value of their illicit trade with Iraq. The US would have to unilaterally compel Iraq's compliance with the elements of containment. Its relations with France, China, Russia, and many of Iraq's neighbors, including US allies, would become more combative. Of course, this would only increase the general unpopularity of containment, as the US would no longer have international backing for its actions. Overall, the military, economic, diplomatic, and political costs of containment would rise dramatically.

Importantly, it is simply not true that it was impossible to restore containment, as some advocates for the war argued. Of course, containment would never be perfect: the US could never be *certain* that Iraq's WMD programs had been dismantled, any more than it had been for the decade of containment up to 2002. And it would cost much more, costs which, in the absence of a good prospect of inciting a successful coup or revolt against Saddam, would potentially have to be borne for a very long time. But it was still feasible. At the very least, the US possessed the military and economic resources necessary to contain Iraq unilaterally. It's just that this course of action would have been exorbitantly costly.

These two conditions—the exploding costs of preserving effective containment and the collapse of the possibilities for inciting internal regime change—were perceived to hold as early as 1999 in the Clinton administration. Why didn't the US invade Iraq then? And if other challenges, such as the Kosovo War, the new round of Middle East peace negotiations, and the Monica Lewinsky scandal diverted the Clinton

administration's attention through the end of its term, why didn't the Bush (Jr.) administration take up war when it entered office?

Certainly, many believed the new administration would do just that. It included, in substantial numbers and at very high levels, officials who had publicly and vehemently advocated a war to change Iraq's regime prior to entering government. And as mentioned, Iraq was the very first issue its foreign policy team considered. So why didn't it invade in 2001? A related question, for which the proposed answer is the same, is: why, in 2003, didn't the Bush administration just continue to kick the can down the road? After all, the US had been putting off serious action on Iraq since at least 1996, when the CIA's client groups in Iraq were defeated by the regime and the institution of Oil-for-Food began the erosion of containment.

The answer is 9/11, or rather the willingness it instilled in the American public to consider drastic action to improve US security. Since at least the years of US nuclear monopoly, when the United States considered war to prevent the Soviet Union from obtaining nuclear weapons, US policy-makers have perceived an unwillingness in the American public to countenance aggressive, unprovoked war to improve US security (Quester, 2000; Silverstone, 2007). According to Pollack, the Clinton administration's consideration of a US invasion of Iraq was always stymied by the belief that the public simply would not support such a war, at least not without grave provocation by Iraq. And even the earliest reported date at which the Bush administration had set a policy of invasion was still months after 9/11.

The Bush administration recognized the window of opportunity that the shock of 9/11 created. In order to gain the public support for doing what many of its members

had long regarded as necessary with respect to Iraq, it needed only to develop a case for the war that would connect the two in the public's mind. There followed the allegations about the possible "nexus" wherein Iraq might transfer WMD to a terrorist organization for use against the US or its allies, the reported politicization of intelligence estimates, the silencing of official voices pointing out the likely high cost of a war and occupation, and the claims that restoring containment was impossible. And, in time, the Iraq War.

Still, even if 9/11 was necessary for the Iraq War to occur, it does not fully explain it. Why did so many members of the Bush administration (and also the former Clinton administration, and Congress, and the independent policy community) see a war as desirable? It was not a matter of the new appreciation of the dangers of international terrorism, because these people had been advocating war well before 9/11. And why was war, rather than revitalized containment (for which the US surely could expect some generosity from the international community in the wake of 9/11) and a renewed commitment to inciting revolt, the chosen solution?

This essay offers answers to these questions. Internal regime chance was no longer regarded as a realistic option, and indefinite containment had grown more expensive than war. 9/11 simply provided the opportunity to overcome the public acceptance of the (eroding) status quo and aversion to the costs and risks of war. I will conclude this section by discussing the implications of this theory for other rationalist accounts of the Iraq War.

Implications for Other Rationalist Accounts of the War

Every other rationalist explanation for the Iraq War of which I am aware has at its core the same drivers of inefficiency as the model I have presented here.[26] First, there is asymmetric information about Iraq's pursuit of WMD (or its possession of WMD, or its intentions with respect to WMD). Second, there are commitment problems in that Iraq cannot commit to not developing WMD or to not using them once acquired. I have argued in this essay that, in general, such explanations for war are not sufficient unless analysts also demonstrate that measures other than war were unavailable, ineffective, or too costly.

This concern may seem superfluous here: after all, I have also argued that in the case of the Iraq War, these other measures (i.e., containment) were indeed too costly. However, ignoring the availability of these measures leads to interpretations of the Iraq War that do not do justice to the facts. Iraq's commitment problems and asymmetric information about its pursuit of WMD were present from the beginning of containment, and so cannot explain why the war did not happen in the 1990s. One could use the argument of the previous section, that 9/11 created a window of opportunity for war, to justify the occurrence of war a decade later, but there were many more propitious windows earlier on. Iraq was caught red-handed with evidence of hidden WMD—including nuclear weapons infrastructure—more than once in the early 1990s. In the same period, the US repeatedly assembled broad international support for strikes against Iraq to enforce its compliance with containment. Why, when it had incontrovertible evidence of Iraq's perfidy and broad international support

---

[26]I know of only three: Baliga and Sjöström (2008); Debs and Monteiro (2010); Lake (2010).

for aggressive action, didn't the US invade Iraq then?

One answer for why the war did not occur until after 9/11 is the so-called "one percent doctrine," promulgated by Vice President Cheney. This held that even a miniscule probability that a terrorist organization might obtain a weapon of mass destruction from Iraq and then use it against the US or its allies should be treated as though it were a certainty, because the consequences would be so extreme. There are a number of practical problems with this argument, including the lack of any reason to believe Saddam would risk such a thing, with which many members of the Bush administration would have been familiar. But there is also a deeper one: why is war necessary to eliminate this possibility?

More generally, it is not clear how asymmetric information about Iraq's pursuit or possession of WMD could lead to war. The development and stockpiling of biological, chemical, or nuclear weapons is not like resolve or the precise balance of military capability, parameters bargaining theorists often take to be asymmetrically known. It might well be impossible for one side to credibly convey its strength or resolve without fighting, because these things are inherently difficult to observe. But WMD and the physical capital and scientists that produce them are physical objects. Given the opportunity to look, it is always possible to see if they are present in a particular place. And it is possible for Iraq to allow the US to look freely for them, anywhere and anytime it likes, for as long as it likes, with as many inspectors and as much equipment as it likes.

It was thus in no sense impossible for the US to become approximately certain that Iraq had no weapons of mass destruction and was not actively pursuing them.

In fact, it did just that in the aftermath of the war. The only thing preventing the US from achieving this certainty was Iraq's refusal to comply with the inspections. However, if there's one thing on which the historical record is clear, it is that Saddam would do anything to preserve his regime. If he became convinced that unless total cooperation with inspections was forthcoming, the US would invade, then he would cooperate. Faced with the choice of a war he was almost certain to lose, and fully revealing and losing his WMD but having a chance to survive, he would surely take the latter.

Of course, Saddam never fully cooperated with the inspectors. He didn't need to: he only needed to give the US sufficient confidence in the rudimentary state of his programs to avoid an invasion. And, as we now know, even though he actually had virtually nothing to hide, there were other values for him in maintaining some ambiguity. Some have proposed this as an explanation for Saddam's refusal to do what was necessary to convince the US that he had abandoned his programs. That is, it is argued that Saddam was willing to run a risk of war in order to maintain the WMD ambiguity necessary to deter Iran and his internal enemies. But risk of war is not certainty of war. Faced with the choice between not cooperating and certainly losing a war with the US, and cooperating and possibly having to fight a much weaker enemy such as Iran or Iraqi rebels, Saddam's preference is still clear.

Thus, asymmetric information about the possession or pursuit of WMD could not cause war, because it could and would be dispelled at Iraq's behest whenever it felt the US threat of war was sufficiently credible. Asymmetric information about preferences is different, since these are not directly observable, but it is hard to believe that after

more than a decade of constant bargaining and limited fighting, either the US or Iraq had any substantial uncertainty about the other's preferences. Hardly anyone in the US doubted that Iraq wanted to get WMD, or that it would accelerate its efforts if containment ended. The evidence from after the war suggests this belief was correct (Iraq Survey Group, 2004).

Instead, the US was perfectly capable of, if not verifiably ending Iraq's WMD programs, then at least assuring itself that they were like to be very small and, in the case of nuclear weapons, unlikely to succeed any time soon. The problem was that this assurance came at a cost: containment was not cheap. By 2002, after four years of rapidly eroding sanctions and without inspections, the CIA was uncertain about precisely what Iraq was up to. But the US rectified the situation by sending a costly signal that it was serious about invading when it commenced a large buildup of military forces in the region in preparation for the war. Naturally, Saddam reacted by letting the inspectors back in, and Iraq cooperated more fully than it had in years.

The Bush administration was charged with ignoring the positive results of the inspections, with making unreasonable demands of Iraq, and with pressuring the intelligence community to produce more threatening estimates. These behaviors are all mysterious if the root of the problem was present uncertainty about Iraq's WMD programs. If the problem is lack of information, why wouldn't the results on inspections and the unvarnished estimates of the CIA be ameliorative? The answer is that both were actually irrelevant to the decision to go to war. In fact the problem was not asymmetric information, it was that the administration had already decided to stop paying the costs of containment and to go to war instead. Even if the inspectors

had reported with confidence that Iraq had verifiably disarmed, and the intelligence had confirmed this, the true rationale for war would be unchallenged.

That said, every piece of evidence of Iraq's continued clandestine WMD programs and support for terrorism, no matter how circumstantial or suspect, would help to build support for the Iraq War among the public and the international community. But this evidence was solely useful for the sales pitch—and this explains why the administration would ignore the unfavorable (to its case for war) inspection reports and try to exaggerate the CIA's estimates.

The bottom line is that the facts are not consistent with accounts based solely on asymmetric information and shifting power (due to the future acquisition of nuclear weapons). Under containment, the asymmetry in information was quite limited, and the expected shift in power quite small, because the suspected small stock of chemical and/or biological weapons would not seriously shift the balance of power, and nuclear weapons were still a long ways off. Instead, the war occurred because maintaining this status quo had become more expensive than war.

The evidence presented here is by no means conclusive. But it does strongly suggest that an account of the Iraq War based on costly peace is more compelling than the existing accounts based on asymmetric information and shifting power. Thus, the costs of arming are demonstrably relevant to understanding the origins of at least one historical war.

2.4 Imposition and Civil Conflict in Post-Gulf War Iraq

Before the US went to war against Iraq in 2003, it tried an alternative method of avoiding the costs of containment: inciting a revolt to change Iraq's regime from within. Its approach was to impose sanctions on Iraq that were implicitly, and later explicitly, intended to encourage regime change. It is clear enough intuitively why the US might want to incite civil conflict to overthrow Saddam, but it is not immediately obvious why imposing sanctions would accomplish this. Why can't Saddam and the (potential) rebels come to a compromise that avoids the destruction of civil war or a coup? The answer, as I will demonstrate in this section, is that sanctions make internal peace between Saddam and the rebels costly; and if this peace is costlier than revolt, then a revolt will occur.

More abstractly, the second empirically common source of costs in peace is *imposition*: penalties imposed or rewards offered by outside actors that depend on the outcome of an interaction between two "inside" actors. These outside actors may be unwilling or unable to fight to affect this outcome, but they can still influence the choices made by the inside actors by making demands of them and offering incentives to meet these demands. If meeting these demands favors one inside actor but disfavors the other, the latter may refuse to do so, and peace will then entail the loss of the incentives offered by the outside actor. If these incentives are large enough, peace will be very expensive, and war will result as the disfavored side tries to hold its ground and the favored one tries to capture the outside incentives.

I am aware of only one model of imposition in the existing literature. Powell (2011) models a negotiation between two actors within a country over the balance of power between them. State consolidation, defined as the achievement of a near-monopoly of power, results in outside rewards that increase the value of the game. Consolidation can be achieved peacefully, but takes time and so delays the receipt of the rewards, making peace costly. If the rewards are large enough, then peaceful consolidation is more costly than war (i.e., violent consolidation), and war occurs. Empirically, the rewards are taken to be investment by foreign firms that is contingent on a sufficiently secure state. However, these outside actors are not modeled, and it is unclear why there is no deal they could work out with the inside actors that would enable investment to be made and war to be avoided, even in the absence of consolidation. Moreover, no empirical example of a war due to imposition is presented, so it is unclear whether this explanation is actually relevant to explaining historical conflict.

Here, I will present a model of war due to the costs of imposition that is tailored to the situation *within* Iraq between the two wars with the United States. For tractability, this model ignores many of the aspects of the US-Iraq interaction considered in the previous section, but it endogenizes the design of sanctions on Iraq and the interaction between Saddam Hussein and an organized opposition (the inside actors) over how to respond to the sanctions. This approach has two advantages relative to that taken in Powell (2011). First, it illuminates not only why sanctions might cause civil conflict, but also why the US (the outside actor) might choose to incite this conflict, and thus why there is no acceptable deal among the outside and inside actors that

would avoid it. Second, the model provides an explanation for the civil conflicts that took place in Iraq between the two wars with the US and for the role the US played in these conflicts. It thus shows that war due to the costs of imposition is empirically relevant.

After describing the setup of the model, I will state and provide intuition for the two results that characterize equilibrium. The section closes with a discussion of the aforementioned civil conflicts in Iraq.

A Model of Imposition and Civil Conflict

Suppose that Iraq is composed of two "internal" actors, labeled S for Saddam's regime and R for the opposition (rebels), who must bargain over control of the polity. The actors have opposed interests over, for example, whether Iraq is democratic or not, the degree of state control of the economy, minority rights and autonomy, the country's alignment with foreign nations, whether to pursue WMD, and perhaps other issues. These issues are represented by the unit interval $[0, 1]$; assume that Saddam favors outcomes closer to 1, the opposition those closer to 0, and that both are risk-neutral over outcomes in the interval, so that a settlement $q \in [0, 1]$ yields payoffs $u_S = q$, $u_R = 1 - q$ for the respective players.

The US also has an interest in these issues. Suppose the US is risk-neutral and favors outcomes closer to 0, just like Iraq's opposition, so that a settlement q gives the US a payoff of $u_{US} = 1 - q$. Note that the outcome 1 does not represent the ideal set of policies for Saddam's regime—rather, it is the set of policies the regime would implement in the absence of any internal opposition, but still taking account

of the need to appease the US. By contrast, 0 represents the policies the opposition would implement in Saddam's absence, and I assume that these are also the policies most favored by the US. Relaxing this assumption and allowing for there to be small differences between US and opposition preferences would not qualitatively alter the results presented here. Moreover, it seems reasonable given that the US took care to support only elements of the opposition that favored its interests (Pollack, 2002, Chapter 3). This support is not modeled explicitly, but to the extent that it is costly to the US and raises the rebels' chance of victory, incorporating it would strengthen the results.

The game described here is assumed to take place with the US already engaging in containment, having decided that this was preferable to a war with Iraq.[27] The game begins with the US choosing a demand to make and a sanction to impose if the demand is not met. The choice specifies the severity of the sanction $\alpha \geq 0$, and a target outcome $q_\alpha \in [0, 1]$. That is, if the internal actors do not implement a settlement in $[0, q_\alpha]$, the US will impose sanctions that cost Iraq α. Notice that this is equivalent, from the point of view of the internal actors, to a reward of size α in exchange for implementing an outcome no larger than q_α. Such a reward might include official development assistance, private direct investment, a security guarantee, access to foreign markets, diplomatic rehabilitation, and so on; the sanctions could entail the denial of any of these.

Imposing sanctions is costly for the US. The cost depends only on the severity of the sanction, and is given by the function $s(\alpha)$. It is assumed that $s(\alpha) > 0$,

---

[27]It would also suffice for the results if the US instead preferred war to containment, but also preferred inciting a revolt to war.

$s'(\alpha) > 0$, and $s''(\alpha) > 0$ for $\alpha > 0$, and that $s(0) = 0$, $s'(0) = 0$, and $s'(\alpha) \geq 1$ for high enough α. These ensure that: imposing non-zero sanctions is costly; the cost to the US of sanctions rises in their severity; the US implements the most cost-effective sanctions first; the first (tiny) measure of sanctions is effectively costless; and the marginal cost to the US of more severe sanctions eventually exceeds the marginal increase in their severity.[28] These costs are described in more detail in the previous section, but to remind, they include economic costs (e.g., reduced international access to Iraq's economy), humanitarian costs (e.g., the impoverishment of Iraq's people), and diplomatic/political costs (e.g., the costs of maintaining international support for the sanctions, or of losing it and being regarded as acting illegitimately).

Since imposing sanctions is costly for the US, its threat to do so if its demand goes unmet may not be credible. Because the internal actors will not suffer the costs of sanctions if they wait until sanctions are actually imposed but then immediately agree to the US demand, they have no incentive to agree to it beforehand. Thus, if the US wants to exert influence over their choice of policies, it must actually impose the sanctions and pay the cost of doing so up front, before the internal actors make their decisions. So, I assume that upon making its demand and threat of sanctions, the US immediately implements those sanctions and pays the cost of doing so.

Once the US has specified its demand and sanctions, Saddam can either start an internal war or make an offer of a settlement of the contested policies to the

---

[28]The formulation given here says nothing about the possibility that there is a limit on the severity of the sanctions the US is capable of imposing, regardless of cost. In principle, the most the US could possibly do is to completely close Iraq to outside commerce of any kind. However, this possibility can be incorporated by simply assuming that the costs of sanctions become exorbitant above some upper bound, so that the formulation used here is completely general.

opposition, which the rebels can either accept, in which case it is implemented, or reject, in which case there is war. War is a costly lottery that Saddam wins with probability p, the opposition with probability $1 - p$, and which costs the internal actors $d_S, d_R > 0$ respectively. The winner is assumed able to implement his choice of policies within $[0, 1]$, while the loser gets nothing. It is important to note here that the structure of the game rules out the possibility that Saddam's regime might start a war with the US in order to stop the sanctions. This is consistent with the fourth calibrating assumption given in the previous section, which ensures that Iraq's value under containment is equal to its war value even given that the US is imposing quite severe sanctions.

So long as Saddam's regime remains in power, the US pays a cost c_{US}, which is the total cost of all the components of containment described in the previous section, except the costs of the sanctions. If the rebels take over, the US no longer needs to pay the cost of containment since the rebels share its interests, but a war within Iraq imposes a cost of $d_{US} > 0$ on the US, taken to be the humanitarian and diplomatic/political costs to the US of the suffering and destruction of a war it had induced.

All parameters of the game are assumed to be common knowledge, and the game is solved for subgame-perfect equilibria.

Analysis of the Model

The analysis of the model is designed to answer two questions. First, under what conditions is the US *able* to use the imposition of sanctions to incite civil conflict

within Iraq? Second, when would the US actually choose to do so? I will state a lemma and a proposition that answer these questions and discuss the intuition behind these answers; proofs are in the appendix.

Lemma 2.1. *If meeting the US demand is worse for Saddam than fighting an internal war, and the cost of sanctions for Iraq, weighted by the chance of ending them in war, exceeds the costs of war for Iraq, then war will occur. If meeting the US demand is better for Saddam than fighting, or the weighted cost of sanctions is less than the cost of war, there will be internal peace.*

The lemma answers the first question: the US would *always* be able to incite a revolt. It need only make a stringent enough demand and impose severe enough sanctions. If the demand is stringent enough, then Saddam would never voluntarily choose to meet it: he would spurn the demand even if it meant suffering a revolt. This means that if Saddam is left in power, the US demand will not be met and Iraq will have to endure the sanctions. Thus, the sanctions make peace between Saddam and the rebels costly. War is also costly, but it offers the possibility of ending the sanctions, should the rebels win, and so reduces the expected cost of peace. If the sanctions are severe enough, so that the costs of peace are higher than those of war, then the value left by the sanctions to divide up between Saddam and the rebels will not suffice to appease both, and war will occur.

Peace is costly because of a commitment problem identified by Powell (2006). In theory, Saddam and the rebels could eliminate the costs of peace relative to war by agreeing to set policies according to a coin flip weighted to reflect the internal balance of power. If Saddam won the toss, he would set his ideal policies and Iraq

would endure the sanctions. If the rebels won the toss, they would set their ideal policies and the sanctions would end, just as if they had successfully revolted. The expected costs from sanctions would then be the same in peace as in war, and thus there would be no need for war. Both Saddam and the rebels would commit to this coin flip if they could. But they cannot so commit: if the coin flip came out against Saddam, he could simply demand another toss, and the same is true for the rebels.

Proposition 2.2. *The US will incite a war within Iraq if and only if the cost of containment is high enough. Incitement becomes more tempting as the costs of war decrease, the chance of rebel success increases, and sanctions get cheaper.*

The proposition answers the second question: the US would choose to incite a war within Iraq when the cost of containment is too high. This is the same reason it would fight a war against Iraq in model of the previous section. There, the US was willing to pay the costs of a war with Iraq whenever these were less than the costs of containment. Here, the US is willing to pay the costs of *inciting* a war *within* Iraq—the cost of imposing severe enough sanctions, and the costs for the US of an Iraqi civil war—when these are sufficiently low relative to the cost of containment. When the costs of a civil war are lower, it is cheaper for the US to incite one, and when the rebels' chance of success is higher, an incited war is more likely to free the US of the cost of containment: both these increase the incentives to incite a war.

With the analysis in hand, we can turn to explaining the historical instances of rebellion that Iraq suffered after the Gulf War.

Sanctions and Revolt in Iraq

In what follows, I offer an analytic narrative of the civil conflicts within Iraq following the (first) Gulf War, guided by the model presented in this section. This model clarifies the origins of several, though not all, of these conflicts, and explains the US role in these conflicts and how it evolved over time. It also substantiates the claim made in Section 3, that the absence of a viable third option involving internal overthrow of Saddam partly explains why the Iraq War happened in 2003 rather than earlier. I rely heavily here on the more conventional historical narrative of the Gulf War and period of containment given in chapters 2 and 3 of Pollack (2002).[29]

To begin, consider the problem faced by the United States at what became the end of the Gulf War. The Bush (Sr.) administration believed that removing Saddam and occupying Iraq would be extremely costly. Moreover, these costs might be avoided if someone else removed Saddam instead. The widely-held perception that the Gulf War defeat had seriously weakened Saddam's hold on power led the administration to expect that Iraq's generals would overthrow Saddam and replace him with someone more amenable to the US (Bush and Scowcroft, 1999, p. 488). If this happened, the US could then negotiate a new and more favorable relationship with Iraq. So, the administration actively encouraged it, with speeches broadcast and leaflets dropped into Iraq that encouraged the Iraqi military and people to rise up against Saddam (p. 48) and bring containment to a close.

Unfortunately, the post-war revolt against Saddam was launched not by the generals, but by the Shi'ites of southern Iraq, followed immediately by the Kurds of

---

[29]In this subsection, I will cite this book only with page numbers for the sake of brevity.

northern Iraq. The Bush administration feared that the success of this revolt would lead to a breakup of Iraq, the same fear that contributed to its decision to halt the Gulf War invasion short of Baghdad (Bush and Scowcroft, 1999, pp. 488–489). It also worried that the revolt would cause the Sunni elites at the center of power in Iraq to re-unite behind Saddam, as indeed it did, and thus reduce the chances of a military coup (pp. 48–49).

In terms of the model, we can think of the immediate post-war revolt as involving rebels whose preferences were perceived to differ substantially from those of the US. Though this was not modeled explicitly, it is equivalent to assuming that, should the rebels succeed, the policies they would then implement were distant from those preferred by the US. So long as the US thought that another set of rebels with interests more amenable to its own (e.g. a military conspiracy that would staunchly favor Iraq's unity) would be more likely to succeed, the success of the current revolt would make the US worse off. This explains why the US refrained from supporting it, even to the point of allowing Saddam the use of helicopter gunships to crush it. Pollack explains the lack of US support for this revolt in much the same way (p. 49).

When it became clear that Iraq's military had rallied behind Saddam to crush the revolt, the Bush administration proceeded to implement containment, with all its attendant costs. For all the reasons described in Section 3, containment would protect US and allied interests from the threats Saddam posed while he remained in power. Moreover, and most importantly, it was expected to do so at a lower cost than full-on war to remove Saddam, the same reason the policy was reaffirmed later on by the incoming Clinton administration (pp. 65–66).

Importantly, the Bush administration's calculation of the cost of containment was predicated on the assumption that containment would not be needed for long (pp. 47–49, 52–53, 55). Senior officials still believed that it was only a short matter of time before a coup was attempted, and that it would be likely to succeed (p. 53, 55). The higher the probability of a successful coup, the lower the expected costs of containment, which need only be endured if the coup fails.

Why did the administration believe that another attempt at regime change was imminent? After all, Saddam had united the military and successfully put down the Shi'ite/Kurdish revolt, and was undoubtedly working to reduce the chances of such an attempt. The model explains why: containment included very severe sanctions, along with a stringent set of demands Iraq had to meet in order to end them. Obviously, Saddam would never voluntarily meet those demands, which included making restitution to countries harmed by the war (e.g., returning billions in stolen property to Kuwait and paying reparations) and abandoning Iraq's WMD and missile programs and support for terrorism. Thus, the sanctions gave any more amenable elites within Iraq enormous incentives to remove Saddam by force. In this way, sanctions make peace within Iraq costlier than revolt, and as Lemma 2.1 shows, revolt should follow.

The Bush administration well understood this, and did not even bother to uphold the pretense that the sanctions were about changing Iraq's policies, rather than its regime (p. 58). Moreover, the US was not about to sit back and hope to get lucky. As early as May 1991, a covert action campaign to "create the conditions for the removal of Saddam Hussein from power" was begun, and the CIA was given carte blanche to accomplish this (p. 59). In practice, this entailed organizing, arming, and training

Iraqi groups that were favorable to the US and opposed Saddam. The rationale for this support to the rebels is clear. If a revolt's success means no longer having to pay the costs of containment or war, then it is entirely in the US interest to take measures to increase its likelihood.

In terms of the model, the conditions were ideal for the US willingness to incite a revolt. First, to the extent that undermining Saddam's military and WMD programs is another benefit of sanctions, the net price of applying sanctions severe enough to incite revolt will be cheaper. Second, the costs of internal conflict in Iraq were declining over time, as Iraq's economy sank. And third, the US was doing its best to raise the chances of success for those groups it supported. According to Proposition 2.2, these three factors should all have contributed to the US willingness to impose sanctions and incite such revolts.

For its part, the first Clinton administration was initially wary of the risks of covert action to unseat Saddam, though it soon accepted them. But it never questioned the value of the sanctions on Iraq. These played a crucial role in undermining Saddam's military power and particularly his WMD programs, but they also provided the principal incentives for revolt. The partial relaxation of the sanctions agreed in May 1996 under the Oil-for-Food program simply traded off some of the costs of containment—it was necessary to counter growing international discontent with Iraq's impoverishment (p. 74).

In fact, no less than eight serious attempts to unseat Saddam were made during the years of containment, and the CIA was only known to be involved with two of

them.[30] Moreover, at least four episodes of open civil conflict occurred, and these were understood to be caused by Iraq's worsening economic situation.[31]

The problem is that all eight attempts failed. In each and every case, Saddam managed to dodge the assassination attempt (though there were close calls), detect the coup early on and foil it (though several involved remarkably high-level conspiracy), and crush the revolt. Each time, the US had to revise downward its estimate of the likelihood of a successful revolt in the foreseeable future.

However, after 1996, when the international consensus on containment was beginning to show serious fractures, the steady-state costs of containment rose rapidly (p. 82–91). Thus, as Proposition 2.2 predicts, the US continued to support the opposition even as the evidence mounted that its chances of success were low (p. 96). As late as September 1998, two years after Saddam had broken or ejected all of the CIA-supported groups within Iraq, the US Congress passed the Iraq Liberation Act calling explicitly for Saddam's overthrow and providing the unprecedented sum of almost $100 million for that purpose. And in 1999, the Clinton administration began trying to resuscitate and re-organize the Iraqi opposition (p. 97).

In time, events such as the Kosovo War, a renewed attempt at Middle East peace negotiations, and the upcoming US election intervened. Given the exceedingly low chances of success, the outgoing Clinton administration shelved aggressive plans for pursuing internal regime change, and the Iraqi opposition fell apart (p. 98–100, 102).

---

[30]There were coup, revolt, or assassination attempts in May 1991 (p. 59), June 1992 (p. 59), December 1993 (p. 68), January 1994 (p. 68), March 1995 (p. 72), May 1995 (p. 75), June 1996 (p. 80), and January 1999 (p. 93). The CIA was involved with the associated groups in the cases of March 1995 and June 1996.

[31]In 1993 (p. 68), 1994 (p. 68), 1995 (p. 75), and 1999 (p. 93).

After eight years spent trying to unseat Saddam, and little in the way of competent opposition left to support, the US gave up on inciting regime change in Iraq. Once it had done so, paying the costs of the still-high severity of sanctions no longer made sense, so the US agreed to the further relaxation of economic sanctions known informally as "Oil-for-Stuff," though it was determined to maintain military sanctions and to demand a new inspections regime in order to preserve the vise grip sanctions had provided on Saddam's WMD programs.

Thus, by the time the Bush (Jr.) administration entered office, there was no option other than indefinite containment or war. By 2002, the Director of Central Intelligence reportedly estimated the chance of success of a new program to overthrow Saddam at 10 to 20 percent (p. 290). Given the disarray of the opposition after 2000, and the historical record of the CIA's attempts with well-organized, well-trained groups, is is hard to believe that this was not a severe overestimate.

The model presented here thus explains why the US would maintain severe sanctions over the course of containment, despite their costs, and why it was (at least while a competent opposition existed) reasonable to expect that this would eventually lead to Saddam's overthrow. Indeed, many serious attempts were made, so in some sense the policy was successful. Though its estimate of the chances of success in this endeavor steadily ratcheted downward in the face of many failed attempts, as the costs of containment rose, it became willing to bet on long shots. Eventually, the plausible avenues for internal regime change were exhausted, and the costly sanctions most essential to encouraging revolt were dropped. By the end of containment, the only remaining option for being rid of Saddam was a US invasion.

2.5 Predation and the American War of Independence

The third and final source of costs in peace that I will analyze is *predation*: the expropriation of value from one actor by another. This expropriation may be costly because it undermines the victim's incentive to produce value. This idea is a very old one in economics, and is central to the economic analysis of taxation, growth, and many other phenomena. And there is a long line of research following Hirshleifer (1991) that explores the relationship among predation, inefficiency, and conflict. However, to my knowledge, the model presented below is the first to make explicit the connection between the costs of predation and the choice to fight a war, as opposed to peaceful (but potentially coercive) bargaining.

Predation occurs when one actor (the "producer") has a comparative advantage in production, while the other (the "predator") has a comparative advantage in coercion. That is, the producer is good at creating wealth while the predator is good at fighting wars. The predator can use his strength to expropriate, under threat of violence, some of the producer's wealth. The problem with this taxation is that it lessens the value the producer gets from his hard work, and so discourages production, or at least production that the predator can steal. This reduces the value that is available for either actor to consume, and so makes peace costly. If instead the two fought, then the producer would have a chance to end the predator's taxation, and the predator would have a chance to directly control the producer's effort. If the costs of peace exceed those of war, then they will fight.

I will explore this source of costly peace with a model that is tailored to the interaction between the colonies that later became the United States (henceforth, the "Colonists") and Britain, in the period following the Seven Years' War. This model is different from most bargaining models of war in that it incorporates an economy that is slightly more realistic than the usual "pie of size 1." Most importantly, what the two actors bargain or fight over is endogenously determined, rather than being fixed by assumption. While this complicates the analysis somewhat, it does seem a promising approach for developing more realistic bargaining models that can speak to the vast literature on the relationship between economic phenomena and war.

I then use the model to provide an account of the American War of Independence (henceforth, the "war" or "Revolution"). The Seven Years War, which eliminated the common interest the Colonists and Britain had in cooperating to oppose France, exposed conflicting interests over the governance of the colonies that ultimately boiled down to how to divide the fruit of the Colonists' labor. Britain attempted to impose a series of unprecedented taxes, and when the Colonists resisted these taxes, Britain responded by initiating efforts to increase its control over colonial governments. The Colonists came to believe that peace would mean tolerating escalating taxation and ceding more direct control to Britain. Because the colonial economy was particularly responsive to taxation, peace would be very costly, and so the costs of peace came to be viewed as exceeding those of war and the Revolution occurred.

This account draws very heavily on Rackove, Rutten and Weingast (2000) and de Figueiredo Jr., Rackove and Weingast (2006), which to my knowledge constitute the only other rationalist account of the Revolution. These companion papers (hence-

forth "FFRW") report the results of a study that did much of the hard work involved in formulating a rationalist account of the Revolution, especially in developing a rationalist interpretation of the deep role that ideas about governance played in the conflict and in providing the evidence to support this interpretation. In fact, I have little to add to their understanding of the issues in contention between the belligerents.

What I can explain, using the model analyzed here and the evidence FRRW gathered, is why war happened rather than coercive bargaining over the future tax treatment of the colonies. This is a question that cannot be answered with the model in de Figueiredo Jr., Rackove and Weingast (2006), because that model does not allow bargaining—the Colonists must either accede fully to Britain's demands or fight. This formulation rules out the possibility that the Colonists and Britain could compromise on some division of policy-setting power between the two. By contrast, the model I present allows for such compromises but explains why they would be abandoned for war.

I'll begin with the setup of the model, then state the results and discuss the intuition, and then provide an account of the war.

A Model of Predation

Suppose we have two actors, Britain (indexed by B) and the Colonists (indexed by C). These actors are embedded in an economy in which there are two kinds of activity. "Transferable" activity produces things that are rivalrous in consumption and can be transferred between the two actors. That is, each product can be consumed by either, but not both, actors. Transferable products include most goods and services traded

in licit markets. In contrast, "non-transferable" activity produces things which are rivalrous in consumption but which cannot be transferred between the two actors. In particular, only the actor that made some non-transferable product can consume it. For example, the standard non-transferable activity in economics is leisure: its production requires some allotment of a worker's resources (e.g., time), and once produced it cannot be moved to another actor. But goods produced and sold outside of the reach of the tax authorities, such as those trafficked locally on the black market, or in foreign trade via smuggling, or outside of authority's remit on the frontier, are also non-transferable.

The economy has a total endowment of resources $R \in \mathbb{R}^+$, which can be allocated to either transferable or non-transferable activity. Assume that the Colonists control the entire endowment R; in other words, C is the producer. This means that whatever Britain consumes must be transferred from the Colonists, and hence can only be drawn from the Colonists' transferable activity. In other words, B is the predator. Of course in reality Britain had productive resources of its own, but these were not under contention historically, so I do not include them in the model. Note that it would make no qualitative difference to the results if we instead assumed that the endowment was initially partitioned between the two players, but it would unnecessarily complicate the exposition of the model.

While Britain derives utility solely from transferable activity, the Colonists derive utility from both transferable and non-transferable products. For tractability we will assume that the Colonists' utility function is additively separable: it is the sum of the utilities derived from each kind of product. Since the two players will bargain

over the disposition of the transferable products, we assume for simplicity that their utilities are risk-neutral over these. These assumptions allow us to write their utilities as:

$$u_C(r, \tau) \equiv (1 - \tau)(R - r) + l(r) \tag{2.4}$$

$$u_B(r, \tau) \equiv \tau(R - r) \tag{2.5}$$

Here, $\tau \in [0, 1]$ is the proportion of transferable products that the predator consumes, r is the amount of resources devoted to non-transferable production, $R - r$ is the amount given to transferable production, and $l(\cdot)$ is the utility of consuming non-transferable products. We assume that $l(r) > 0$, $l'(r) > 0$, $l''(r) < 0$, and $l'''(r) > 0$, for $r \in (0, R)$, and $l(0) = 0$, $l'(0) = 1$, and $l'(R) = 0$. These are standard assumptions from economic theory that ensure that: C likes consuming non-transferable products; more consumption is better than less, but there are diminishing returns; C has decreasing absolute risk aversion; and C's optimal allocation between transferable and non-transferable production will always involve producing at least a little of both whenever the tax rate is neither 0 nor 1.

The sequence of moves is as follows. First, the Colonists make an offer to Britain or start a war.[32] If C makes an offer, B either accepts this offer or starts a war. If an offer is accepted, the producer chooses an allocation of resources between transferable and non-transferable production, production occurs, the agreed transfer is made, all products are consumed, and the game ends.

---

[32]It might be more realistic to assume that Britain was making offers to the Colonists, or that they were exchanging offers. However, this bargaining protocol simplifies the exposition, and the results do not depend on it, as whether war occurs does not depend on the allocation of the bargaining surplus.

The Colonists' offer is to transfer a certain fraction ($\tau \in [0,1]$) of whatever transferable products are made to Britain. In economic terms, C has to choose a linear tax on transferable products, the proceeds from which will be given to B. This is an important assumption. It rules out non-linear tax schemes, in which the marginal rate of tax might depend on how much is produced. Allowing non-linear taxes that vary in the amount produced would not change the qualitative results, which require only that the tax scheme distorts C's allocation between transferable and non-transferable products and hence introduces inefficiency (i.e., costs) into peace, something that any such scheme would do. I discard such taxes for mathematical convenience.

However, a non-linear tax scheme that does not vary in the amount produced—a lump-sum tax—would not distort the producer's allocation and so would not make peace costly. The problem with such taxes, as economists well know, is that they create moral hazard. The tax collector (here, the predator) wants to assign taxes to an activity that represent some fraction of its full realizable value, but it often will not know what that value is, even in aggregate. If the participants in that activity anticipate that a lump-sum tax will be assessed, they have an incentive to realize less than its full value, so as to reduce the amount of tax that will be assigned. This creates inefficient distortion just as in the case of taxes that do vary in the amount produced. In principle, the tax collector could permanently assign the tax at the beginning, and promise not to alter it. If the participants believed this promise, there would be no more distortions and the full value would be realized. The problem here is that if this full value were above the tax collector's estimate, he would have an incentive to renege on his promise and raise the tax, which is precisely why participants wouldn't

believe that promise and the distortions would continue. In other words, lump-sum taxes would end up varying with the amount produced in practice. I discard them and focus on linear taxes in favor of the simplicity of the latter, but the main result derived from a model in which lump-sum taxes and moral hazard were included would be qualitatively the same as the one given here. All that matters is how costly the resulting distortions are.

War is modeled as a costly lottery, won by the Colonists with probability p and Britain with $1 - p$, with costs $d_C, d_B > 0$ regardless of outcome. If the Colonists win, then there will be no transfers of any kind to Britain. If Britain wins, then it is assumed to take direct control over the Colonists, their resources, and the tax rate, at some cost $s \geq 0$, which is the cost of exercising this control. I assume that $s \ll R$, so that control doesn't eliminate a large portion of the total value. In effect, the Colonists fight to be free of Britain's predation, and Britain fights to control the colonial economy.

In this formulation, the only thing the two players bargain over is the tax rate. We have presumed that it is not possible for Britain to take control over any portion of the Colonists' allocation of resources, except by victory in war. In theory, rather than imposing a tax, Britain could enslave some fraction of the colonial population, expropriate a portion of their capital, and occupy part of their land. While it would thereby eliminate the economic distortions that arise from a tax on *products*, this predation on the *factors* of production generates other distortions, in decisions to invest in and expand these factors over time. A one-time transfer of a portion of the Colonists' factors to Britain won't satisfy the latter, because as the Colonists

create new resources, Britain would also demand a share of these. But this is equivalent to what I model here, in that Britain imposes a tax (now on factors instead of products) which distorts incentives for creating value. Whether we are talking about the disposition of transferable products or that of productive factors, the point is that predation can distort incentives for production or investment and thereby make peace costly. War solves this problem, because if the Colonists win, there is no further predation and thus no distortion, and if Britain wins, it gains control of *all* of the Colonists' economic activity and is able to direct both the creation and employment of resources, at a cost.

Furthermore, we assume that the Colonists would not commit themselves to any allocation of resources, other than the one that is optimal under the agreed tax rate. Essentially, the only way the colonial governments can force their constituents to choose an allocation other than the one that is best given the tax rate is to exert direct control over them, just as Britain would if it won a war. However, doing so would leave these constituents powerless and unable to secure any utility for themselves in the face of predation by their own government. They would no more acquiesce peacefully to their own government doing this than they would to Britain's attempt.

Finally, all parameters of the game are assumed to be common knowledge.

Analysis of the Model

The analysis of the model aims to answer just one question. Under what conditions will the tax (predation) necessary to satisfy Britain make peace so costly that the Colonists would prefer to revolt? I will state the answer in a proposition and then

discuss the intuition for it. The proof is in the appendix.

Proposition 2.3. *War will occur if and only if the sensitivity of the Colonists' allocation between transferable and non-transferable activity to taxation is high enough. When the costs of war are lower, the threshold sensitivity that separates war and peace is also lower.*

In this model, the only way that peace is efficient is if Britain agrees to a tax rate of zero. Any positive tax rate encourages the Colonists to shift resources into non-transferable activity, which produces less value per unit of resource than transferable activity. As the tax rises, Colonists shift more resources into even less productive non-transferable activity, and the costs of these distortions grows at an accelerating rate. Thus, given that a tax rate of zero would not satisfy Britain, peace will be costly.

Given that there is a tax rate that would appease Britain, whether war or predation occurs depends only on which has the higher costs. If the destruction of war and the cost of Britain exerting direct control over the colonial economy if it wins exceed the distortionary costs of the lowest tax rate that would appease Britain, then the Colonists will tolerate predation. But if the reverse is true, then the Colonists will seek to throw off Britain's predation, violently if necessary, and Britain will fight to take direct control of the economy.

Whether war or predation is more costly depends principally on how sensitive the Colonists' allocation of resources between transferable and non-transferable activity is to taxation. If even large tax rates would only lead to small shifts in allocation, then the distortion caused by the tax is small, the Colonists continue to produce close

to the full realizable value of their resources, and there is a tax that both Britain and the Colonists prefer to war. If, on the other hand, even modest tax rates lead to large shifts in the Colonists' allocation, then the distortion caused by taxation is large and the Colonists will produce much less than the full realizable value of their resources. Then, the tax that is necessary to satisfy Britain will have to be higher in order to account for the faster reduction in total value, and even the least such tax will not appease the Colonists.

Equilibrium here is inefficient whether it involves predation or war. The fundamental drivers of this inefficiency are an interlocking set of commitment problems and (unmodeled) asymmetric information. First consider lump-sum taxes. As discussed previously, in theory these would not create any distortions, because they don't alter the value of a given allocation on the part of the Colonists relative to any other. If Britain could commit to an initial assignment of these taxes, the Colonists would have no reason to alter their allocation from the efficient one, and the surplus from the lack of distortion makes possible a set of lump-sum taxes that both would strictly prefer either to varying taxes or war.

However, Britain cannot commit not to adjust these taxes in response to unexpectedly high value creation by the Colonists. So, if in setting the initial lump-sum taxes, Britain underestimates their productivity, then the Colonists will realize less than the full value so as to avoid revealing higher productivity and being subjected to a tax hike, and peace will be costly. This problem could be avoided if there were not asymmetric information about the Colonists' productivity and incentives to misrepresent it.

Even if Britain acknowledged this problem and switched to using taxes that varied in the amount produced (e.g., the linear taxes modeled here), this doesn't suffice to ensure that peace is costly. If the Colonists could commit themselves to the efficient allocation, then even varying taxes would not create costly distortions and both players could be made strictly better off than war or inefficient predation. They might be able to commit to this by having their government, or Britain, take direct control of their allocation decision. However, they would only cede this control if their government or Britain could commit not to take advantage of this control by taking all their transferable products. If either could commit, they would, but of course they cannot.

Thus, the presence of commitment problems and asymmetric information ensures that equilibrium must be inefficient, whether it involves distortionary predation or war. I turn now to applying these results to the American Revolution.

Stylized Facts of British-Colonial Relations

Here, I will first describe a set of stylized facts about the interaction between Britain and the Colonists in the 18th century. The evidence and argumentation for these facts is collected and presented in Rackove, Rutten and Weingast (2000) and de Figueiredo Jr., Rackove and Weingast (2006) ("FRRW"). I then argue that the model analyzed here explains why disputes over principles of taxation could lead to a large, expensive war. This section closes with a consideration of the alternative explanations for the war given by FFRW.

I will make use of four stylized facts that FFRW substantiate in their account,

enumerated and described below. All that I have added to their arguments is a summary of the dispute between Ireland and Britain over the former's legal status, which undoubtedly provided a recent historical analogy the Colonists could use to develop expectations about the ramifications of Britain's new policies toward them.[33] Readers seeking fuller documentation of the facts given below are referred to FFRW. For each, I also provide an interpretation of the fact in terms of the model presented here.

1. Before 1763, disputes over imperial policy (and authority) arose, but were always resolved peacefully, even though the underlying conflict was already present. There were three reasons for this. First, life for the Colonists was hard up to the early 18th century, so that there was relatively little surplus for Britain to extract. Second, the Colonists composed the vast majority of Britain's empire, and imperial policy-making was therefore attuned to their needs. Third, and most importantly, France was also powerful in North America, and formed an ever-present and very dangerous threat to Britain's colonies there. In opposing France, the Colonists and Britain had a very strong common interest, and the presence of the French in North America would have made open conflict between the Colonists and Britain very expensive because it would have rendered them vulnerable.

In the various wars fought with France and its Native American allies, the Colonists contributed substantially to their own defense, including quartering Britain's troops and raising their own militias to fight alongside them. They also paid great costs

---

[33]This summary is drawn from the entry on "Declaratory Act, 1766" in the Gale Encyclopedia of US History, published by Cengage in 2007.

through the Navigation Acts, intended to halt colonial trade with France, though these also led to enormous smuggling.

Nonetheless, there were serious disputes over colonial policy during this era. In particular, Britain's Board of Trade sought reforms that would allow it to assert greater control over various colonies, in part to clamp down on rampant smuggling. These were protested by the Colonists, and failed to pass Parliament, presumably in order to maintain a unified front against France.

In the model, the restrictions of the Navigation Act can be thought of as implicit taxes on the Colonists, because they effectively transferred value from the restricted traders to Britain by undermining the traders' business but also hurting France. However, as long as France remained powerful in North America, these restrictions also improved the general well-being of the Colonists, because they too benefited from the effort to contain France. In other words, these implicit taxes were effectively used to provide public goods for Britain and the Colonists. Thus, both *governments* had good reason to tolerate them, though individual colonists might not. The fact that many traders nonetheless shifted resources into non-taxable activity—smuggling—in such high volume indicates a high sensitivity on the part of the Colonists to taxation.

2. Trends in place before 1763 exacerbated the underlying conflict of interests, and laid the groundwork for its eventual exposure. First, the 1707 Acts of Union melded Scotland and England into Great Britain by uniting their Parliaments, but left Ireland out. Soon after, Britain passed the Declaratory Act of 1719, subordinating Ireland to both the Crown and Parliament of Britain, which could freely revoke Ireland's own Parliament, and asserting control over all legal actions to be taken in Ireland. The

resulting abuse of Irish interests motivated Drapier's letters (written pseudonymously by Jonathan Swift), which paralleled the later pamphleteering of American radicals like Thomas Paine, and considerable unrest in Ireland.

Second, the Great Awakening of the 1730s and 40s greatly strengthened religious observance and religious diversity among the Colonists, and increased their distance from the Church of England, which at that time was reasserting itself in Britain. This raised the value of the religious liberties entailed in the colonists' self-government, liberties that were considerably greater than those enjoyed in Britain.

Third, the Colonists were rapidly getting richer. Their population expanded enormously over the 18th century, and the Colonists swiftly expanded frontier settlements, putting new lands into production and developing the continent's vast natural resources. The net result was that both overall wealth, and income per capita, grew substantially over this period.

Fourth, and most importantly, over the course of the 18th century, the long series of wars among the major powers, fought both in Europe and in the areas of colonization, resulted in radical increases in the power, extent, and indebtedness of Britain's empire. Securing this empire was immensely expensive and would require a much more capable and assertive imperial bureaucracy than had existed theretofore. In particular, to maintain its dominance, Britain would have to draw on the wealth of its colonies.

Of special import to the Colonists was Britain's victory in the Seven Years' War. The settlement of the war included the expulsion of the French from the bulk of North America, and largely ended any near-term threat to the Colonists from France. (It

was precisely for this reason that it was safe for the rebelling Colonists to seek France's help during the war.) In doing so, it eliminated the strong interest the Colonists had previously shared with Britain in opposing France, and greatly reduced the costs of open conflict between the Colonists and Britain, for both.

In the model, the trends toward religious diversity and increased wealth among the Colonists can be thought of as increasing the value (R) that is potentially available to be "taxed," while Britain's new indebtedness increased the value to it of obtaining new sources of revenue. Britain might impose restrictions on religious practice, which would harm (transfer value from) colonial practitioners but benefit Britain in the form of the Church of England, or it might imposes taxes on the Colonists' new wealth. The treatment of Ireland after the Declaratory Act provided an example of what might happen to the Colonists if they allowed Britain to exert direct control over their affairs. But most importantly, the elimination of the French threat meant that the implicit (and high) taxes of the Navigation Acts were no longer providing public goods—they no longer benefitted the Colonists.

3. Britain well understood the risk of a colonial rebellion following the Seven Years' War, but it also needed money to pay for empire, had the power to demand it, and might need much more in the future. It therefore set out to reform colonial administration with an eye to its future needs. Thus began a series of attempts to impose unprecedented taxes on the Colonists. These measures were explicitly designed to accommodate the Colonists to new forms of taxation: the levies were very small; the taxed activity was chosen to be less visible in order to minimize popular aggravation and less evadable in order to minimize distortionary costs. Presumably,

the intent was to prepare the Colonists for later increases in these taxes as Britain required them.

However, each new attempt brought unrest, evasion, resistance, and boycotts of British goods from the Colonists. Britain initially reacted by repealing one measure and passing a new one, hoping to find something the Colonists would accept. When this failed, it began to enforce compliance. Britain responded to resistance by punishing those leading the resistance and asserting more direct control over colonial governance. Its punitive responses included suspending the assemblies of New York and Massachusetts, and closing the port of Boston and later blockading all of Massachusetts. Its assertions of control included taking over payment of colonial officials' salaries, and establishing a new customs board with British staff to enforce trade restrictions and new courts ruled by British appointees without juries to try smugglers.

In particular, the Declaratory Act of 1766, passed along with the repeal of the Stamp Act, held ominous portents for the Colonists' future. It was patterned nearly word-for-word after the earlier Declaratory Act for Ireland, discussed above. The Colonists understood exactly where this might lead. Thomas Jefferson, for instance, in response to the later Coercive Acts, argued that Britain's actions "too plainly prove[d] a deliberate and systematical plan of reducing us to slavery," which is the status the Colonists perceived Ireland to have.

Also, the Quebec Act of 1774 legally enshrined Catholicism in the newly acquired French-Canadian territories and expanded the remit of those territories into areas of the frontier adjacent to the Colonists'. It thus implicitly threatened the religious

liberty of the Colonists, and also showed that Britain was willing to impose substantial harm on the Colonists' prospects for growth in order to satisfy its other interests of placating the French-Canadians and the Native Americans.

In sum, the days of conciliation in order to preserve unity against France were over.

In the model, Britain's gingerly attempts at increasing taxation can be thought of as trying to find ways to minimize the Colonists' subsequent transfer of resources into non-taxable activity. Britain's efforts to enforce compliance were attuned to the need to exert more direct control over the colonial economy, by clamping down on the local assemblies and making colonial officials directly responsible to Britain. These were initial steps toward the more thorough-going solution Britain presumably would have implemented had it won the war later on, and the new Declaratory Act made this intention clear. Finally, the real wealth of the Colonists lay not in their current prosperity, but in the vast opportunities for development represented by the territory beyond the frontier. By curtailing these opportunities, the Quebec Act effectively transferred a large amount of value from the Colonists to Britain (in the form of a different province of the Empire). It thus made clear that Britain would not limit itself to the small direct taxes (e.g., on sugar and tea) it tried to enact, but would also impose much larger (though implicit) taxes.

4. The core of these disputes was not the actual taxes imposed, which were modest overall, but the principle that legitimated them. Both sides' material interests were inextricably entwined with the ideas of the Enlightenment and of the Imperial Era that Britain was entering, concerning proper forms of governance and definitions of

sovereignty. The Colonists supported a form of imperial federalism they viewed as the earlier status quo, in which the Colonists had control over all internal affairs, political (including the selection of local legislatures), economic (including taxation and expansion), and social (including religious freedoms), while Britain had control over all foreign affairs, especially trade and security. In this view, freedom from tyranny lay in local governance and the principle of no taxation without representation.

By contrast, Britain supported an absolute form of parliamentary sovereignty: Parliament itself was the main bulwark against tyranny, and was therefore "unlimited and unlimitable." The ancillary doctrine of virtual representation, developed in part in response to colonial agitation, held that Parliament provided legitimate representation even to those (such as the Colonists) who had no role in the selection of its members.

Of course, each side's ideology was entirely consistent with its material interests. Sovereign Parliament could freely and rightfully extract wealth from the Colonists to pay the expenses of empire; self-governing Colonists could freely and rightfully enjoy their own wealth. The real taxes contested by the two sides were not the piddling measures Britain tried to enact, but the much larger taxes that might follow these under whichever principle reined supreme.

The model provides only the crudest representation of the ideological debate between Britain and the Colonists, in the form of bargaining over the implemented tax. However, this stylized fact makes clear that taxes—interpreted broadly to mean measures that transfer value from the Colonists to Britain and distort colonial affairs— were indeed the heart of the issue. In bargaining over taxes, the Colonists were

concerned not with Britain's initial efforts at predation but with where the process might end: in effect, the final agreed tax.

————————————————————

These stylized facts reveal how the conflict of interests between Britain and the Colonists took shape. They explain why relations between the two were relatively pacific for most of the 18th century, and why these issues came explosively to the fore after 1763. They also render a clear rationalist interpretation of the disputes over principles of governance that connects these principles to implications for the future prosperity of the Colonists and predation upon them, and thereby resolves the mystery of why the Colonists would be provoked to the point of violence by the present, small taxes Britain tried to impose. However, despite all this, they do not make clear why the war happened.

Explaining the War Itself

Why couldn't the Colonists and Britain find a suitable compromise? After all, it's not as if Britain was unwilling to accomodate the principle of no taxation without representation. Earlier in the century it had unified with Scotland by combining the Scots Parliament with the English one. And it's not as though the Colonists were unwilling to compromise on a strong central government: not long after the Revolution, the weak Confederation was replaced with a much stronger United States government. Britain could have added some colonial representatives to Parliament, in accordance with the power and importance of the Colonists, in exchange for which the latter could have agreed to Parliament's sovereignty. Or, the two sides could

have agreed to retain imperial federalism, with the Colonists governing their internal affairs, but with a specified set of activities that Britain would be free to tax. Yet neither of these, nor any other reasonable alternative, were given serious consideration by either side.

The model presented here explains why. Any settlement Britain would agree to would have to transfer enough value from the Colonists to at least equal the value Britain would expect to get from a war to reassert its sovereignty. And any settlement that appeased Britain would have to leave enough for the Colonists to meet the value they would expect to get from a war of independence. But transferring enough to satisfy Britain imposed costs on the colonial economy, in the form of resources shifted out of transferable activity. These costs were high, and the anticipated costs of war were low, so no mutually satisfactory compromises existed and the Revolution occurred.

Obviously, the validity of this account turns on whether the costs of peace, in the form of the distortionary costs of the taxation that Britain was expected to eventually demand, actually exceeded the costs of war. There are two good reasons to believe that they did.

First, the Colonists were adept at escaping the reach of authority. The colonial governments had nothing like the sophisticated imperial bureaucracy that was then coming into being in Britain. Settlers expanded the frontier, whether permission was given or not. The frontier also provided a ready escape for anyone fleeing the authorities. Traders routinely disregarded customs law and smuggled in volume, evading the implicit taxes of the Navigation Acts with gusto. The prevalence of these

behaviors implies that the Colonists could quite easily shift their productive efforts into untaxed activity when needed. This in turn implies that they would respond sensitively to more taxation, and thus that the costs of peace in the near term would be high.

Second, and more importantly, the vast majority of colonial prosperity lay in the future, something the Colonists well knew. The expansion of settlement, the clearing of new land for agricultural and later industrial production, the extraction of natural resources, and the development of canals and other infrastructure to support the movement of goods, were the keys to the colonies' wealth. And yet at the time of the Revolution, all of these had only just begun. This fact has important implications for the costs of both war and peace.

With regard to the costs of war, it implied there was much less to be destroyed in a war begun at the time. War might be expensive in the near term, and indeed it was. But war posed no threat to the colonies' future wealth, as it was not even fought in the territory that contained most of this wealth, and the settled areas had much less capital and infrastructure in existence than they would later. Thus, at worst a war could destroy a small fraction of the colonies' total value, meaning that the costs of war for both Britain and the Colonists would be modest in the long view.

With regard to the costs of peace, it implied that the long-term sensitivity of the colonial economy to taxation was very high. Realizing the wealth tied up in America's interior would require immense entrepreneurial effort that heavy taxation might discourage. Since any negative impact on this growth would compound exponentially over time, even a slight decline in the natural growth rate would be extremely costly.

Indeed, this presumably explains the very low tax burden the Colonists imposed once they had won their independence and maintained until at least the Civil War.

Certainly, the Colonists could anticipate that Britain's policies would affect the growth rate. The Quebec Act demonstrated Britain's willingness to hand the Colonists' future growth opportunities over to more acquiescent provinces. Moreover, Britain continually hounded the colonial authorities to limit territorial expansion, largely because of its need to avoid a substantial, ongoing commitment of military resources to fighting wars with Native Americans on the frontier. This suggested that, in the future, colonial expansion would be delayed as necessary to bring peace to the frontier and enable Britain to respond to urgent military needs elsewhere. Given the experience of Britain's many wars in the first half of the 18th century, these delays could occur regularly.

This is not to say that, had Britain retained the colonies of the future United States, it would not have invested heavily in their development. This development was in Britain's interest, too, and indeed it was partly why Britain had a policy of "salutary neglect" toward the colonies in earlier times. However, if the colonies remained in the Empire, this interest would sometimes be traded off against Britain's other interests. In particular, the Colonists could expect to be regularly squeezed to help pay for Britain's many wars. The problem was that, after the defeat of France in the Seven Years' War, these wars were not relevant to the Colonists' well-being and thus constituted a pure transfer that could only distort the colonial economy.

The bottom line is that, both now and in the future, the Colonists had good reason to anticipate that the taxation that would be needed to satisfy Britain would impose

great costs on their own prosperity. Greater even than the costs of a war, now, to win independence and end future predation, which could only become more expensive over time. So, they declared independence. For its part, Britain could anticipate that the Colonists would resist its extraction of revenue, and had already begun to increase its control over the colonies. Faced with a declaration of independence, Britain chose to fight a war to take full control and end resistance to its taxes.

Alternative Explanations

FFRW give two subtly different explanations for the war in their two papers. I will describe each in turn, and then discuss the fundamental flaw that is common to both. The problem is that the reasoning underlying these explanations rules out bargaining between the Colonists and Britain. This leads to war, but for empirically implausible reasons, as I will explain.

First, the explanation in Rackove, Rutten and Weingast (2000) relies on a commitment problem. In this story, the stability of Britain's governance of the Colonists is built on the continuation of a long-standing policy of benign neglect. This policy is interpreted by the Colonists as an endorsement of their philosophy of imperial federalism, itself derived from the ideas developed in the time of Britain's internal struggle over self-government, and so is satisfactory to them. However, changes in Britain's interests, deriving mainly from its war victories and the concomitant expansion of its empire, lead Britain to take steps that depart from tradition and violate imperial federalism, though they are fully consistent with Britain's actual philosophy of parliamentary sovereignty. These steps break tradition and thereby undermine

the Colonists' confidence in Britain's commitment to imperial federalism. Thus, the Colonists worry that Britain cannot commit not to intervene willy-nilly in their internal affairs, something they could not tolerate. They then initiate a war to secure their independence.

The second explanation, in de Figueiredo Jr., Rackove and Weingast (2006), offers a similar story, but is focused more on understanding how it could (rationally) be true that both the Colonists and Britain were each surprised by the vehemence of the other's reaction to what they viewed as rightful, uncontroversial behavior. In this telling, the Colonists were unaware that Britain's philosophy had shifted from imperial federalism to absolute parliamentary sovereignty over the course of the 18th century, and remained so as long as the common interest of opposing France led Britain to treat the Colonists lightly. The Colonists thus believed that they played a kind of repeated prisoner's dilemma with Britain, in which mutual cooperation consisted of the Colonists' loyalty to the empire and Britain's non-intervention in their internal affairs. When the French threat was removed, Britain's interests changed and it imposed modest, though novel, taxes that were in keeping with its philosophy and that it thus still saw as cooperative. Because the Colonists incorrectly believe that Britain still supports imperial federalism, they view this as (surprising) defection and so respond with defection: a declaration of independence. And because Britain is unaware that the Colonists have false beliefs, it is surprised by their defection and also responds with defection: war.

Both of these explanations for the war are seriously flawed, and for the same reason: they do not allow the two sides to bargain over a new arrangement for gover-

nance. In both explanations, the choice of this arrangement is binary. Britain either intervenes or it doesn't. In the first explanation, the commitment problem is not the one to which Rackove, Rutten and Weingast (2000) refers. If the only two choices are intervention or not, and the intervention is anticipated to be substantial, then Britain would strictly prefer a war to no intervention, and the Colonists would strictly prefer a war (the result of declaring independence, given Britain's preference) to intervention. Neither would commit to an option other than the one it chose, because the other option would leave it strictly worse off. War happens because the disputed issue is completely indivisible. This is still war due to a commitment problem: the two sides would prefer to flip a coin to determine the outcome, but they can't commit to respect the result.

In the second explanation, war happens because of asymmetric information about the game being played and about Britain's preferences. However, there is no incentive to misrepresent this information on the part of either side. Thus, if they had the opportunity to communicate their understandings in the course of bargaining over a settlement, they would no longer have false beliefs about the other's interests, and the only thing standing in the way of peace would be the binary choice of settlement.

It seems highly implausible to assume that only two settlements were available to the Colonists and Britain. What would stop them from considering a compromise, in which Britain would be given some authority over the Colonists (e.g., a specified set of activity that could be taxed, or specified revenue or rates of tax), and the Colonists would retain some autonomy (e.g., over religious affairs and frontier policy)? If an agreement existed that both sides preferred to war, then they could overcome ini-

tially incorrect beliefs about the other's interests in the course of bargaining. They would then be able to identify this agreement (or another like it), and its implementation would be stable (within the bargaining range) because each side could credibly threaten war if the other violated it. Thus, FRRW cannot explain why the war itself occurred.[34]

The model I have presented in this section can: no such agreement existed, not because it was not physically implementable or because of incorrect beliefs, but because the anticipated distortionary costs of the taxation necessary to satisfy Britain reduced the value of the peace to the point that war, which would eliminate these costs, was preferred.

2.6 Developing the New Explanation

The principal intent of this essay was to convince the reader that costly peace is an empirically relevant rationalist explanation for war. To that end, it analyzed three sources of costly peace—arming, imposition, and predation—and showed how, in each case, underlying commitment problems or asymmetric information forced actors into inefficient equilibria, and war occurred whenever its costs were lower than those of peacefully coexisting with arming, imposition, and predation. It used these three sources to provide analytic narratives of three different cases: the Iraq War, the civil conflict in Iraq between the wars with the United States, and the American War of Independence. Thus, costly peace is empirically relevant: it explains at least one

---

[34]More generally, two-by-two models of conflict such as the prisoner's dilemma are, by construction, unable to illuminate why two actors would choose war over a peaceful settlement.

inter-state war, one intra-state war, and one extra-state war better than the available alternatives.

It is my hope that scholars of war will take from this the main implication: costly peace should be added to the rationalist toolkit for explaining the origins of wars. It is useful for understanding the wars studied here, and it might be useful for understanding many more. A second implication is that, even if there is some war for which scholars have confidently identified the presence of a commitment problem, asymmetric information, or both, these do not suffice to explain the occurrence of the war. To have a complete account of its causes, it must also be established that costly measures *other* than war, which the belligerents could have taken to address the identified drivers of inefficiency, either did not exist, were insufficiently effective to suppress the underlying driver, or were simply more costly than war.

For those interested in using the bargaining theory of war to make informed predictions about the war-proneness of some dyad, or to perform statistical analysis of the historical record of wars, there is a third implication. Analysts must beware of using only indicators of either commitment problems or asymmetric information as predictors, without also taking account of the possible presence of measures available to the players that, while costly, might be cheaper than war and hence prevent war from occurring. Ignoring these other measures will bias the resulting predictions toward war.

Finally, for those who seek to advise policy-makers on how to respond to the threat, initiation, or continuation of war around the world, there is a fourth implication. Wars that occur due to costly peace are not like those that occur in the absence of costly

peace—some interventions which would improve the lot of the (potential) belligerents in the absence of costly peace will actually harm them in its presence. Preventing or stopping such a war may simply force the belligerents to suffer even greater costs than those of the war. A decision to prevent war or impose peace should only be made if there is some belief that the war is not being caused by costly peace.

For both scholars and practitioners, then, it is valuable to be able to distinguish which wars occur due to costly peace. Unfortunately, and whatever the merits or flaws of the analytic narratives I presented in this essay, divining the principal cause of any war is most certainly a black art. That said, its value is such that I believe scholars ought to do more of it, and in the most rigorous and thorough fashion possible. To that end, I will conclude this essay with a brief set of suggestions for how to go about identifying other wars that are plausibly due to costly peace, and a speculative list of some possible candidates for such wars.

How to Distinguish War due to Costly Peace Empirically

In practice, there are at least three criteria that should be met in order to distinguish whether a war can plausibly be explained with reference to the costs of peace. These are not exhaustive, but they are in some sense minimal requirements, and each corresponds to a potential pitfall for a scholar eager to attribute some war to costly peace.

First, the participants had to have the means to overcome any commitment problems or asymmetric information without going to war. This is potentially a very challenging requirement, because it is hard to think of a war that does not involve at

least *some* movement in the balance of power and *some* asymmetry of information. And, since most wars are preceded by substantial spending on arms, it is also difficult to think of a war in which costly peace played *no* role in its causes. When more than one of these explanations appears to contribute substantially to the occurrence of a particular war, then it is very hard to identify one or another as being the dominant cause. To have any confidence in an attribution of a war to costly peace, there must be a reasonable case that, had the participants not gone to war, peace would not entail much shifting power or asymmetric information.

Second, peace must actually be substantially costly *overall.* It is not enough that one actor sees peace as costly because he does not get his way, or because he suspected the other actor was taking advantage of him, or because of any reason that he alone expected to gain from war. Both actors must anticipate that peace between them will yield substantially less than the total value it could, were it not for some source of costs.

Third, and this is one is the most tricky, war must somehow, in expectation, *reduce* the costs of peace. This criterion is much harder to meet than it might first appear. A story must be told for why there was no deal the actors could agree to, no mechanism they could implement, that would mitigate the source of costs in peace, other than war. In other words, how exactly does war solve the problem of these costs, and why can't this solution be implemented peacefully?

Other Wars due to Costly Peace

I make no claim that the three sources of costly peace identified here—arming, imposition, and predation—are exhaustive empirically. They are only the ones I have thought of so far, and there are probably more. But, in addition to the particular wars discussed in previous sections, there are a number of other wars that might be attributed to these three sources. It seems likely that analyzing these wars through the lens of costly peace would be a valuable exercise.

Arming: Wars of consolidation, secession, and succession might derive substantially from arming costs. Fearon (1995) offered the possession of a given throne as a possible example of an indivisible issue, but then noted that, while in principle a throne is divisible—it can be shared or alternately held, or its concomitant territory divided, by two potential occupants—in practice, the norms of monarchy make it appear indivisible. An alternative explanation that either supplants norms or provides an implicit explanation for where the norms come from is that splitting up territory or sharing a throne requires that both occupants maintain the militaries necessary to preserve their claims. If there are economies of scale in the maintenance of such militaries or in the protection of one's territory from outside powers, then the added costs of multiple, smaller armies may exceed the costs of war to retain a unified throne. Thus the costs of arming prevents territory from being divided too far below a certain efficient scale, and when this scale is reached, war will be preferred to further division.

Imposition: Could some of the civil wars that occurred during the Cold War have

been due to the imposition of penalties and rewards by the two superpowers? In the global competition for states not solidly in either superpower's camp, each superpower routinely supplied substantial military and other aid to its favored elements within such states, which was intended to increase the leverage of this group relative to the other favored by the opposing superpower. If the peaceful coexistence of the two groups implies continuing costs for the superpowers of arming and support, then both superpowers may try to incite their clients to fight a war to consolidate control over the state and eliminate these costs. The same can be said more generally for the desire of hegemonic states to avoid bearing the costs of war to eliminate some nuisance by encouraging a local party to do it for them.

Predation: Wars over raiding and piracy—essentially, land- and sea-borne predation—have occurred since antiquity, and continue in the rangelands of Africa and potentially off the coast of Somalia. Do the costs of predation cause these wars? If the producers fight to end the predation, do the predators fight to temporarily erode the producer's defenses and so take full control over its resources while its strength recovers? Also, many wars have been fought over prohibition, from the international opium wars of the 19th century to more recent intra-state wars in Colombia and Mexico. Because the social costs of (certain) drugs' use are thought to exceed the individual benefits, drug traffickers effectively prey on productive society. Is this what causes these wars?

Chapter 3

The Modern Economic Peace

Abstract:[1] Rationalist theories of interstate conflict do a poor job of explaining some international relationships, especially those in which peace seems most comprehensive. I argue that this is because these theories have little to say about a factor they nonetheless identify as central to war and peace: the presence and magnitude of underlying conflicts of interest. I begin to develop a theory of the origins of disputes among nations, based on the idea that the *economic* conflict of interests between two states is determined by the benefits and costs of transferring wealth (by whatever means) from one state's economy to the other's. Whether such a transfer happens in equilibrium depends not only on the military situation between the two states, but also on the characteristics of their economies and governments. Nations that have very sensitive, highly integrated economies—what are commonly referred to as "modern" economies—of comparable size and that are ruled by highly representative

governments have little to fight or coerce one another over. If this theory is correct, it could potentially explain not only the warmth of relations among such nations, but also the very-long-run decline in organized violence and the rapid global economic liberalization and political democratization of the past two centuries.

3.1 Introduction

If you ask an analyst or official in the US foreign policy community, "why do the United States, Japan, Canada, Australia, the United Kingdom, and the member states of the European Union get along so well with each other?", their first answer is usually "they're allies." Upon pointing out that this reply simply begs the question, their second answer is often "common interests." Asked to elaborate, the subsequent list typically includes stemming the proliferation of advanced weapons, suppressing terrorism and civil conflict, preserving international peace, and encouraging the spread of political freedom and economic development.

Not a few academic political scientists would offer the same set of responses, and yet they are hardly satisfactory as an explanation for the deep comity among these nations. The rich states of the imperial era, including many of the same countries included in the question above, shared a quite similar set of interests, with the exception that the ideology they sought to spread was not liberty but rather the superiority of the colonizers. Yet these common interests did not suffice to prevent these states from engaging in wars, arms races, proxy conflicts, and assorted other adversarial behaviors with each other. In general, the phenomenon of a group of wealthy states that share an interest in doing what it takes to preserve and entrench their success

is nothing new, but they usually still compete with each other, often intensely. But unlike in previous eras, the members of today's group do not war with each other, do not arm against each other, and rarely even have serious diplomatic disputes. Why are today's states different?

When pressed, and especially after the question is more pointedly stated as "why doesn't the United States use its superior military power to take advantage of the others?", both policy types and academics will give a stronger answer. Namely, what's good for Europe (say) is good for the United States, for the most part. This answer certainly feels right. It accords with the warmth with which the governments of these countries regard each other, and the durability of this warmth—disputes over strategy (e.g., with regard to Iraq and climate change) do occur, but these are not taken as the result of fundamental underlying differences in interest, and any tension is quickly dispelled.

That this answer is so intuitively satisfying should not distract us from the fact that we have no precise idea why it might be correct. Why are the common interests among these nations so deep that the suggestion that the US use its military might to extract wealth from the others is almost universally regarded as absurd?

In this paper, I will argue that rationalist theories of international relations do not answer this question. In particular, simply pointing to certain attributes of these nations and noting the well-established correlation between these attributes and peace—whether democracy, capitalism, or others—does not suffice. It again simply begs the question. Though there are available answers for how these attributes lead to peace among nations that display them, the postulated mechanisms offer no convincing ex-

planation for the *depth* of the peace these nations share. In fact, I will show that these mechanisms cannot explain this kind of peace *even in principle.*

Instead, I propose that answering this question requires that we develop a theory for why disputes occur at all. That is, we must find explanations, not for war, but for conflicts of interest among nations. The absence of this understanding may be the source of our inability to explain the deep comity among certain nations. It also raises doubts about the completeness of our explanations for international relations, even when our theoretical predictions do match observation, and renders suspect many of the conclusions drawn from statistical studies of the historical record of conflict.

To begin the development of a theory of the origins of conflicts of interest, the problem must be greatly simplified. This study deals only with *economic* conflicts of interest, between only two states, endowed with highly abstracted economies and political regimes. A model constructed in accordance with these drastic simplifications nonetheless leads to interesting results. The most important of these is that two states endowed with democratic regimes and economies of comparable size that are highly sensitive to taxation and thoroughly integrated with each other have very little to gain from engaging in coercion, and thus little to no economic conflict of interest. They should therefore regard each other essentially without fear or malice. By contrast, the absence of any of these characteristics generates incentives for coercion and the possibility for war and other costly behaviors.

The model inspires conjectures about possible resolutions to other long-standing puzzles in the study of international relations and of politics more generally. First, since the prevalence of democracy, the average sensitivity of economic activity, and the

depth of global economic integration have gone up over time, especially in the last two centuries, the theory predicts that the magnitude of interstate economic conflicts of interest has decreased, and so provides an explanation for the observed very-long-run decline in the prevalence of interstate violence. Second, this general decline in economic conflicts of interests should also reduce the prevalence of other costly responses to bargaining problems, such as the imposition of trade and investment barriers, and so may also help to explain the advent of liberalization of the international economy over the last two centuries. Finally, the sensitivity and integration of economies also has implications for the stability of more egalitarian (e.g., democracy) versus more monopolized (e.g., dictatorship) political regimes. Thus, the spread of democracy in the last two centuries may itself have been encouraged by changes in the identified economic factors.

The next section establishes some stylized facts about the warm relations among certain countries in the last few decades and argues that these cannot be explained by rationalist theories based on the bargaining framework. It suggests that this explanatory hole may be related to the absence of any underlying theory of where interstate conflicts of interest come from. Section 3 then proposes and defends a highly simplified model from which the development of a theory of the origins of conflicts of interest might proceed, based on the principle that coercion, even in the absence of the use of force or preparation for it, isn't necessarily free. Section 4 analyzes the model to determine how the economic and political structures of two states affect the magnitude of their conflict of interests, and then offers two empirical examples intended to illustrate the results. Section 5 concludes with a survey of the possible

applications of this theory to several other empirical questions.

3.2 The Puzzle of Deep Interstate Comity

Here I will discuss the empirical phenomenon that is the central motivation for this paper. Namely, some countries share a particularly deep friendship, as evidenced by four properties of their relations that are described below. However familiar and intuitive these properties seem, each confounds the expectations generated by the available rationalist theories of international conflict and is thus mysterious. In fact, the problem is worse than that: I proceed to show that *any* rationalist theory based on the extant bargaining framework and thereby reliant on international differences in bargaining failures or the costs of war will fail to dispel the mystery. I suggest that a new avenue, based on differences in the underlying conflicts of interest, which have heretofore not been explored in the rationalist literature, offers a way to solve the puzzle.

Stylized Facts

1. *The balance of military power sometimes differs radically from the distribution of wealth between two nations.*

The most obvious examples are between the United States and its closest friends: Canada, Australia, New Zealand, Japan, the members of the European Union, Iceland, and Norway. Even during the Cold War, most of these countries spent substantially less on their militaries than did the US, relative to GDP, and in absolute

terms the difference was tremendous. Since the Cold War's end, most of these countries have greatly reduced their military spending relative to GDP, while the US has maintained relatively high spending, despite occasional dips. In addition to the difference in raw spending, the US also gets considerably more military power out of each dollar spent. First, US military technology is considerably more advanced than that of its friends, and the technological gap has widened substantially since the Cold War ended. Second, the US military is a fully professional (volunteer) force, whereas much of the military manpower of its friends is composed of short-term conscripts. Third, the US military has extensive, recent experience of intense combat; most of its friends' do not. And fourth, the US maintains substantial military forces within the territory of its friends, but they do not station forces in the US. The net result is that the balance of military power greatly favors the United States.

At the same time, these US friends are among the richest countries in the world, in both per capita and total terms. In fact, several of these countries are richer than the United States, in per capita (Liechtenstein, Luxembourg, Norway) and total (the EU as a whole) terms. Even the poorest (Romania, Bulgaria) are among the richest third of nations in total terms, and in per capita terms still have the same order of magnitude of wealth as the US.

So, in every case, the balance of power is overwhelmingly weighted toward the US, while the distribution of wealth is less lopsided, and in some cases is particularly disjoint (e.g., Japan, Canada, Norway, Germany).

2. *The most powerful state advocates for and materially contributes to the strengthening of the others, in political, economic, and military terms.*

The military preponderance of the United States would be even greater if the US did not share much of its technology and training with its friends. Although the most advanced technologies (e.g., stealth) are usually not shared, this seems to be aimed more at preserving the secrecy of the details of these technologies from outside enemies than at preventing allied militaries from gaining the advantages associated with them. Moreover, both during the Cold War and since, the US has constantly exhorted most of its allies to increase their military spending, incorporate new technologies, and enhance training to improve the efficacy of their militaries. It also encourages its European friends to more deeply coordinate and integrate their militaries, so as to improve overall military power. And it also encourages Japan to allow its military to take a more active role in foreign conflicts.

Economically, the US encourages its friends to undertake reforms that it believes will increase the productivity and growth rate of their economies relative to its own. It also allows many of the scientific and technological advances developed in the US to be freely shared with its friends (and others).

Politically, the US encourages the integration and consolidation of its friends, as well as their involvement in international affairs. The US has strongly supported the political consolidation of Europe, and the expansion of institutions such as the European Union to include Eastern European countries, and generally supports the cooperation of European countries in participating in international institutions and diplomacy. It also supports the "normalization", which is to say increased activism, of Japan's foreign policy.

In short, the US encourages and materially supports the power of its friends.

3. *Many states delegate the protection of their vital interests to the most powerful state, but pay no tribute in exchange.*

Beginning by World War II, and increasingly since, many of these friends have delegated the protection of their interests—even vital interests—to the US. This, despite the rising vehemence with which the US advocates that they strengthen their own militaries. Perhaps most importantly, many of these friends have explicit guarantees that, should they be attacked, the US will come to their aid with whatever force is necessary, up to and including the use of nuclear weapons against a nuclear-armed adversary. Moreover, these friends seem to believe that these guarantees are credible.

Remarkably, these states appear to have no fear that the US will take advantage of their decision to entrust their protection to it. Their own militaries, such as they are, do not train a wary eye on the US forces stationed in their home territories. They do not conceal their plans for territorial defense from the US, nor do they appear to even plan for the possibility that the US might turn against them. What military technology they develop is often shared with the US, and failures to do so seem to be due to the desire to preserve commercial, rather than military, advantage. Their forces, when they train for serious combat, often do so with US forces, enabling the US to learn their strengths and weaknesses.

Finally, there appears to be little payment, implicit or explicit, in return for US protection. Some friends do help the US to defray the location-based costs of stationing forces in their territory, but this ignores the budgetary costs of hiring and training the associated manpower and buying and maintaining the associated equipment and the opportunity costs of not stationing these forces elsewhere. There

is an expectation that these friends will support the US in international diplomacy, but this expectation appears based more on the perception that such support is actually consistent with the friends' interests, even in the absence of the alliance, than on the need to repay the US for its protection. Certainly, the US exacts no recognizable tribute from these states; it is not running a protection *racket*. On the occasions when the US has sought to redeploy its forces out of one friend's territory and into one where those forces might be more useful, the first friend often resists the move and the US usually attempts to conciliate it for the move. Undoubtedly, this has to do both with the perceived security benefits of locally-stationed US forces, as well as the positive stimulation of the local economy that comes with them. Indeed, since the end of the Cold War the US has sought to bring more of its forces home, not to rescind or lessen its provision of allied security, but to avoid the expense of foreign basing.

4. *The disputes in which these states engage are over stakes of trifling value, relative to their total wealth.*

The relations among these countries are, of course, not entirely without strife. Serious disputes over how to respond to Iraq's suspected weapons programs, locations for the foreign basing of US troops, appropriate climate change policies, commercial openness to genetically modified foodstuffs, and the subsidization of civilian aircraft manufacture, among others, have occurred. And, on any given day, there are many disputes among these friends over commercial and other policies being tried at international institutions such as the World Trade Organization. However, it is important to observe several common properties of these disputes.

First, many of these disputes are over means, not ends. For example, these countries all sought to ensure Iraq's disarmament, but disagreed over whether war was the best way to do so. All agree that open commerce and more productive agriculture is best, but disagree about the safety of genetically modified organisms. This latter example, and many other regulatory disputes, are based largely on epistemic differences—about whether the evidence for anthropogenic climate change is sufficiently conclusive to undertake costly regulation, not about whether lessening the extent of climate change, if possible, is desirable.

Second, and relatedly, these disputes are often waged and settled by persuasion rather than by coercive threats. Precisely because many of the disputes are driven by disagreements about facts rather than conflicts of underlying interests, attempts by each side to convince the other of the case for its desired policy are a common means of negotiation. Even in the case of war with Iraq, perhaps the most serious recent dispute, the US invested serious diplomatic resources in convincing its allies of the case for war, but by and large refrained from issuing threats to coerce those that did not support war.

Third, when threats are used, they are quite mild in nature and do not escalate. Some commercial disputes among these friends involve threats to refer the case to the WTO and to take any retaliatory measures the WTO sanctions, and these threats are sometimes carried out. However, the adjudication of cases at the WTO, and the retaliatory measures that sometimes follow it, are very low cost compared to the wealth of the disputants, and these retaliations never escalate to anywhere near a real trade war. Diplomatic representatives are never expelled; threats of violence,

either implicit or explicit, are never made and would be regarded as an egregious overreaction.

Fourth, these disputes, even while ongoing, rarely affect the overall warmth of relations among the involved nations. Most of these disputes do not even enter the public consciousness in these countries. The most serious disputes, such as over the war in Iraq, do lead to public tension, but this always quickly blows over, and the general harmony of other aspects of relations is rarely undermined.

Fifth, and most importantly, the resolution of these disputes one way or another, even all together, would represent quite modest alterations to the distribution of prosperity. If all of the disputes ever lodged at the WTO involving a particular country were decided in its favor, this would hardly tilt the scales of prosperity at all. Even support for the Iraq War, whose total cost is liberally estimated to range into the low trillions of dollars, would represent only one or two percent of the total US economy in any year. These disputes are not all drops in the ocean, but most are no more than drops in a puddle.

Overall, disagreements among these states tend to be well-mannered and over modest interests.

Existing Explanations

It is very difficult to explain these stylized facts using the available rationalist theories of interstate conflict. I will argue that these theories offer no convincing explanation of any single one of these four facts, though this is partly a matter of judgment. However, even if we were to be as generous as possible in applying the

extant theories, none of them can explain all four facts together, *even in principle*. The reason, as I will show, is that the abstract mechanisms drawn upon by these theories place severe limits on just how friendly two nations could ever be.

As an explanation of these facts, many will quickly point to the democratic, commercially open, financially integrated, and culturally liberal nature of these countries, as well as their numerous memberships in international institutions, as there is substantial evidence that these attributes are associated with peace. These empirical observations are termed the democratic, commercial, capitalist, liberal, and Kantian peaces, respectively. Denote this set of observations, and that of the empirical association of any other attribute with peace, as the *attribute peaces.* Of course, simply noting that the countries in question share these attributes is not at all enlightening unless we can specify *how* these attributes cause peace, and particularly how they explain the four stylized facts.

There are some theories of various attribute peaces in the literature, which I will refer to as *attribute peace theories.* The rationalist theories rely on the bargaining framework for understanding international disputes put forward by Fearon and others. This framework implies that war will occur only when two sides suffer from a bargaining problem (either a commitment problem or asymmetric information and incentives for misrepresentation) that is severe enough relative to the cost of war. So, one commercial peace theory holds that, since war disrupts trade, nations that trade more face greater disruption should war occur, so that the cost of war among such nations is higher and thus it should occur less often. One democratic peace theory holds that the presence of organized, vocal opposition in legislatures enables democracies to

credibly reveal what would otherwise be asymmetric information about resolve, thus avoiding war. More generally, every rationalist attribute peace theory of which I am aware postulates that the specified attributes lead to peace by increasing the cost of war and/or avoiding or mitigating the occurrence of bargaining problems.

To see why these theories are inadequate to account for our stylized facts, we need some elements of the bargaining framework. Suppose we have two states, A and B, and that the set of stakes over which they might have conflicting interests is represented by the unit interval, $[0, 1]$. If either state should use force to try to resolve the dispute unilaterally, then A's expected value is p, minus a cost $c_A > 0$ since the attendant destruction reduces the value of the stakes; similarly, B's expected value is $(1 - p) - c_B$. The parameter p represents the balance of power between the two sides; the militarily stronger A is relative to B, the higher is p. At any given time, the disposition of the stakes can be represented by $q \in [0, 1]$, where A possesses the subset $[0, q]$ and B has $[q, 1]$.

Ordinarily, in a peaceful dyad we would expect to find q in the neighborhood of p: somewhere in $[p - c_A, p + c_B]$. If we take the set of stakes $[0, 1]$ to be the total pool of wealth available to states A and B, then q represents the balance of wealth, p represents the balance of power, and we expect the two to be similar to each other. This presumes, of course, that A and B have similarly averse or neutral attitudes toward risk, which seems reasonable in the case of rich, stable countries like those discussed here. It also assumes that there is no expectation of a rapid shift in the balance of power, as otherwise q should be shifted away from p in the direction favored by the declining state in order to compensate it for the other state's rise. But this,

too, seems reasonable to suppose.

One explanation for the first stylized fact, that among the named countries and particularly in dyads involving the United States, q is not near p, is that the relevant strategic context may not be merely dyadic. That is, there might be an external enemy or set of enemies so threatening that a dispute among the named countries about the distribution of wealth would be too dangerous to be worthwhile. However, since the end of the Cold War, it is very hard to see who this enemy might be. China is rising; nuclear weapons are (slowly) proliferating; and terrorism has become a more salient threat in recent years. But these threats are piddling compared to the threats the more powerful of the named countries might pose to others within the group. Moreover, what dispute might there be with the external threat that would be more important than the wealth of the group, composed of some of the world's richest states?

A seemingly much more promising explanation for a radical difference between dyadic balances of power and of wealth is in terms of the costs of war. If the costs of a war to reduce or eliminate the mismatch between power and wealth are large enough, they may overwhelm the expected benefits of the war for the comparatively wealth-deprived (and power-rich) side. Then, no war will occur, any threat of war will not be credible, and the mismatch can be maintained.

There are at least four reasons to doubt this explanation. First, the cost of war is not entirely exogenous. States that wish to resolve their conflicts violently have some (joint) control over how ugly the fighting gets. Nuclear weapons have only been used in one war, even though countries possessing them have fought and even lost numerous

wars; similarly for chemical and biological weapons. The Geneva Conventions serve to limit the costs of war among adherents; the so-called American way of war is defined principally by its severe aversion to inflicting non-military damage. The phenomenon is not limited to the modern era: belligerents in the Middle Ages often took care to avoid disrupting important trade routes. So it is not obvious why the costs of war can simply be assumed to be large enough to account for the mismatch between p and q.

Second, and related to the first reason, if the two sides have some influence over how costly a war would be, then in equilibrium the costs must have something to do with the value of what is at stake. Certainly, the two sides would not choose to escalate a war to the point that the costs they suffered exceeded the benefits they expected to win or defend. Moreover, the states under discussion are all very wealthy, so that the expected benefits for the comparatively wealth-poor side could be quite large in absolute terms. Thus, even if the exogenous component of the (absolute) cost of war is high, war might still be worthwhile on net given the size of the power-wealth mismatch.

Third, even if the cost was anticipated to be very high, a belligerent needn't threaten war, but only the *chance* of war. War between the United States and the Soviet Union was anticipated to be extremely costly due to the large nuclear arsenals each possessed, but this did not prevent the resolution of conflicts of interests with recourse to coercive threats. The threats employed were of the possibility that one or the other government would lose positive control over its military forces, so that a war would be started by accident and then might escalate to catastrophic proportions.

Because war only happens with some modest probability, the anticipated cost of carrying through on a threat to create the risk of war is not prohibitively high, and thus it is still possible for the participants to use threats to get their way.

Finally, would such a war really be so costly even if the cost is mostly exogenous? Most of the named countries would fall rather quickly to a US attempt at conquest, and this would be even more true had the US not shared much of its military technology with the other states (more on this momentarily). Moreover, the US needn't even occupy a targeted state once the latter's military had been vanquished—it could simply demand some of its wealth (in the form of a tribute, perhaps) and punish any refusal, without fear of serious retaliation.

Thus, the explanation of the first stylized fact based on a claim that the costs of war among these states are high is not convincing. But we can push the argument even further. Suppose, very generously for this candidate explanation, that it is somehow true that the cost of war is large enough to overwhelm the observed mismatch between power and wealth, though not larger than the total wealth of a particular dyad. Even so, it still cannot explain the second and third stylized facts. A shift in the balance of power moves the range $[p - c_A, p - c_B]$ and thus potentially erodes the future position of the weakened state if some exogenous factor (e.g., shocks or trends in each country's economy) shifts the balance of wealth. Thus, we would never expect to see both sides contributing to their own relative weakening, in the form of the US exhorting the others to build stronger militaries and the others relying ever more heavily on US power. These are the opposite of the behaviors we would expect to see: the US trying to increase its preponderance of power, and the others countering

its efforts.

Finally, suppose we are so generous as to allow that the true cost of war is extreme enough to overwhelm the total wealth of the dyad. Even so, these states obviously have other means of unilaterally affecting the balance of wealth: they can threaten to impose trade barriers, tax or confiscate other states' investments in their home territory, refuse to cooperate on other issues, and so on. Indeed, these measures are sometimes used by these countries, but as we established with the fourth stylized fact, the disputes to which these measures are applied are over stakes of quite small value. They could instead use these cheaper, but still costly means of coercion, or threats thereof, to settle larger disputes over the balance of wealth. Because these disputes offer greater potential benefits, it would be worthwhile to use these cheaper means, if necessary, and thus the threats to do so would be credible. Instead, and contrary to the implications of assuming even outlandish costs, we only observe them disputing trifles. Thus, the possibility that the costs of war among these states are large cannot explain this set of facts, *no matter how large these costs are.*

We can more quickly dispatch the possible explanations based on lessened bargaining problems. Even if the involved states were perfect bargainers, in the sense that they could credibly reveal asymmetric information at will and/or avoid generating it and could costlessly commit, completely and forever, to agreements that are mutually beneficial at the moment, this would not explain our stylized facts. The occurrence or absence of bargaining problems affects whether war or other inefficient resolution of disputes occurs, but they do not in and of themselves affect the settlement that should result in equilibrium. If the costs of war are modest, then regardless

of bargaining prowess, the distribution of prosperity should still reflect the balance of power. If instead the costs are large, the second and third stylized facts would still be the opposite of the expected behavior. And, if the costs are extreme, we would still expect to see bargaining over large disputes backed by threats to resort to the cheaper means of coercion.

The bottom line is that the existing theories of interstate conflict, based on the effects various national attributes have on bargaining problems and/or the costs of war, don't give compelling explanations for any of the four stylized facts, and cannot possibly explain the set of them. If the bargaining framework and the rationalist theories derived from it to date were right, then the EU should not so wholeheartedly trust the US to secure Europe's massive wealth, and the US should not be complaining about but rather taking advantage of this incredible arrangement to extract tribute from the EU. Many dyads among this group of nations should have vehement arguments over the distribution of prosperity among them, and at various times issue and perhaps make good on threats to redistribute prosperity involuntarily. Yet we see none of these things. Plainly, something is missing from our theories that is very important for understanding the behavior we observe among these states.

Differing Conflicts of Interest?

If the usual mechanisms can't explain deep interstate comity, then we might profitably ask whether there is some other mechanism—some relevant feature that varies across countries—that could. The avenue explored in this paper is that the magnitude of interstate conflicts of interest might differ across dyads. In the subsequent

sections, I will offer a theory for these differences, but first, it is instructive to consider how the possibility of differences in underlying interests have been dealt with by the extant literature, and why it is important to develop a theory of these differences.

Bluntly, there is no explicit rationalist theory of the origins of conflicts of interest among nations. Of course, the rationalist models all include a representation of the conflict of interests between two countries, typically in the form of the diametrically opposed preferences over the division of a unit interval. And they even implicitly allow the severity of the dispute to vary: though the interval is always of size 1, the costs of war are defined relative to it, so that the parameter(s) for the costs actually represent the ratio between the costs of war and the value of the disputed stakes. If the cost parameter is higher, it could be because the value of the disputed stakes is lower and thus there is a less severe conflict of interest. But there is little in the way of actual theorizing over what *particular* stakes that interval represents, why preferences over it are opposed, or why the severity of the conflict of interests might vary from case to case.

Many empirical studies do attempt to include proxies for differing conflicts of interest in their regressions. These proxies include measures of the similarity in a dyad's alliance portfolios or the sides' voting records in the United Nations, among others. However, in the absence of an understanding of why there were or were not serious opposed interests, how are we to know whether these actually are good proxies for common interests? There are, after all, plausible reasons they might not be. Some states may conspire to see to it that the issues which most severely divide them never come to a vote, so that their voting records present an inaccurate picture of the level

of commonality in their interests. And some states that allied in order to fend off a severe external threat might nonetheless be at each other's throats the moment that threat recedes. Even if we somehow convince ourselves that these are good proxies, we have simply moved the goalposts back one step, in much the same way as the attribute peaces don't explain war and peace so much as expose the question of why certain attributes are associated with peace or war. That is, why do some nations ally, but not others? Why do some nations vote together at the UN, but not others?

The absence of theoretically-driven measures of conflicts of interest from these studies raises concerns about the validity of their conclusions. Suppose that there is substantial variation in the magnitude of opposed interests (call this variable *magnitude*) across dyads and over time. In tests of or predictions based on bargaining theory, if *magnitude* is correlated with bargaining problems and anti-correlated with the cost of war, then estimates of the effects of bargaining problems or costs of war on the occurrence of war will be biased, though the sign of the estimates will remain correct. But if *magnitude* is correlated with the absence of bargaining problems and the cost of war, then estimated effects will be biased and even their signs may be wrong. Even if *magnitude* is uncorrelated with the other two, the presence of dyads with low *magnitude* but high bargaining problems and/or cost will lead to attenuation of the estimated effects.

These concerns and the empirical puzzle discussed earlier suggest that the development of a theory of the origins of conflicts of interest is an important enterprise for theorists. But this would be true even if these concerns were somehow assuaged and there were no identifiable empirical counterexamples to the extant rationalist

theories. After all, there would *still* be a gaping hole in our theories, and we would still have to suspect that these theories' predictions were right for the wrong reasons. Since there *are* serious disconnects between the predictions of these theories and empirical behavior, the value of developing a theory of the origins of conflicts of interest seems greater, but it would be an important task for students of international relations regardless.

3.3 A Simple Model of Economic Conflicts of Interest

In the remainder of this paper, I will offer a starting point for a rationalist theory of conflicts of interest, based on a model of interstate (and possibly intrastate) bargaining over economic interests. This model is tailored to the problem of explaining why some modern countries get along so well, and will be analyzed in the following section. Here, I describe its elements and assumptions and defend the choices made in the construction of the model.

There are two states, labeled A and B. Each state is composed of a government (labeled G_A and G_B), a citizenry (C_A and C_B), and an associated domestic economy. A government is a unitary actor that may be either of two types: a dictatorship or a democracy, with differences in interests to be defined below. Each citizenry is also a unitary actor. The stake over which the actors must bargain is the aggregate wealth of both states' citizens: they must divide among them the sum total of the private goods produced by both states' economies.

For concision, I will refer to these as the actors' "economic interests," but it is important to note that this definition of the stake under dispute is broader than it seems. A state's economy is composed of the production and consumption of all utility-bearing things by the state's citizens. "Private goods" are simply those utility-bearing things whose consumption is both excludable (one actor can prevent others from consuming them) and rivalrous (if one actor consumes them, others cannot). This category includes conventionally economic sources of value such as factors of production and consumer goods and services. But it also includes less obviously economic things such as religious adherence (e.g., whether certain people will be forced to practice a certain faith) and government policy (e.g., whether members of a certain ethnic group can be discriminated against).

Private goods are the natural candidates for economic conflicts of interest. By construction, actors would prefer to consume more goods rather than less, all else equal, because goods bear utility. If the good in question is not rivalrous in consumption (i.e., a public or club good), then the desires of all actors to have more are not incompatible.[2] If the good in question is rivalrous but not excludable in consumption (i.e., a common good), then the actors' preferences for more are incompatible, but no actual conflict arises because nothing can be done about this incompatibility: no actor can alter the degree to which it or another consumes the good. By contrast, private goods are precisely those utility-bearing things that *must* be divided (because

---

[2]Of course, actors might argue over who is to *provide* public goods, if doing so requires some private goods to be sacrificed as payment for the provision, but the underlying issue is still the division of private goods. Also, actors might bargain over inclusion into a good-providing club, but the only reason to exclude anyone is in order to motivate the sacrifice of private goods required to create the club good in the first place, so again the question is of the division of private goods.

they are rivalrous) and *can* be divided (because they are excludable). Thus, actors'
preferences for more private goods are both incompatible and actionable.

This is not so much a departure from the existing rationalist models as a spec-
ification of what exactly is meant by the "stakes" under dispute. In these models,
the stakes are typically represented as a unit interval that must be divided between
the two sides, who have opposed preferences over the division. This implies that the
stakes are private goods. Moreover, while the stakes in these models are not neces-
sarily defined to include *all* of the private goods available to the actors, the absence
from these models of the possibility of side payments is incompatible with assuming
that the actual stake is other than the sum total of all private goods available to the
actors. Thus, the stakes specified here are the same as the stakes assumed—usually
implicitly—by extant models.

Now, if the actors do not intervene to alter the division of private goods, then a
"natural" division arises, in which A's citizenry consumes goods of total value v_A and
B's citizenry consumes goods of total value v_B, and the two governments consume
nothing. The interval $[0, v_A + v_B]$ thus represents all the private goods the states'
economies produce and consume, in the absence of intervention by any actor, and
so plays the role of the unit interval in this model. The division of private good
consumption between the two states is labeled q, with $[0, q]$ consumed by A and the
rest consumed by B, so that the natural division is $q = v_A$. The natural division plays
the role of the status quo disposition of the stake in this model. It can be thought
of as a function of the two states' economic fundamentals: their endowments, their
productivity, their commerce, and so on.

Actors have the power to alter the natural division of wealth via either of two means. First, any actor may start a war, which ends the game with a costly lottery. The value of this lottery to each player is non-negative and depends neither on who started the war nor on what took place before the war occurred. For the governments and citizenries of the two states, these values are labeled W_A^G, W_B^G, W_A^C, and W_B^C respectively. War is costly, so $W_A^G + W_B^G + W_A^C + W_B^C < v_A + v_B$. If state $i \in \{A, B\}$ is ruled by a democracy, then it is assumed that $W_i^G = W_i^C$, for reasons that will be obvious shortly.

Second, the actors may make use of transfer instruments. A transfer instrument is a policy, together with an apparatus for its enforcement, that is mutually agreed among the actors and that results in the transfer of private goods from one state's economy to an actor. Transfers may be made within a given state and also between states in either direction. For the economy of state $i \in \{A, B\}$ and an actor $j \in \{G_A, G_B, C_A, C_B\}$, a transfer from i to j is represented by $\tau_{i,j} \in [0, 1]$, where $\tau_{i,j}$ is the fraction of state i's wealth that will be transferred to actor j.

These transfer instruments are most easily conceptualized as taxes that each government can impose on its own economy or that of the other state (if all actors agree on this) and whose proceeds can then be consumed by the taxing government or distributed to other actors for their consumption. However, "taxes" is broadly defined here. Empirically, governments have many instruments to choose from should they seek to extract value from their own or another state's economy, not all of which are conventionally labeled taxes: income taxes, poll tax, sales taxes, tariffs, land tax, confiscation of goods, cession of territory, enslavement of individual citizens. The

common feature of all these instruments is that they can be used to alter the division of private goods. For simplicity, we will assume that for any permissible i, j, the associated transfer instrument $\tau_{i,j}$ is unique.

Once again, this is not really different from extant models. Instead, it is simply a more detailed specification of what exactly is meant by agreed revision of the status quo disposition of the contested stake. In previous models, such revisions are just assumed to occur once agreed. Here, the actors must actually impose a policy designed to implement a revision.

The real departure from previous models is to recognize that the utilization of any plausible transfer instrument is itself *costly.* First, it is directly costly to implement a transfer instrument. Taxes must be somehow assessed and collected, in whatever form they come. Income must be reported and checked; the value of land must be measured; imports must be inspected and cleared through customs; confiscated goods must be transported; ceded territory must be occupied.

Second, and often more importantly, transfer instruments generate side effects that create indirect costs. Individuals subject to a particular tax generally seek to avoid (by doing less of the taxed activity) or evade (by concealing the taxed activity) it. Income taxes lead people to work less or to accept payment in-kind or "off the books"; wealth taxes cause people to stockpile less of whatever forms of wealth are taxed or to utilize offshore tax havens; tariffs lead to less trade or more smuggling; confiscation of land (or any other productive factor) discourages investment aimed at increasing its productive potential. Because these phenomena involve individuals deviating from the actions that would be most valuable to them in the absence of a

tax, they generally reduce the value of the taxed economy and so are costly.[3] For the purposes of this paper, we will term any such effects on the taxed state's economy "distortions."

The use of transfer instruments can also impose indirect costs on the other (untaxed) state's economy, which we will call "externalities." A tax on one economy will generally have effects on another if the two are at all integrated. People in one country who work less due to income taxes or trade less due to tariffs will produce less to export and consume fewer imports, and thus will force people in another country to deviate to consuming and producing other products. The confiscation of productive factors in one country leads not only domestic investors, but also foreign ones, to invest less in improving these factors. Less work and/or less investment will lead to less innovation in science, technology, and organization and thus less growth in both economies. Because these phenomena also involve individuals (in the untaxed economy) deviating from the actions that would be most valuable to them in the absence of a tax, they also reduce the value of the untaxed economy and so are costly.

We will assume for the sake of simplicity that the total costs borne by each economy depend only on the overall rate of tax levied on each, not on which actor imposes the tax or which receives the proceeds. That is, the costs depend only on $\tau_A \equiv \sum_j \tau_{A,j}$ and $\tau_B \equiv \sum_j \tau_{B,j}$. We can then represent these costs as functions of the

---

[3]Pigouvian taxes, which are designed to discourage individuals from generating public bads such as pollution or congestion, generate distortions that actually *increase* the overall value of the taxed economy. But they are still directly costly to implement, and the total revenue generated from such taxes is typically only a very modest fraction of the economy, so we ignore them. Also, tax revenue may be used to provide public or club goods, which may increase the value of the taxed economy enough to make up for the costs of taxation and so be on net beneficial. However, the revenues from the taxes considered here are assumed to be consumed as private goods by the receiving actor, and so do not yield these benefits.

tax rates, denoted $c_A(\tau_A, \tau_B)$ and $c_B(\tau_A, \tau_B)$. Then, the value of the two economies under a given tax vector is given by the functions $v_A(\tau_A, \tau_B) = v_A \cdot [1 - c_A(\tau_A, \tau_B)]$ and $v_B(\tau_A, \tau_B) = v_B \cdot [1 - c_B(\tau_A, \tau_B)]$. We assume that $c_A(0,0) = c_B(0,0) = 0$, and that $c_i(\tau_A, \tau_B) > 0$, $\left.\frac{\partial c_i}{\partial \tau_i}\right|_{\tau_A, \tau_B} > 0$, and $\left.\frac{\partial^2 c_i}{\partial \tau_i^2}\right|_{\tau_A, \tau_B} > 0$ if $\tau_i > 0$. We say that the two economies are not integrated if $\left.\frac{\partial c_i}{\partial \tau_j}\right|_{\tau_A, \tau_B} = 0$ for all $j \neq i$ and (τ_A, τ_B). Otherwise, we say that the economies are integrated (to at least some degree) and assume that $c_i(\tau_A, \tau_B) > 0$, $\left.\frac{\partial c_i}{\partial \tau_j}\right|_{\tau_A, \tau_B} > 0$, and $\left.\frac{\partial^2 c_i}{\partial \tau_j^2}\right|_{\tau_A, \tau_B} > 0$ for all i, j, and $(\tau_A, \tau_B) \neq (0,0)$. In plain English: if no taxes are imposed then there are no costly distortions or externalities; if taxes are imposed on an economy, they are costly to that economy and the cost increases, at an increasing rate, in the aggregate rate of tax; if the economies aren't integrated at all, then taxes on one have no affect on the other, and if they are at least a little integrated then any tax is costly to both economies and the cost increases, at an increasing rate, in the aggregate rate of tax.

Each citizenry's utility is equal to its consumption of private goods. A government's utility depends on its type. A dictatorship's utility is equal to its consumption of private goods, but a democracy's utility is equal to its citizenry's consumption. In other words, a dictatorship is greedy and seeks only to maximize its own self-interest, whereas a democracy is completely devoted to its citizenry and seeks to maximize its (aggregate) welfare. Obviously, real governments are not like this, but they do fall on a spectrum of representativeness of which the two types considered here represent the extremes. These extreme types are chosen so as to render starkly the differences among regimes, but the results presented in the next section can easily be extended to more fine-grained representations of governance.

The order of moves is as follows. The game begins with the natural division of wealth discussed above. The government of state A either starts a war, or offers to the other actors a revision of the natural division based on a vector of transfer instruments. If he makes an offer, each of the other actors must either accept it, or reject it and start a war (the order of these moves is irrelevant). If all the actors accept it, the vector of taxes is implemented, distortions and externalities result and the attendant costs are borne, and each actor consumes his post-tax allocation of private goods. Information is perfect.

This completes the description of the model; we now turn to its analysis.

3.4   The Political Economy of Coercion

We will now analyze the model to determine the actors' incentives to engage in "coercion," defined for our purposes as the use of power (in the form of bargaining power or the threat of war) to alter the natural division of wealth. The model does not have closed-form solutions for the equilibrium use of transfer instruments, but we can still compute comparative statics. We will focus on several parameters that affect the equilibrium level of coercion: the military balance of power (the actors' war values); the sensitivity of each economy (roughly, how quickly the cost of distortions in the taxed economy rises in the rate of tax); the integration of each economy (roughly, how quickly the cost of externalities in the untaxed economy rises in the rate of tax); the size of each economy; and the type of government. These enable us to develop a rudimentary theory of the political economy of coercion. We will see that under some conditions there are *no* incentives for coercion, and thus no conflict of economic

interests among actors. Finally, we will work through two empirical examples that illustrate the conditions that maximize or minimize the incentives for coercion, or equivalently the severity of the conflict of economic interests, and discuss how these might provide an explanation for the stylized facts of Section 2. The proofs for the results below may be found in the appendix.

We will concern ourselves mainly with the taxes that the government of A (G_A) will impose on its domestic economy and the foreign (state B's) economy. To facilitate this, we will assume that state B is governed by a democracy, so that it is effectively a unitary actor—the preferences of B's government are identical to those of its citizenry—and thus we needn't worry about taxes internal to B.

Assumption 3.1. *B is governed by a democracy.*

Next we need to simplify the problem by isolating the types of tax vectors that can actually occur in equilibrium. It turns out that it is enough to study only the total rate of tax that ends up being levied on each economy, and that in equilibrium there cannot be any "redundancy" in taxation.

Proposition 3.1. *Any tax vector is equivalent in outcome to one in which the government of A imposes taxes τ_A, τ_B on the two economies, and then distributes $x_A \geq 0$ of the resulting revenue to the citizenry of state A, $x_B \geq 0$ of the revenue to state B, and consumes the rest itself. Under this representation, in equilibrium, at least one of τ_i and x_i is zero, for any $i \in \{A, B\}$.*

The first statement follows immediately from our assumption that the cost of a tax to an economy depends only on the aggregate rate of tax imposed on each

economy. If it doesn't matter who imposes the tax or who receives the revenue, then a particular tax vector that results in a particular final allocation of private goods to each actor can be represented by one in which the same overall tax rates are levied entirely by G_A, who then distributes the revenue so that the final allocation is the same. So, we can subsequently speak only of G_A imposing domestic and foreign taxes and allocating the revenue, without loss of generality.

The second statement has a simple intuition. If G_A taxes an economy (meaning, collects wealth from the citizenry of the associated state), and then disburses some of the overall tax revenue to the associated citizenry, then the overall level of taxation is higher than it needs to be to achieve the same final allocation of private goods among the actors. Because a higher tax entails a higher cost to the taxed economy, G_A would do better to lessen the redundancy—that is, to reduce both the tax and the disbursement in a way that leaves all other actors equally well off—and then pocket the surplus from the lower overall cost of taxation. This implies that a peaceful equilibrium has one of three forms: positive taxes on both economies with all revenue consumed by G_A; a positive tax only on A's economy, with a possible disbursement to B; or a positive tax only on B's economy, with a possible disbursement to the citizenry of A.

We will make one other simplifying assumption. For the purposes of this paper, we will exclude any consideration of whether war would occur. (It is easily shown that war will occur whenever there is no tax/disbursement offer that will satisfy all actors, or equivalently whenever any offer that would satisfy two of the actors would impose more costs than war.) This is a topic for another paper; here the focus is on whether

transfers are demanded in equilibrium, and how they are determined. The following assumption removes the possibility of war in equilibrium, by simply restricting the sum of the actors' war values to be lower than the total value of the economies, even under the assumption that G_A collects as much revenue as possible and thus creates the highest plausible costs of taxation. Since these values are independent parameters, no restrictions on other parameters are implied.

Assumption 3.2 (No-War Assumption). *War is more costly than peace even under any revenue-maximizing set of taxes on both economies.*

Now we can discuss the main results. In what follows, the disbursement to state B is zero unless otherwise stated, and the disbursement to the citizenry of A is zero if G_A is a dictatorship and all the tax revenue if G_A is a democracy, unless otherwise stated.

Proposition 3.2. *Suppose that A is governed by a democracy.*

- *If $v_B \geq W_B$, then $\tau_A = 0$ and $\tau_B = \min\{\tau_B^*, \tau_B^w\}$, where:*

$$v_A \left.\frac{\partial c_A}{\partial \tau_B}\right|_{\vec{\tau}=\left(0,\tau_B^*\right)} + \tau_B^* v_B \left.\frac{\partial c_B}{\partial \tau_B}\right|_{\vec{\tau}=\left(0,\tau_B^*\right)} = v_B \left[1 - c_B\left(0,\tau_B^*\right)\right] \qquad \Rightarrow \tau_B^* \tag{3.1}$$

$$W_B = (1 - \tau_B^w)v_B \left[1 - c_B\left(0,\tau_B^w\right)\right] \quad \Rightarrow \tau_B^w \tag{3.2}$$

- *If $v_B < W_B$, then $\tau_A = \tau_A^B$, $\tau_B = 0$, and $x_B = W_B - v_B\left[1 - c_B\left(\tau_A^B,0\right)\right]$, where:*

$$W_B - v_B\left[1 - c_B\left(\tau_A^B,0\right)\right] = \tau_A^B v_A\left[1 - c_A\left(\tau_A^B,0\right)\right] \qquad \Rightarrow \tau_A^B \tag{3.3}$$

To understand this result, first consider what would be the ideal policy of the democratic government of A. Since its utility is equal to that of its citizenry, it would want to impose as little tax on its own economy as possible, since such a tax just lessens the wealth available to its citizenry. By contrast, it would seek to collect as much revenue as possible from the foreign economy, since it does not care about the other state, although if the two economies were integrated, it would also have to worry about the externalities its foreign tax imposed on the domestic economy. How well it can achieve these goals depends on how powerful state A is relative to B. If the natural division gives B more wealth than it would expect to retain in a war ($v_B \geq W_B$), then state A is externally "strong," in the sense that it is comparatively power-advantaged and wealth-deprived relative to B. If the opposite is true, then state A is externally "weak."

If A is externally strong, then it can and will extract wealth from B's economy. How much it can get depends on whether its ideal tax rate—the one that maximizes the consumption of its citizenry—leaves B enough to avoid war. If it does, then A's government will set this ideal rate (τ_B^*), and if it does not, then the tax will be *just* low enough to leave B enough wealth to prefer acquiescence to war (τ_B^w). In both cases, the tax rate A imposes decreases as the cost thereby imposed on B's economy ($c_B(0, \cdot)$ and $\partial c_B / \partial \tau_B$) rises, as this decreases the revenue that is collected at any particular tax and lowers the tax rate at which domestic consumption is maximized. If the two economies are not integrated, then the way to maximize domestic consumption is to simply maximize the revenue collected from the economy of B. However, if the two economies are integrated to some degree, so that a tax levied on B's economy

also imposes externalities on A's economy (i.e., $\partial c_A / \partial \tau_B > 0$), then A's ideal tax will be lower than the revenue-maximizing one because the democratic government of A fully internalizes the costs of the externalities to its own citizenry's consumption. This effect is more pressing if the size of A's economy is large relative to B's, so that the additional externalities of a higher tax quickly come to overwhelm the additional revenue the tax brings. By contrast, if A's economy is very small relative to B's, then the externalities will be negligible compared to additional tax revenue.

The tax A imposes decreases as B gets stronger relative to its wealth, and eventually B becomes the comparatively power-advantaged player, so that instead of taxing B, G_A must tax its own economy and transfer the proceeds to B (when $v_B < W_B$). Since G_A prefers to take away as little of its citizenry's wealth as possible, it collects just enough tax to satisfy B and avoid war (τ_A^B). The tax G_A must levy on its citizenry to satisfy B increases as B gets stronger relative to its wealth or the externalities on B's economy of the tax on A's economy $(c_B(\cdot, 0))$ rise, because A must do more to satisfy a more powerful B or one that suffers bigger side effects from A's effort to pay tribute. It also increases as the size of A's economy, relative to B's, shrinks or the distortions imposed by taxes on A's economy $(c_A(\cdot, 0))$ rise, as G_A has a smaller pool from which to pay B and thus must levy a higher rate of tax.

The most important result here is that the costs of transferring wealth from one state to another place limits on how large a transfer can actually occur, and thus on the extent of coercion. That is, the equilibrium post-transfer balance of wealth will reflect the balance of power, but only within certain limits. For example, even if A were completely militarily dominant, such that $W_B \approx 0$, A would not take all of B's

wealth, because at a high enough rate of tax, any increase in the tax generates *less* total consumption for A because it imposes such large costs on the two economies. More generally, the more sensitive the two economies are to taxation (i.e., the faster the taxed economy's costs ramp up as the tax on it increases) and the more integrated they are (that is, the faster the untaxed economy's costs ramp up as the tax on the other increases), the narrower these limits on equilibrium transfer will be, and the larger the resulting mismatch between the balance of power and the equilibrium balance of wealth might be. These costs thus narrow the scope for coercion to occur.

This is very different from the standard models, in which, so long as no shifts in power are expected and information is symmetric, the equilibrium disposition of the stake can differ from the balance of power only to the extent that war is costly. As argued in Section 2, in the standard models, a large equilibrium mismatch between power and wealth requires the costs of war to be implausibly high. In the model given here, a large mismatch can result even if war is relatively cheap. We will return to the empirical importance of this in the examples to be discussed later, but for now we will study how things might change if A were instead governed by a dictatorship.

Proposition 3.3. *Suppose that A is governed by a dictatorship and the two economies are not integrated.*

- *If $v_B \geq W_B$ and $v_A \geq W_A^C$, then $\tau_A = \min\{\tau_A^*, \tau_A^w\}$ and $\tau_B = \min\{\tau_B^*, \tau_B^w\}$,*

where:

$$\tau_A^* v_A \left. \frac{\partial c_A}{\partial \tau_A} \right|_{\tau_A = \tau_A^*} = v_A \left[1 - c_A \left(\tau_A^* \right) \right] \qquad \Rightarrow \tau_A^* \qquad (3.4)$$

$$\tau_B^* v_B \left. \frac{\partial c_B}{\partial \tau_B} \right|_{\tau_B = \tau_B^*} = v_B \left[1 - c_B \left(\tau_B^* \right) \right] \qquad \Rightarrow \tau_B^* \qquad (3.5)$$

$$W_A^C = \left(1 - \tau_A^w \right) v_A \left[1 - c_A \left(\tau_A^w \right) \right] \qquad \Rightarrow \tau_A^w \qquad (3.6)$$

$$W_B = \left(1 - \tau_B^w \right) v_B \left[1 - c_B \left(\tau_B^w \right) \right] \qquad \Rightarrow \tau_B^w \qquad (3.7)$$

- If $v_B < W_B$, then τ_A is the same as above but $\tau_B = 0$ and $x_B = W_B - v_B$.

- If $v_A < W_A^C$, then $\tau_A = 0$, τ_B is the same as in the first case, and $x_A = W_A^C - v_A$.

To explain this result, we will focus mainly on how the behavior of the government of A changes when it is a dictatorship rather than a democracy. Consider what a dictatorship's ideal policy would be. Because it cares neither for the other state nor for its own citizenry, but only for its own consumption, it would seek to collect as much revenue as possible from both its own economy and B's. When the two economies are not integrated, G_A's taxation of its own economy does not affect, and is not affected by, its taxation of B's economy, so the problems of what tax or disbursement to offer to the citizenry of A, and what tax or disbursement to offer to state B, are completely independent. Because these two problems are independent, and because both a dictatorship and a democracy have in common that they will try to get as much revenue as possible from B (or offer B as little as possible in order to avoid war), the external behavior of A does not depend on the type of its government. From B's point of view, in the absence of integration, whether A is governed by a dictatorship or a democracy does not matter at all.

Without integration, the only way the dictatorship and the democracy differ is in their internal behavior: that is, with respect to their own citizenry. The democracy never wishes to tax its own citizenry, as this reduces their consumption and thus its own utility, and will not do so unless it is externally weak and so has to pay tribute to B to avoid war. But the dictatorship would rather consume its citizenry's wealth itself, and thus would always like to collect as much revenue as possible from the domestic economy, whether it is externally weak or not. How much internal revenue the dictatorship can get depends on how powerful it is relative to its own citizenry (C_A). If the natural division gives C_A more wealth than it would expect to retain in a war $(v_A \geq W_A^C)$, then the dictatorship is internally strong in the sense that it is comparatively power-advantaged and wealth-deprived relative to C_A. If the opposite is true, then the dictatorship is internally weak.

If the dictatorship is internally strong, then it will set the revenue-maximizing internal tax rate (τ_A^*) unless this would be rejected by its citizenry in favor of war, in which case it sets a tax rate that leaves the citizenry just enough wealth to cause it to prefer peace (τ_A^w). Either way, the tax rate the dictatorship levies decreases as the cost imposed on its own economy $(c_A(\cdot, 0)$ and $\partial c_A/\partial \tau_A)$ rises, as this decreases the revenue collected at any particular rate and lowers the rate at which revenue is maximized. The tax the dictatorship imposes on its own economy decreases as its citizenry gets stronger relative to its wealth, and eventually the citizenry becomes the power-advantaged player, so that the dictatorship must stop taxing its own economy, and instead transfer some of the proceeds of its tax on B's economy to the citizenry of A.

Thus, when the government of A is a dictatorship, the costs of taxation place limits not only on external transfers (between states) but also on internal transfers (within A), and so on the extent of intra-state coercion. Even if the dictatorship has untrammeled power within its own state, it will not transfer all of its citizenry's wealth to itself, and the more sensitive to taxation his associated economy is, the smaller the transfer will be. Even so, overall a dictatorship will always impose more taxes—making the equilibrium more inefficient—than would a democracy, because the two tax B the same, but the dictatorship imposes higher internal taxes.

Next we explore the differences between dictatorship and democracy when the two economies are integrated.

Proposition 3.4. *Suppose that A is governed by a dictatorship and the two economies are integrated.*

- *If $v_B \geq W_B$ and $v_A \geq W_A^C$, then $\tau_A = \tau_A^*$ and $\tau_B = \tau_B^*$, where:*

$$\tau_A^* v_A \left. \frac{\partial c_A}{\partial \tau_B} \right|_{\vec{\tau} = \left(\tau_A^*, \tau_B^*\right)} + \tau_B^* v_B \left. \frac{\partial c_B}{\partial \tau_B} \right|_{\vec{\tau} = \left(\tau_A^*, \tau_B^*\right)} = v_B \left[1 - c_B\left(\tau_A^*, \tau_B^*\right)\right] \qquad and$$

(3.8)

$$\tau_A^* v_A \left. \frac{\partial c_A}{\partial \tau_A} \right|_{\vec{\tau} = \left(\tau_A^*, \tau_B^*\right)} + \tau_B^* v_B \left. \frac{\partial c_B}{\partial \tau_A} \right|_{\vec{\tau} = \left(\tau_A^*, \tau_B^*\right)} = v_A \left[1 - c_A\left(\tau_A^*, \tau_B^*\right)\right] \quad \Rightarrow \left(\tau_A^*, \tau_B^*\right)$$

(3.9)

- *If $v_B < W_B$, then $\tau_A = \min\left\{\tau_A^+, \tau_A^w\right\}$, $\tau_B = 0$, and $x_B = W_B - v_B\left[1 - c_B\left(\tau_A, 0\right)\right]$,*

where:

$$\tau_A^+ v_A \left. \frac{\partial c_A}{\partial \tau_A} \right|_{\vec{\tau}=\left(\tau_A^+,0\right)} + v_B \left. \frac{\partial c_B}{\partial \tau_A} \right|_{\vec{\tau}=\left(\tau_A^+,0\right)} = v_A \left[1 - c_A \left(\tau_A^+,0\right)\right] \qquad \Rightarrow \tau_A^+$$

(3.10)

$$W_A^C = \left(1 - \tau_A^w\right) v_A \left[1 - c_A \left(\tau_A^w,0\right)\right] \quad \Rightarrow \tau_A^w$$

(3.11)

- *If* $v_A < W_A^C$, *then* $\tau_A = 0$, $x_A = W_A^C - v_A \left[1 - c_A \left(0, \tau_B\right)\right]$, *and* $\tau_B = \min\left\{\tau_B^+, \tau_B^w\right\}$,

  where:

$$v_A \left. \frac{\partial c_A}{\partial \tau_B} \right|_{\vec{\tau}=\left(0,\tau_B^+\right)} + \tau_B^+ v_B \left. \frac{\partial c_B}{\partial \tau_B} \right|_{\vec{\tau}=\left(0,\tau_B^+\right)} = v_B \left[1 - c_B \left(0, \tau_B^+\right)\right] \qquad \Rightarrow \tau_B^+$$

(3.12)

$$W_B = \left(1 - \tau_B^w\right) v_B \left[1 - c_B \left(0, \tau_B^w\right)\right] \quad \Rightarrow \tau_B^w$$

(3.13)

We will concentrate on the effects of integration on the dictatorship's behavior, as well as the effect integration has on the differences in behavior between democracy and dictatorship. The first thing to observe is that integration decreases both the internal and the external tax rates levied by the dictatorship, relative to the case without integration. The reason is that, in the presence of integration, both taxes entail higher costs, because each creates externalities in the untaxed economy $(\partial c_i / \partial \tau_j)$, and this lowers the revenue-maximizing rates (τ_i^*, τ_i^+) regardless of whether the dictatorship is internally or externally strong or weak.

However, the lowering of equilibrium tax rates is more pronounced when the dictatorship is internally or eternally weak. In either case, the dictatorship must tax one

actor in order to pay off the other and avoid war. If the tax on one actor negatively affects the other actor through its externalities, then the dictatorship will have to pay the latter actor enough to fully compensate it for these externalities if war is to be avoided. For this reason, the dictatorship fully internalizes the externalities its tax impose. By contrast, when the dictatorship is externally and internally strong, it only cares about the effects (both distortions and externalities) its taxes have to the extent that they reduce its overall revenue. Because it does not have to compensate either other actor for these effects, it only partially internalizes them. This can be seen in the equations above: when the dictatorship is internally and externally strong (the first bullet), the externalities of each tax are weighted only by the tax rate the dictatorship imposes on the affected economy (e.g., $\tau_A^* v_A \partial c_A / \partial \tau_B$ in the first equation), whereas when the dictatorship is weak with respect to some actor and must pay it tribute, the externalities on that actor are given full weight (e.g., $v_B \partial c_B / \partial \tau_A$ in the third equation).

Even so, when it is internally strong, the dictatorship still sets an internal tax higher than the democracy would choose. Though the higher costs (due to externalities on B) lower the equilibrium rate of internal tax, there is still revenue to be collected, and the dictatorship will do so in excess of whatever may be required to pay tribute to B, just as in the case without integration.

However, unlike in the case without integration, the dictatorship's external behavior may now differ from that of the democracy. When either is externally weak, their external behavior is the same—they simply pay the minimum tribute required to satisfy B. There is also no difference when the dictatorship is internally weak, though

the two types of government have different motives. Both are trying to extract as much revenue as possible from the foreign economy, while taking full account of the externalities this extraction imposes on their own economy. But the democracy does so because it is trying to maximize its citizenry's consumption; the dictatorship does so because it must fully compensate its citizenry for these externalities in order to avoid war. Their behavior differs only when the dictatorship is both externally and internally strong. Then it has the power to levy tax on B, and the power to consider the externalities this imposes on its own economy only to the extent that they lessen its revenue from its domestic taxation. Thus, unlike the democracy, the dictatorship does not fully internalize the domestic externalities of foreign taxation, and so the dictatorship will levy a higher foreign tax than the democracy.

Overall, integration strengthens the limits the costs of taxation place on the extent of inter- and intra-state transfers, and so on the degree of coercion, because it raises these costs. However, its effects are more pronounced for democracies and weak dictatorships than for strong dictatorships, because the latter do not fully internalize the domestic externalities of foreign taxation. Regardless, internally strong dictatorships always impose higher internal taxes than do democracies, and thus dictatorships generally impose more overall taxation and inefficiency than do democracies. With these results in hand, we will consider two empirical examples.

Empirical Examples

The theory's concepts of sensitivity and integration are necessarily quite abstract, as are the propositions that flow from them. But this abstraction leaves open two

questions about the empirical applicability of the theory. First, what do sensitivity and integration actually mean when applied to real economies? Second, and related, even if the theory is correct as stated, are the effects it predicts large enough to actually explain major differences in international behavior? Here, we will discuss two empirical examples that, according to the theory, should have substantial differences in the extent of coercion and so serve as a concrete illustration of the theory's concepts. I will argue that the difference in the theory's predicted effects is large enough to explain the observed difference in behavior between the two examples. The conjecture advanced here is that the deep interstate comity discussed in Section 2 is substantially explained by the high sensitivity, thorough integration, and democracy of those states. The paper's title derives from this conjecture: it is among these states that the modern economic peace has arisen.

Example 1:

Imagine a group of states, each of which has the following characteristics. The bulk of the state's economy is composed of subsistence agriculture, because the productivity of most of its citizens is not high enough to generate much of a surplus above survival requirements. In part because of this, but also because of relatively high costs of trade arising from poor shipping technology and imperfectly controlled banditry on trade routes, the state engages in very little international commerce. It is ruled by an autocratic leader, who has amassed considerable power over the citizenry and is interested mainly in bolstering his own wealth and power.

This description fits most states around the world for most of human civilization. Though it is still approximately true of some states today, these properties were more

common before the 19th century (when some economies shifted away from farming and into manufacturing), and most common outside the realm of various empires that could substantially lower the costs of trade (e.g., the Roman Empire via its control of the Mediterranean Sea and its coasts, or the British Empire via its control of the oceanic trade routes).

Under these conditions, both sensitivity and integration were very low. To see why, consider what an autocrat might tax in such a state. He could "tax" (that is, seize) land or labor, the principal factors of production. Given the absence of human capital or any abundance of undeveloped arable land, this taxation would not lead to a reduction in the supply of either factor because there are no investment decisions to be distorted. The autocrat could also tax crops, the principal good produced in the economy. These are easily requisitioned in-kind, so that the direct costs of collecting the tax are not high. Moreover, farmers cannot afford to work less in response to the tax, because their reliance on their own produce for subsistence implies that doing so would put their survival at risk. Thus, given the necessary military power, an autocrat would be able to raise substantial taxes, in the form of levies on crops or direct seizure of land or labor, without much reduction in the total value of the taxed economy.

What few costly effects such taxes would have would be unlikely to bleed over to an untaxed economy. Since the state engages in very little commerce, the taxed economy is largely independent of other economies, and thus these are likely to be unaffected by the autocrat's predation on the taxed economy. Thus, the autocrat's taxation is unlikely to reduce the value of any untaxed economy.

When sensitivity and integration are very low, the theory predicts that the incentives for coercion will be maximized. The exaction of tribute causes little in the way of distortions or externalities, and thus the costs of taxation impose few limits on what can happen in equilibrium. Moreover, since the states' governments are dictatorships, any costs to the domestic economy wouldn't be fully internalized by the tax authority anyway. Coercion should be rife, and the balance of power should closely match the balance of wealth.

More broadly, the theory has implications for the nature of relations among these states. They exist in a rapacious world, in which autocrats view each other with justified suspicion and engage in constant scheming to conquer more valuable territory. States give close attention to the balance of military capabilities, and jealously guard their own security. Wars occur from time to time as commitment problems or asymmetric information arise and become too severe. Alliances between states are purely Machiavellian, in the sense that they are motivated by opportunity and greed rather than comity. They are always subject to repudiation or betrayal as national self-interest dictates. There is no real "friendship" among states, and certainly no deep comity of the kind observed in Section 2.

Example 2:

Now imagine a very different group of states, sharing the following characteristics. Each state's economy is composed mainly of manufacturing and services. Much of the economy's total present value derives from investment and innovation in science, technology, and organization that increase the future value of the economy. The productivity of most citizens is much higher than the survival level. Each state

engages in extensive international commerce, with trade in goods and services forming a substantial percentage of the economy. Additionally, each receives large inflows of direct and portfolio investment from abroad, and its own citizens invest heavily in foreign ventures. Many of the firms doing business in the state are multi-nationals. Finally, its government is highly representative, with leadership selected by vote of the majority of the citizenry.

Beginning at least by the Industrial Revolution, some states around the world shifted away from the characteristics of the first example and toward those listed here. In recent decades, the latter apply most clearly to the United States, the states of the European Union, Canada, Australia, New Zealand, Japan, Iceland, and Norway. And these characteristics are increasingly approximated by other states around the world as their economies develop and their polities liberalize.

For these states, both sensitivity and integration are very high. Because their economies are diversified and relatively free, it is relatively easy to avoid or evade any specific tax. To constrain avoidance and evasion, any significant level of taxation must therefore be collected from a broad base. Because citizens' survival is generally not at risk, they can respond to such broad taxes by working and/or investing less. This decreases the value of the economy in any given period, but it also generally reduces its growth rate, because citizens put less work effort or less investment into generating innovation. Any reduction in the growth rate results in an exponentially increasing loss of value to the economy over time, so that even a quite small decrease in growth has very large costs. Thus, the costs of the distortions associated with any significant level of tax are substantial.

These effects are not likely to be contained within the taxed economy. Because these states are heavily involved in international commerce, there are many channels by which the effects of taxation can bleed into other, untaxed economies. Employees and shareholders of firms that export goods to the taxed economy would be hurt by reduced consumption of these goods due to reduced wealth in the taxed economy. Consumers that buy products imported from the taxed economy would have to pay higher prices or turn to alternative, next-best suppliers of these goods and services. Investors would have a reduced set of opportunities in which to invest and reduced capacity for diversification. Firms that own foreign affiliates or purchase and sell products abroad would share in the taxed economy's costs. Fewer innovations in the taxed economy would be available to spread to and increase the value of the untaxed economy. And all these externalities would compound over time due to the lower growth rate of the taxed economy. Thus, the costs of the externalities suffered by the untaxed economy would also be substantial.

Both of these economic characteristics are widely recognized in these states. The importance of property rights and free, uncoerced participation in markets to the growth and well-being of society is fundamental to the prevailing understanding of the origins of prosperity in the "West." Elites within these states argue over the extent of public goods provision (and thus taxation to fund these goods) by the government, but most agree that the effects of taxation on growth are central to the debate. And it is conventional wisdom that economic downturns in one or several of these states will negatively affect the others. Participation in all the institutions of modern international economic cooperation—the World Bank, the IMF, central bank

coordination—is motivated in part by the perception that helping particular states to accelerate growth and recover rapidly from economic crises is beneficial to all the other states.

When sensitivity and integration are higher, the theory predicts that the incentives for coercion will be reduced, relative to the first example. The levying of taxes on one state's economy both substantially distorts that economy and also generates costly externalities in the taxing state's economy. Moreover, because the states discussed here are democracies, the costs of these externalities are more thoroughly internalized by the governments of these states. These distortions and externalities thus place hard limits on the size of transfers that can occur in equilibrium. Coercion should thus be less prevalent, relative to the first example, and the balance of power might deviate from the balance of wealth among these states.

However, even if the direction of the theory's predicted effect is correct, it remains to be seen just how large the effect might be for the states of this example. A thorough examination of this is well beyond the scope of this paper, but a simple numerical exercise is instructive about what magnitudes are plausible. Suppose that there are just two states in the group discussed in this example, A and B, not unlike the United States and the European Union (considered as a single state). State A is much more militarily powerful than B, sufficient to enforce the collection of any tax it chose to levy on B's economy, and is considering imposing a broadly-based tax at a rate of 10% on economic activity in B. The natural division of wealth gives equal value to the two states' economies, with an initial GDP of \$15 trillion, and the natural (i.e., untaxed) rate of growth of each is 3% per year, roughly comparable to the non-recessionary

values for the US and EU. Each discounts the future at 3% per year. Then, in the absence of any coercive taxation, the total present value of each economy is $\sum_{t=0}^{\infty}(1-.03)^t\cdot(\$15\text{ trillion})\cdot(1+.03)^t = \$15\text{ trillion}/(1-.97\cdot1.03) \approx \$16.7$ quadrillion.

If A levies the proposed tax, then B's economy will suffer a static loss—a year-on-year constant-fraction reduction in value—of just 1%, and the growth rate of B's economy will decline by a certain fraction γ. For its part, A's economy will suffer costs equal to only one-fifth of those borne by B's. So, under the tax, A's economy will generate a total present value of $\sum_{t=0}^{\infty} .97^t \cdot (\$15\text{ trillion}) \cdot (1 - .2 \cdot .01) \cdot [1.03 \cdot (1 - .2\gamma)]^t \approx \$15\text{ trillion}\cdot.998/(.0009 + .19982\gamma)$. A will get to consume this value, but also the total present value of the revenue from the tax, which is $.1\cdot\sum_{t=0}^{\infty} .97^t\cdot(\$15\text{ trillion})\cdot(1-.01)\cdot[1.03\cdot(1-\gamma)]^t \approx \$1.5\text{ trillion}\cdot.99/(.0009+.9991\gamma)$.

Should A impose the tax? Since A is a democracy, all that matters is whether the wealth its citizenry receive in the natural (no tax) division exceeds what they would get from the taxed division and the revenue from B. Obviously, that depends on γ. If it were (implausibly) zero, then the tax would not alter either economy's growth rate and would barely affect either's per-period value, and imposing the tax would simply make A's citizenry almost 10% richer, leaving them with a value of \$18.3 quadrillion. If it were (implausibly) one, then the tax would completely preclude any growth in B's economy, which would drastically reduce A's growth rate as well, and leave A's citizenry vastly poorer in the long-run, with a value of only \$76 trillion—less than 1/200th of the wealth they would have if A didn't tax B.

A little arithmetic leads to the remarkable result that, even if the proposed tax would reduce B's growth rate by a mere .04% (decreasing it from 3% per year to

2.9988), *A* would be better off not imposing it. Under the tax, *A*'s citizenry would receive only $16.4 quadrillion. Because the costs of taxation ramp up as the tax rate goes up, if it on net costly for *A* to impose a 10% tax, then a higher tax would be even worse for *A*. And though a smaller tax might be profitable, at some point the amount being transferred is negligible.

The lesson here is that any significant reduction in the taxed economy's growth rate—even one so tiny that it would likely be undetectable by economists—renders taxation undesirable, because the losses mount up so quickly over time that they overwhelm the tax revenue. This, regardless of whether a state has the power to impose the tax. In growth-based economies such as the ones discussed in this example, it is plausible that any non-negliglible rate of tax would induce significant, though tiny, reductions in growth rates. Thus, it is also plausible that the levels of sensitivity and integration seen in these states, together with their democratic governance, are sufficient to virtually *eliminate* the incentives for their governments to engage in interstate coercion. In other words, the theory's predicted effects could be so large empirically that coercion among these states *never occurs.*

If this is so, then the group of states discussed in this example should be relatively harmonious. Their governments should trust one another, and generally believe in and abide by the principle of "what's good for them is good for us." Mind you, the economic interests of these states are still opposed in principle, because each citizenry wants more wealth, and at any given time there is only so much wealth to go around. But their interests are compatible in practice, because taxation to transfer wealth is net costly for the side imposing it, so that the only way for anyone to get richer is

through growth, and integration ensures that this enriches everyone. Thus, all states can agree that each and every state should focus on its own, organic growth.

I conjecture that this theory of the "modern economic peace," so named because it is theorized to derive mainly from the high sensitivity and integration common to what are conventionally termed "modern" economies, explains the deep interstate comity observed in Section 2. Among the United States, the European Union, Canada, Australia, New Zealand, Japan, Iceland, and Norway, in particular, the balance of power sometimes differs radically from the distribution of wealth because the former is irrelevant to the latter. The US has no desire to revise the distribution of wealth in line with its predominant power, because the taxation necessary to do so would on net reduce the economic well-being of the US citizenry. Alterations in the distribution of wealth are driven by underlying economic fluctuations, not interstate coercion. The US has nothing to lose if the other states get stronger, because they will not use this power to revise the distribution of wealth, and it gains more powerful allies against external threats. This is why it contributes to the empowerment of these states. Just the same, these states rely heavily on the US for their security, because they know it has no reason to take advantage of them. Quite the opposite: they needn't even pay for protection, because the costs the US would suffer if they were coerced by an external power—deriving from the sensitivity of their economies and their integration with the US economy—are enough to motivate it to defend them for free. When these states argue, it is not over the balance of wealth, but instead to do with the value of, and best way to provide, international public or club goods, such as counter-proliferation and climate stability. Debate and persuasion resolve these

arguments, not force. In the rare cases when threats and punishment (e.g., referral to the WTO) are used, they are always mild and do not disturb otherwise warm relations, because both sides are mindful of the overwhelming commonality in their general interests. In particular, wars among these states are unthinkable, because they would not accomplish anything meaningful.

I will close this section with a brief summary of the theory. The incentives for states to engage in coercion depend on three factors: first, the degree to which taxation will cause distortions in the taxed economy that lessen its value, called sensitivity; second, the degree to which taxation will cause externalities in the untaxed economy, called integration; and third, the degree to which the government imposing the foreign tax internalizes its externalities on the the domestic economy. The costs of taxation narrow the size of transfers that are profitable to the taxing government, and so constrain the transfers that can occur in equilibrium. Insensitive, unintegrated dictatorships have the strongest incentives to engage in coercion when their power permits, because the resulting taxation imposes few costs on the taxed or their own economy, and the government partly ignores the costs to its own economy because it does not care about its citizenry's prosperity. Sensitive, integrated democracies have the least incentives to coerce each other, because the concomitant taxes are very costly to both economies and are fully internalized by a government that will be held accountable by its citizenry. Most countries for most of history were of the first type, and so coercion and war were rife and interstate relations watchful and marked by greed. Since the Industrial Revolution, those countries that have come to rely on growth and commerce to increase prosperity have few if any incentives to coerce each

other and so have harmonious relations. This explains why some nations get along so well, the puzzle that served as the principal motivation for this paper, but it also has the potential to answer other long-standing empirical questions. We turn to these next.

3.5 Some Implications of the Theory

If the theory of the modern economic peace is correct, then it suggests new answers for several standing empirical questions. First, it offers the potential for a new account of the origins of different regime types and their frequencies in different historical eras. It also promises an explanation for the observed long-run decline in organized violence, as well as the rise of international economic liberalization in the industrial era. Finally, it offers the possibility of a deep connection between rationalist and constructivist theories of international relations.

We will discuss each of these in turn, but a disclaimer is merited. The point of this discussion is not to argue for the correctness or even probability of the conjectures described therein. Developing and testing these derivative theories would require many more papers. Instead, the point is to show the versatility and potential power of a well-developed theory of conflicts of interest, such as this paper has endeavored to begin. Disclaimer issued.

Economic Origins of Dictatorship and Democracy (Redux)

As modeled in this paper, regimes differ only in the extent to which they serve themselves versus their citizenry, and the sole implication of this difference is that the

"democratic" regimes impose the least domestic taxes possible, and fully internalize the domestic externalities caused by any foreign taxes they impose. Regime type itself is taken to be exogenous and independent of the domestic economy's sensitivity or integration with other economies. But what if these economic characteristics also affected the viability of dictatorship versus democracy?

In the model developed here, the citizenry of a state is a unitary actor, and so it is intuitive to think of a democracy as attempting to maximize the (single) citizenry's well-being. More realistically, the citizenry can be broken down into groups—rich and poor, different ethnicities, and so on. The situation among these groups is similar to that among states: there is a distribution of political (and possibly military) power, and a distribution of wealth, and the issue is whether comparatively power-advantaged, wealth-deprived groups will use coercion to improve their lot. If, say, two groups within a state had high incentives to engage in coercion, then a political system with a relatively egalitarian distribution of power, such as democracy, would be inherently unstable. Each group would look for an opportunity to seize power and extract wealth from the other. This contest for supremacy could easily lead to a coup or civil war, and the subsequent replacement of democracy with the elevation of the victorious group into power over the other. Thus, in the presence of high incentives for coercion, there is no strong common interest between the groups for the government to serve and any non-discriminatory government such as democracy is unlikely to survive. Dictatorship, on the other hand, is more stable because it entails a concentration of power in one group, which can more easily resist any attempt by the subjugated group to gain dominion.

By contrast, suppose that groups within a state had low or no incentives to engage in coercion, because of high sensitivity and integration between their associated economies. Much like the states discussed in the second example in the previous section, these groups would have a strong common interest in preserving the natural distribution of wealth, so that the surplus thereby created could be enjoyed by all. In this environment, a dictatorship is inherently unstable. Because the dictatorship does not fully internalize the domestic costs of taxation and so imposes higher taxes and higher costs than would a democracy, the citizenry would have a common interest in replacing the dictatorship with a less inefficient democracy, freeing up the surplus from reduced taxation. The democracy itself would be stabilized by the presence of a strong common interest in preserving this surplus.

This answers a puzzle in the study of democracy and dictatorships that has been made ever starker by the flood of empirical analyses suggesting that democracies grow faster, become richer, and have a variety of other advantages over dictatorships. Namely, if democracies are so great, why are they so rare prior to the last two centuries? It's not as though the ideas of direct and representative democracy were discovered only with the writing of the US Constitution. And yet, since the rise of agriculture and settled civilization, most states have been autocratic. The answer suggested here is that, prior to the Industrial Revolution, the incentives for states *or intra-state groups* to engage in coercion were high, so that democracy was less stable than autocracy and so less prevalent. With the advent of industrialization and growth-based economies, dictatorships grew more and more inefficient relative to democracies, and began to be replaced from within by democracy.

The Long-Run Prevalence of Violent Conflict

If the origins of dictatorship and democracy are not independent of sensitivity and integration, but rather consequences of them, then the theory described in this paper has entirely economic foundations. In particular, the observed long-run decline in organized violence, whether intra- or interstate, is only proximally explained by political factors. Deep comity, whether intra- or inter-state, is at its roots driven by the changing economic characteristics of sensitivity and integration that render democracy more desirable as a system of governance, and as democratization spreads, the full internalization of costly externalities that it brings only speeds the trend toward comity.

As a side note, this is potentially a very broadly applicable theory. If individuals' efforts to maximize their welfare leads to technological progress that increases productivity, specialization, and trade, and this process reaches a "take-off" point where productivity rises fast enough, then the spread of egalitarian political systems and political comity are not far behind. This description applies as well to micro-organisms as it does to humans, and thus this is a theory, not of comity and conflict among states or groups, but of cooperation and conflict among any population of evolving entities. And it is an altogether different theory from those currently prevalent in theoretical biology, based on populations playing prisoner's dilemmas where the issue of how to divide the surplus from cooperation (peace, in the human context) does not arise, and conflict is never more than a fleeting temptation to defect away.

The Liberalization of the International Economy

The prevailing treatment of many topics in international political economy is to view costly barriers to trade, investment, currency exchange, and so on as responses to intra-state differences in economic interests. For example, import-competing firms lobby for protectionist tariffs, even though these reduce the value of the domestic economy. From the point of view of the theory developed here, these are just non-military instances of coercion, and their prevalence should respond to sensitivity, integration, and democratization in the same way as the prevalence of the exaction of tribute or the seizure of territory. This resolves a puzzle in the history of international economic relations, similar to the one already discussed in the history of democratization. If international economic openness is so beneficial, then why are barriers to international commerce so prevalent prior to Britain's liberalization in the 19th century? One answer commonly given is that, as economic surplus and intra-state specialization grew and the costs of trade fell, the potential gains from trade increased, and thus the incentives to reduce barriers grew. But this is not fully convincing: if the gains from commerce went up, but nothing else changed, the incentives to coercively alter the natural distribution of these gains also should have increased, and we should see more barriers, more trade wars, etc. But if sensitivity went up at the same time (as it did, since the advent of liberalization coincided with the spread of the Industrial Revolution), then the incentives for coercion might have on net decreased. The subsequent relaxing of trade barriers would result in increased integration, which would spur the process onward.

Material Interests and Moral Sentiments

Finally, though this paper has focused on rationalist theories of international behavior, there are non-rationalist explanations for deep interstate comity. In particular, constructivists argue that these nations simply hold different conceptions of international politics, in which material interests and power politics give way to moral sentiments guided by norms of cooperation and mutual assistance. These states have constructed identities for themselves as moral actors, and thus coercion and war are simply out of the bounds of appropriate behavior for them.

The question, as with the other empirical puzzles discussed in this section, is, if we assume the constructivists are right, then why is it that cooperative conceptions and moral identities did not arise earlier? After all, the behavior that appears to result from these conceptions is, in the aggregate much less costly than the politics, red in tooth and claw, that animated earlier states. Why was the social construction of peaceful relations preceded by thousands of years of savagery?

The answer suggested here is that the essential moral sentiment of "what's good for them is good for us" and the concomitant belief in the value of collective security and mutual assistance did not win out—and could not have spread—until the economic conditions of some states began to align their material interests with their moral sentiments. Once sensitivity and integration rose in some states, the incentives for coercion among them fell, and conditions became propitious for the rise of an intellectual movement in support of moral governance and a more pacific international community. In turn, the spread of these ideas would have further decreased the incentives for coercion, as this would now entail moral as well as material costs, and

accelerated the process.

Fleshing out the answers the theory of modern economic peace suggests for the empirical puzzles discussed here would seem an important directions for future research. Together, these extensions of the theory comprise a program for developing a more general theory of the origins of conflicts of interest, and the roles the presence or absence of such conflicts play in structuring domestic and international politics.

Chapter 4

A Model of Arms Proliferation and Prevention, with Muhammet Bas

Abstract:[1] We develop a formal model of bargaining between two states, where one can invest in developing nuclear weapons and the other imperfectly observes its efforts and progress over time, and use it to analyze the occurrence of proliferation and war, the viability of non-proliferation agreements, and the role of intelligence-gathering and estimates. We show that surprise proliferation, sporadic crises over the uncertain progress of a proliferant's efforts, and "mistaken" preventive wars can all arise endogenously in the model. We find that much of the over-time variation in behavior is driven, not by exogenous factors like the costs of war and the effects of proliferation, but by stochastic elements such as when the proliferant's program will make progress and when the other state will discover this. Moreover, while exoge-

---

nous factors do influence the probabilities of war, proliferation, and non-proliferation, their effects are often counter-intuitive and non-monotonic. We also find that non-proliferation can be undermined by the possibility of a "better deal" once the proliferant's program has made progress, suggesting that some states invest in a program as much to secure a more favorable non-proliferation deal as to actually get nuclear weapons.

4.1 Introduction

Iraq, North Korea, Libya, and Syria are historical enemies of the United States and have all pursued nuclear weapons, a technology that could radically shift the balance of military power with the US in their favor. The US launched a decisive war in 2003 in part to prevent Iraq from ever obtaining nuclear weapons. But the US did not stop North Korea from doing so, despite long negotiations and at least one crisis in which war was threatened, in 1993. And the US neither attacked nor even threatened Libya and Syria with war as each pursued its own nuclear weapons program. What explains the radically different outcomes across these cases?

Moreover, in the cases of Iraq and North Korea, there was also substantial variation in the relationship between each and the US over time. The final outcomes of war and successful proliferation occurred only after a long period of negotiations and threats. During this period, the US intelligence community obsessively monitored each state's nuclear weapons efforts, and sporadic crises arose in which the US prepared itself for war, and sometimes even seemed on the verge of striking, only to end with a fading threat of war and continued negotiation. What explains this drawn-out

process and the occurrence and evanescence of crises?

Now the focus of many commentators on US foreign policy has shifted to Iran.[2] Will the US (or its ally Israel) attack Iran in order to halt its nuclear progress? Or can the two sides find a non-proliferation deal that would satisfy both? Should the US simply tolerate Iran's efforts? How will these decisions be influenced by the progress of Iran's nuclear efforts, and the estimates of this offered by the US intelligence community?

Of course, these problems are not new (Gavin, 2004, 2009/10). They have recurred since 1942, when the Soviet Union initiated a program to develop nuclear weapons and the United States consider preventive strikes in response (Trachtenberg, 1988). In fact, the questions for strategy posed here are general to any era in which a new military technology is developed by one or several countries and begins to spread to others. Nuclear weapons are a recent, consequential example, but the spread of firearms to pre-colonial societies contacted by European explorers is also believed to have caused or exacerbated numerous wars in the 19th century (Bas and Coe, 2012*a*). And this process may be ongoing—there is some evidence of a link between arms transfers to developing countries and civil conflicts there (Craft and Smaldone, 2003; Krause, 2004).

What explains the variation in the observed behavior of states during these episodes? We try to answer this question by analyzing a formal model in which two states bargain over disputed issues, while one potentially invests in and makes progress toward acquiring a new technology (nuclear weapons, for concreteness) that, once deployed,

---

[2]For recent examples, see Goldberg (2010); Kroenig (2012); Kahl (2012); Waltz (2012).

would increase its bargaining power, and the other imperfectly observes its investment and progress. Because weapons development and observation occur over time, the model enables us to understand and make predictions about the sources of empirical variation in behavior across both countries and time. It also allows us to study the conditions under which agreements such as non-proliferation deals might be stable.

Surprisingly, we find that much of the variation in behavior over time, and possibly also much of the variation across countries, is essentially due to chance. In the absence of a non-proliferation deal, the specific instantiations of two stochastic elements—a proliferant's halting progress toward acquiring nuclear weapons, and the US's noisy observation of its efforts and current stage of development—can make the difference between a final outcome of war or peace, prevention or proliferation, and also determine whether the road to this outcome is quick and calm or long and tense. These variables can lead to periods of slowly increasing apprehension about a proliferant's nuclear progress, peaking in crises which may end in war or merely a repeat of the cycle. Because, at an early stage of progress, there is little reason for the US to worry about a proliferant's efforts, the random elements also make it possible for the US to be surprised by a state's proliferation. In short, these variables can easily have as large an impact on the outcome as more obvious factors such as the anticipated costs of war and effects on the balance of power of a state's acquisition of nuclear weapons.

War in the model is always preventive, but it can occur in two different situations. First, the US might observe that the proliferant has reached a late stage of progress, so that it is very close to acquiring nuclear weapons and not much time remains

for the US to prevent this. Second, even in the absence of any hard intelligence, the US might become so suspicious that the proliferant has made it to a late stage that it felt it could not wait for confirmation. This latter situation may lead to "mistaken" wars, in the sense that US suspicions would turn out to be wrong after the war, and the probability of a war being mistaken can be quite high for plausible choices of the parameters. Although improvements in US monitoring lead the US to bide its time and thus lower the probability of a mistaken war, they can increase or decrease the overall probabilities of war or proliferation, depending on whether they make possible an enforceable non-proliferation deal. In the absence of a deal, a proliferant that is expected to more quickly master the technological prerequisites of building nuclear weapons, whether by virtue of indigenous sophistication or outside assistance, will lead the US to be more suspicious. This makes the US more willing to attack even if it is less confident that intervention is required, and the probability of war and mistaken war will rise. Thus, counterintuitively, the probability of eventual proliferation is lower for a more sophisticated proliferant.

The existence of an enforceable deal strongly depends on the US ability to detect cheating. More surprisingly, an increase in the expected speed at which a proliferant will develop nuclear weapons can make a deal possible, because it may render the US threat to go to war if the proliferant is caught cheating on the agreement more credible. However, under other conditions it can undermine the existence of a deal because it increases the proliferant's temptation to cheat. If an enforceable deal at a late stage of progress exists, non-proliferation may be impossible to enforce at an earlier stage, and the existence of a later deal may even undermine an earlier one

and raise the overall probability of proliferation. Remarkably, a "proliferant" may see securing a better non-proliferation deal, won by the progress it has made in its nuclear program, as a possible benefit that can motivate its early efforts as much as the prospect of actually obtaining nuclear weapons.

Our work joins a spate of recent formal analyses of the strategic problems inherent in states' arming and observation of one another's militaries, several of which are focused on nuclear proliferation in particular. Our model is the first to simultaneously allow: bargaining over the disputed issues; the interaction to repeat indefinitely over time; the choice to arm to be endogenous; and the observation of these choices to be imperfect. This combination of features is required to study over-time variation in these interactions, and especially the role of intelligence gathering and estimates therein. We also uniquely do not restrict equilibria to Markov strategies, which rule out the possibility of arms control (e.g., non-proliferation deals) based on the threat of future punishment, unlike other analyses of over-time arming, such as Fearon (2011*a*); Jackson and Morelli (2009*b*); Powell (1993). With respect to nuclear weapons or any other capability that requires substantial time for research and development before deployment, ours is the first to allow development to progress over time and to allow observations to be made of both the choice to proliferate and the state of a proliferant's program. Empirically, these features appear central to understanding the over-time variation in the interaction between the US and particular proliferants, and we will show that they may also be important to understanding the variation in final outcomes across proliferants.

The downside to studying this new and empirically rich combination of features

is that we must abstract away from other important aspects of this complicated phenomenon. We ignore the possibility of costly, non-game-ending alternatives to war, such as containment or limited preventive strikes; these are analyzed in Coe (2012). We do not allow for asymmetric information about the preferences of the actors, nor do we endogenize the decision to reveal or conceal weapons efforts, as Baliga and Sjöström (2008) does. Asymmetric information in our model arises, not from the use of mixed strategies of uncertain empirical plausibility, as in Debs and Monteiro (2010); Fearon (2011a); Jackson and Morelli (2009b); Meirowitz and Sartori (2008), but from the inherent stochasticity of mastering nuclear weapons technology. We do not allow for private signals of the proliferant's weapons program, but assume the signals are publicly, credibly revealed; Coe and Vaynman (2012) analyzes the effect of public versus private signals on non-proliferation deals. Finally, we abstract away from the issue of the direct (budgetary) cost of a nuclear weapons program, taking these to be negligible; the cited studies all allow arming to be costly.[3] Finally, we eschew two other common features of these models due to their not being relevant, in most cases, to the particular phenomenon of nuclear proliferation. First, we assume that proliferation is one-sided, since the US already has nuclear weapons; Bas and Coe (2012a) analyzes two-sided proliferation. Second, we assume the choice to invest in nuclear weapons development is all or nothing: states do not try half-heartedly

---

[3]We have analyzed an extension of our model in which a weapons program is costly. The results reported here are qualitatively unchanged, but some new results arise. The costs of a nuclear program make it easier to enforce a non-proliferation deal since the proliferant would, all else equal, like to avoid those costs. But they also increase the probability of war in the absence of a deal, since a costly program eliminates some of the surplus of peace over war. These results are available from the authors upon request; they are omitted here due to space constraints, as well as the desire to make the points demonstrated herein as clear as possible.

to proliferate. The cited studies that are not specific to nuclear weapons allow for a choice of levels of arming, consistent with their principal goal of modeling conventional military buildups.

The next section describes the elements of the model in more detail, as well as the assumptions we make. Section 3 analyzes equilibria featuring proliferation and/or war. Section 4 studies equilibria featuring a non-proliferation agreement. Section 5 concludes and offers implications for further theoretical and empirical research on proliferation. Proofs of all the propositions, as well as specifications for the simulations we run for part of the analysis, appear in the appendix.

4.2   Setup of the Model

We model the interaction between two states, A and B, which we will sometimes refer to as "the US" and "the proliferant," as they bargain over revisions to a prior division of a composite of disputed issues, represented by the unit interval. In the first of infinitely many discrete periods of time, A first chooses whether or not to start a war with B. If A attacks, the game ends with a costly lottery. The value of this lottery to each player depends on the balance of military power between them, represented by A's probability of victory in the war. The winner receives the entire contested stake in this and all future rounds; the loser gets nothing. Regardless of who wins, each player pays a positive cost of war, c_A and c_B respectively, in this and all future periods.

If A chooses not to attack, then he must make a take-it-or-leave-it offer to B of a revision to the status quo disposition of the contested issues. If B rejects the offer,

war results, in which case the game ends with the same costly lottery. If he accepts the proposal, the revision is implemented immediately and the associated payoffs are realized.

Peaceful acceptance by B of A's offer is followed by an opportunity for B to invest in developing a new military technology, "nuclear weapons" for concreteness, which we assume A already possesses. To simplify the analysis, we assume that B's development effort is all or nothing—the choice to pursue nuclear weapons is binary. Our focus is on the choice of whether to develop the technology and the possibilities for war or agreements to prevent this development; we are less concerned with a state's optimum choice of the precise level of resources that should be invested.

B must master a series of technological prerequisites before he can actually deploy nuclear weapons. These might include such hurdles as enriching uranium to a sufficient degree, re-processing plutonium from spent reactor fuel, building a viable implosion device, and reducing a warhead to deliverable size. For simplicity of presentation, we assume there are only two prerequisites, so that there is a first stage of development labeled s_1 where B has mastered neither, a second stage s_2 where B has mastered the first prerequisite, and a third stage n where has mastered both and is assumed to possess nuclear weapons. However, the results can be easily (but tediously) extended to any finite number of stages. We assume that B begins the game in s_1, and that this is common knowledge.

The overcoming of these hurdles is partly a result of trial-and-error, so that the time at which B will master one and then the next cannot be perfectly predicted by either player. If B begins a round in s_1, and chooses to invest in that round, then he

advances to s_2 in that round with probability $\epsilon \in (0, 1)$, goes all the way to acquiring nuclear weapons in that round with probability $\epsilon \lambda \in (0, 1)$, and remains at s_1 with probability $1 - \epsilon$. If B begins a (later) round in s_2 and invests, then he advances to acquiring nuclear weapons in that round with probability $\lambda \in (0, 1)$, and remains at s_2 with probability $1 - \lambda$.

This representation of the weapons development process is the central analytical innovation of this paper; many of our results flow from it. It can be thought of as the simplest possible representation of the empirical fact that the development of any complex technology is both progressive and stochastic. Although B's chances of advancing to a given stage depend only on his current stage and his decision to continue trying, his probability of acquiring nuclear weapons will increase over time, and in the absence of good intelligence, A's *estimate* of his time to acquiring nuclear weapons will decrease over time.

If B's development effort is successful and he acquires nuclear weapons, then the balance of power in the next period (A's probability of victory in war) shifts. Before B has the weapons, the balance of power is p; after, the balance is p_n. Naturally, we assume that $p > p_n$, so that having nuclear weapons in war is better than not. We also assume that B's successful acquisition of nuclear weapons immediately becomes common knowledge (e.g., because of a successful and easily observable test detonation).

The first period ends after B's progress or lack thereof is determined. The next period, and every subsequent period, differs in structure from the first only in that it begins with two signals, which are received by A but assumed to be common

knowledge. The first signal indicates whether B invested in the last period or not; the second indicates B's current stage of progress. If B did invest in the previous period, then with probability $\tau > 1/2$ A receives a signal that he did, and with probability $1 - \tau$ A receives a signal that he did not. If B did not invest, then A receives a signal that he did not with probability 1. Thus, A's intelligence on B's investment is noisy, but for simplicity there are no false positives. A will receive a true signal of B's current stage with probability σ, and a "null," uninformative signal with probability $1 - \sigma$. Thus, A's intelligence on B's progress is spotty, but accurate. We will show later that allowing for false positive signals of investment, or false signals of stage, would complicate the analysis but not qualitatively alter the results.

Each player's utility is assumed to be linear in his share of the value of the contested issues. Settlements are labeled by the share going to A: the settlement in which A receives q and B receives $1 - q$ is called q. The players are assumed to discount future consumption at a common rate of $\delta < 1$ per period. Players' preferences and all the exogenous parameters of the game are common knowledge.

4.3   Proliferation and War

We use backward induction to find the Perfect Bayesian Equilibria (PBE) of the model. We will first show that the subgame in which B has acquired nuclear weapons has a unique equilibrium outcome. We then consider earlier subgames, restricting ourselves to equilibria in which there is no non-proliferation agreement, called "no-deal" equilibria; these agreements will explored in Section 4. First we analzye the prior subgame in which B has reached the second stage and A knows this (because

of an earlier signal that B was in the second stage). We then consider the "initial" subgames in which B is not known to be at the second stage.

B Has Obtained Nuclear Weapons

We begin with the subgame where both players have nuclear weapons. Proposition 4.1 gives the unique equilibrium of this subgame:

Proposition 4.1. *Suppose that B has acquired nuclear weapons. In every period, A will offer $q = p_n + c_B$, and B will accept $q \leq p_n + c_B$. No war will occur.*

The intuition here is that once B has acquired nuclear weapons, there will no further shifts in the balance of power and no further cause for B to invest, and any further intelligence signals will be irrelevant. Thus, A does best by giving B just enough to cause him to prefer peace to war, and A keeps the rest for himself. Even though A holds B to this minimum, B still benefits from having the weapons. Because they shift the balance of power in his favor and thus raise his value for war, then to avoid war, A must give him a larger share than he would get if he didn't have the weapons. Thus, the essence of the interactions in earlier subgames, before B has nuclear weapons, is that B has an incentive to pursue them while A has an incentive to prevent B from getting them.

B Is Known to Be in Second Stage

Now back up to earlier in the game, when B has not yet acquired nuclear weapons, but is known with certainty by A to be in the second stage—because at some previous point in the game, A received a signal of progress that B was in the second stage. We

will show that, in the absence of a deal, B will always invest in a nuclear program. We also analyze the conditions under which A will either tolerate this program or go to war to stop it, and show that war can occur due to B's inability to commit either to avoid taking advantage of weapons once he has them, or to pursuing them when he does not.

A "no-deal" equilibrium is defined as one in which A's equilibrium strategy does not depend directly on whether B invested in any previous period and the signals A receives thereof. By contrast, a "deal" equilibrium is one in which A reacts directly to whether (he observed) B investing in the past, as such investment can be thought of as cheating on a (possibly implicit) deal. Proposition 4.2 characterizes the no-deal equilibria when B has reached the second stage, and this is known to A.

Proposition 4.2. *Suppose that B is in the second stage and this is common knowledge. If:*

- *$p - c_A + \frac{\delta\lambda}{1-\delta}(p - p_n) < 1 + \frac{\delta\lambda}{1-\delta}(c_A + c_B)$, the unique no-deal equilibrium outcome is steady investment by B that is tolerated by A, resulting in eventual proliferation;*

- *$p - c_A + \frac{\delta\lambda}{1-\delta}(p - p_n) > 1 + \frac{\delta\lambda}{1-\delta}(c_A + c_B)$, the unique no-deal equilibrium outcome is immediate war;*

- *$p - c_A + \frac{\delta\lambda}{1-\delta}(p - p_n) = 1 + \frac{\delta\lambda}{1-\delta}(c_A + c_B)$, both types of no-deal equilibrium exist, but no others.*

To understand this result, consider B's perspective. In the absence of a deal, A will not reward non-investment, or equivalently, penalize investment, so it makes sense for B to invest as long as there is some benefit from acquiring nuclear weapons (recall

we assumed away any direct costs of a nuclear program). As Proposition 4.1 shows, once B is nuclear-armed, A will offer B just enough to make him indifferent between war and peace. Before B has nuclear weapons, he is weaker (that is, he expects to do less well in a war), and A will concede even less, because B will still prefer this to war and, in the absence of a deal, there is no reason for A to do otherwise. Thus, B can expect that acquiring nuclear weapons will bring bigger concessions from A, and so it *always* makes sense to invest in the absence of a deal.

Now consider A's perspective. In a hypothetical world without nuclear weapons, A would offer B just enough of a compromise on their disputed issues to make B indifferent between accepting the offer and going to war—there is simply no reason for A to be any more generous than that. From A's perspective, the problem with B pursuing nuclear weapons is that, once B has acquired them, A will have to make a more generous offer in order to avoid war. In the absence of a deal, A really only has two choices for what to do. First, A could attack B in order to try to prevent the latter from ever getting the weapons. Second, A could tolerate B's nuclear program. If A chooses toleration, then until B gets nuclear weapons, A can offer B even less than he would in the hypothetical non-nuclear world, because he can take advantage of B's expectation that the nuclear program will eventually be successful and lead to larger concessions from A. If instead A goes to war, there will be no proliferation and subsequent concessions to B, but A will no longer have the chance to peacefully exploit B's hopes, and will lose the surplus from peace (the avoided costs of war) he would have enjoyed even once B became nuclear-armed.

Proposition 4.2 specifies this tradeoff for A. The left-hand side of each condition

represents the immediate payoff of war for A $(p - c_A)$ and the gain from avoiding the change in concessions A would have to make in perpetuity after B had obtained nuclear weapons, weighted by the probability of B's program succeeding in the current period $(\frac{\delta \lambda}{1-\delta}(p - p_n))$. The right-hand side represents, in its first term, the immediate payoff of toleration: 1 is the most A could get in this period if B was willing to accept zero concessions for now in order to avoid war. Its second term represents the surplus A will enjoy from peace if B's investment this round is successful $(\frac{\delta \lambda}{1-\delta}(c_A + c_B))$. If the left-hand side is larger, then A is better off tolerating B's program and taking advantage of B's hopes in the meantime. If it is smaller, then A is better off going to war to stop it before it is successful.

The root cause of war, when it occurs, is a set of linked commitment problems. Because, in the absence of a deal, B always has something to gain from acquiring, and therefore developing, nuclear weapons, B cannot commit either to not taking advantage of the weapons once acquired to extract more concessions from A, or to not pursuing them. Because this shift in power will be disadvantageous for A, B will have to be willing to agree to less concessions while the nuclear program is ongoing in order to satisfy A and avoid war. But B cannot accept anything less than zero concessions; he cannot yield any more than the whole set of disputed issues at a given point in time. If the shift resulting from B's proliferation is large enough, this will not suffice to avert war.

War is more like to occur when A is stronger (p is higher). Then there is less for A to gain from taking advantage of B while his program is ongoing, and more to lose when it bears fruit. War is also more likely the smaller the costs of war; there is less

surplus to lose from war, and if c_A is smaller there is also less for A to gain from taking advantage of B prior to proliferation. More surprisingly, the effect of the probability that B's program is successful in a given period (λ) is not always to raise the likelihood of war. One might think that, the more likely B's program is to succeed, the sooner A will expect B to obtain the weapons and extract a better offer, thus making the commitment problem more severe and war more likely. But this only happens if the shift in power from proliferation exceeds the costs of war ($p - p_n > c_A + c_B$). When instead $p - p_n < c_A + c_B$, the higher likelihood of proliferation is outweighed by the increased advantage A can take of B's more optimistic hopes for success while the program is ongoing, and thus the incentives for war relative to toleration decrease.

Next we turn to the previous subgames, in which B may or may not have reached the second stage, but this is not known with certainty by A.

B Is not Known to Be in Second Stage

We divide the analysis of earlier periods in the game into two parts. First we will consider what might happen, if once B has reached the second stage and A knows this, there will be toleration rather than war. We will show that this means that war will never happen in equilibrium, and in the absence of a deal, proliferation is inevitable. Second, we will examine what might happen at earlier times if the discovery by A that B is at the second stage would lead to immediate war. We will show that, without a deal, A will tolerate B's investment initially, but in the absence of contrary evidence will grow increasingly suspicious that B has covertly reached the second stage. If A becomes sufficiently convinced B has done so, there will be war.

Whether this happens, whether the war is "mistaken," and the character of the road to war or eventual proliferation will depend on the (stochastic) information that A's efforts to monitor the progress of B's program produce.

Proposition 4.3 shows that, if A is willing to tolerate B's program when it is known to have reached the second stage of development, then he will tolerate it earlier as well. In this case, there is no chance of war occurring in equilibrium: B will pursue nuclear weapons and eventually obtain them, without violent interference from A.

Proposition 4.3. *Suppose that war cannot happen in any equilibrium when B is in the second stage and this is common knowledge. Then steady, tolerated investment is the unique no-deal equilibrium of the game.*

There is a simple intuition for this result. When A thinks B might still be in the first stage, then his estimate of the probability of proliferation in a given period is lower than if he knew B to have reached the second stage, because if B is in the first stage he would have to master *both* stages to get nuclear weapons. The lower the per-period probability of B obtaining nuclear weapons, the less severe is the commitment problem, because A expects the disadvantageous shift in the balance of power to occur later than it otherwise would. Thus A is more willing to tolerate B's program as opposed to attacking it, relative to the situation where A is sure the program is in the second stage. So, if A is willing to tolerate a known second-stage program, then he will find it even easier to tolerate a possibly less-advanced program. Given this, B will certainly invest, knowing that this will lead not to war but to better concessions from A. In the absence of a deal, eventual proliferation is inevitable.

What about when the discovery that B's program has reached the second stage of

development would lead to an immediate attack by A? How will this affect behavior earlier in the game? To answer these questions, we will assume that, should A detect B at the second stage, there will be war in equilibrium. That is, from Proposition 4.2, we assume that $p - c_A + \frac{\delta\lambda}{1-\delta}(p - p_n) < 1 + \frac{\delta\lambda}{1-\delta}(c_A + c_B)$. Also, we will once again exclude consideration of deal equilibria in which A's equilibrium strategy depends directly on whether B invested in any previous period and the signals A receives thereof. These will be studied in the following section.

Proposition 4.4. *Suppose that when B is in the second stage and this is common knowledge, immediate war is the equilibrium outcome. Then in the unique non-deal equilibrium of the game, B always invests, given the chance. In this equilibrium, A tolerates B's program until he receives either a signal that it has reached the second stage, or some $k \in \mathbb{N}_0 \cup \infty$ consecutive null signals of its stage, and then goes to war.*

Proposition 4.4 shows that this leaves only equilibria in which B always invests, given the chance. Faced with B's steady investment, A bides his time so long as he is sufficiently confident that B's program remains in the first stage of development. But when A detects that B's program has reached the second stage, or when A has waited long enough without intelligence on B's stage, then A attacks rather than wait any longer and bear further risk of proliferation.

To see how the latter can happen, consider A's perspective. If A receives a signal that B remains in the first stage, then he can assume that B's program is not going well and safely wait and see (unless the probability that B jumps from the first stage to nuclear weapons, or $\epsilon\lambda$, is so high that it makes sense for A to attack at the very beginning of the game). If he receives a signal that B is in the second stage, then

he can immediately attack and suffer no more risk of proliferation. But what if he receives no (a "null") signal of B's stage? Then A will not know for sure what stage B's program is in. Instead, he must estimate the probability that B has at some point covertly advanced to the second stage.

Corollary 4.1. *After receiving $i \in \mathbb{N}$ consecutive null signals of B's stage since the beginning of the game or the last signal that B is in the first stage, A's estimate of the probability that B's program has reached the second stage is $\frac{\sum_{j=1}^{i}(1-\epsilon)^{i-j}\epsilon(1-\lambda)^j}{(1-\epsilon)^i+\sum_{j=1}^{i}(1-\epsilon)^{i-j}\epsilon(1-\lambda)^j}$. If $\epsilon \geq \lambda$, then this probability converges to 1 as $i \to \infty$. If $\epsilon < \lambda$, it converges to $\frac{\epsilon-\epsilon\lambda}{\lambda-\epsilon\lambda}$.*

Corollary 4.1 specifies A's estimate after a given time without a stage signal. To understand this result, consider that in each period without a signal of B's stage, A must weigh two contradictory pieces of evidence. On the one hand, B has not yet gotten nuclear weapons, which suggests that in the last period his program was in the first stage rather than the second. On the other hand, time has passed and A knows B has been trying, so it is possible his program has mastered the first stage and moved to the second. Which of these weighs most heavily depends on whether it is easier for B to master the first stage or the second (i.e., whether ϵ is greater than λ). If the first stage is easier, then over time, in the absence of an informative signal of B's stage, A will eventually become almost sure that B has reached the second stage: intuitively, it is increasingly likely that B has mastered the easy stage but gotten stuck in the hard one. Once A has waited long enough without intelligence, and is confident enough that B has covertly reached the second stage, he attacks. If instead the second stage is easier, then even after a great deal of time, A will still not be sure which stage B's program is at: he could be stuck at either stage, and the
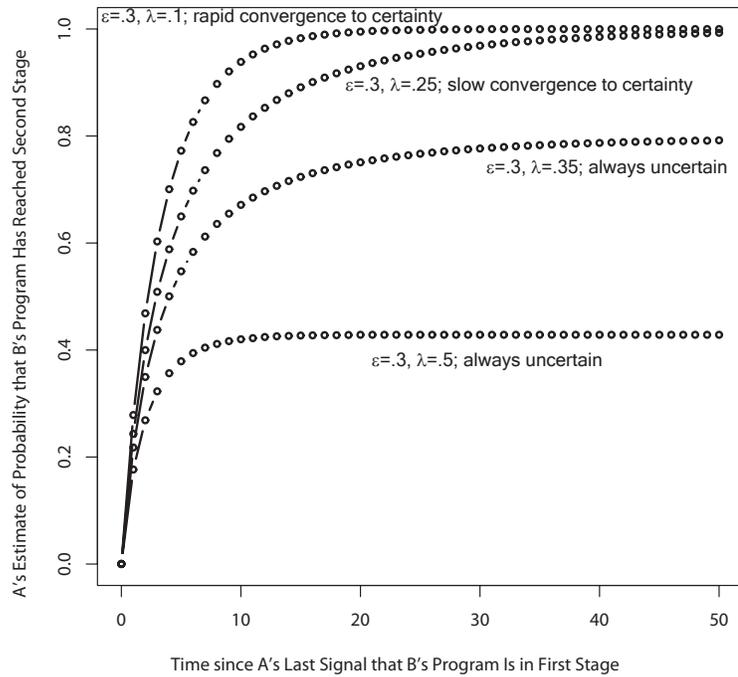
Figure 4.1: *A* grows more suspicious over time, but may never be certain

harder it is to master the first relative to the second, the more likely it is he remains at the first. Then it may be the case that although *A* is increasingly confident over time that *B* has reached the second stage, he may never be sure enough to justify attacking. Figure 1 illustrates the change in *A*'s estimate as he waits without new stage intelligence for several different combinations of ϵ and λ.

The relatively simple description of this equilibrium conceals the variation in behavior that can occur as it unfolds. This variation is driven entirely by the chance successes of both *B*'s nuclear program and *A*'s intelligence-gathering. As a result of these stochastic elements, the game can end peacefully or violently, and with the occurrence of proliferation or its prevention. The final outcome can be reached quickly

or slowly, and relations in the meantime can be pacific or crisis-prone. There are four generic kinds of paths along which the equilibrium might travel, described here and illustrated in Figure 2.

1. A tolerates B's investment for the few few years as it is unlikely to be successful soon. But B's program masters both stages of development unexpectedly quickly, and B acquires nuclear weapons. The process is calm and ends quickly in peaceful proliferation. This fits the pattern of the Soviet Union's nuclear weapons program from 1945 to 1949.

2. A is content to tolerate B's program initially, but receives intelligence that the program has advanced to the second stage and immediately attacks to stop the program. The process is quick and seems calm, but ends violently. This fits the period of Iraq's program in the late 70s and early 80s, and Syria's in the mid 2000s, when their completion of a reactor from which weapons-grade plutonium might be derived led to attack by Israel.

3. Lacking recent intelligence on B's program, A grows increasingly apprehensive about the prospect of its imminent success. A crisis arises and war occurs. The process is potentially long and ends violently. This might correspond to the period of Israel's program in the 60s, when Egypt attacked, perhaps because of fears that Israel nearly had the bomb.

4. Lacking recent intelligence, A's suspicions become persuasive and a crisis arises. War is threatened and appears imminent, but the arrival of intelligence that B's program remains in the first stage defuses the crisis, and the process continues.
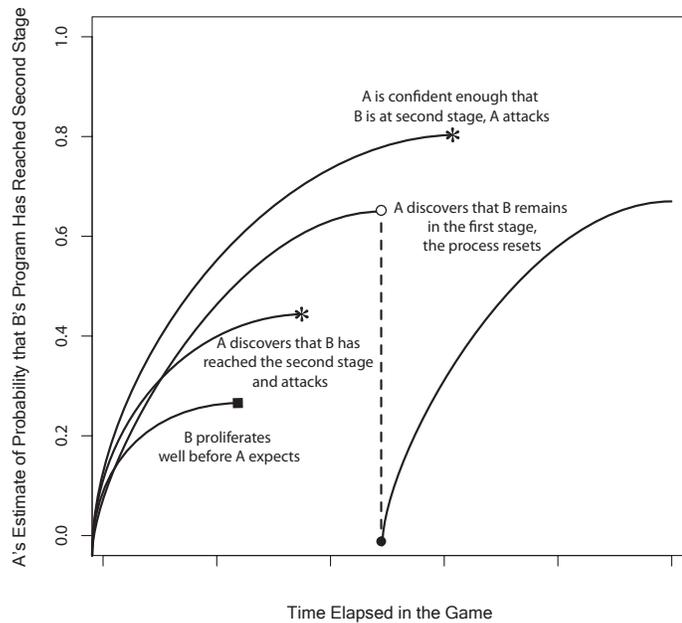
Figure 4.2: There are many possible paths through the equilibrium

The process is drawn-out, tense, and dangerous. This might fit the pattern of Iraq's suspected programs during the 1990s, when the US prepared for war until inspectors revealed that these programs remained rudimentary.

Notice that when war occurs in this equilibrium, its purpose is always to prevent proliferation, but there are two different situations in which it can happen. First, A might receive definitive intelligence that B's program has reached the second stage of development, leading A to attack. Second, A might lack reliable intelligence on B's program, but estimate that its likely progress meant that B would soon go nuclear, and attack based on this (uncertain) estimate. The common thread between the two is that, whether by virtue of positive intelligence or the passage of time, A has

become sufficiently confident that B's program has reached the second stage that he feels compelled to strike. The difference is that, in the first situation, A can be certain that it is better to attack now than to wait, whereas in the second, it might turn out that B's program remained stuck in the first stage so that, in retrospect, it would have been better for A to wait and defer the costs of an unnecessary war. Going to war based on an estimate, rather than reliable information, is rational for A but poses the risk that the war was a "mistake." To be clear, if this happens A is not mistaken about B's intentions or efforts—B is trying to obtain nuclear weapons, after all—but is wrong about the progress he has made and thus the imminence of his success. A might well regret such a war, even though it was rational.

Note also that the stochastic elements of the game (nuclear program progress and intelligence-gathering) determine which kind of path is actually observed and are therefore central to explaining the variation we see empirically. Because the exogenous parameters (σ, ϵ, λ, p, p_n, c_A, c_B, δ) are relatively stable over time, and all the other equilibria (as we shall see) display stable non-proliferation after some initial period, the model suggests that much of the over-time variation in behavior observed between the US and particular proliferants is driven by the stochastic elements. This might also be true for variation in behavior *across* proliferants, but the exogenous parameters do change substantially across cases and so may explain some of the cross-country variation. These features of the equilibrium carry important implications for the empirical study of proliferation, and the limitations thereof, to which we will return in the concluding section.

The reader might worry that these claims are an artifact of our simplified repre-

sentations of the development of nuclear weapons and the gathering of intelligence about it, but this worry is unfounded. Recall first that we assumed a program had just two stages to master prior to obtaining nuclear weapons. In fact, allowing for more than two stages would not alter our conclusions. All that matters is that there is a stage in which B could be discovered that would (in some equilibrium) lead to immediate attack by A, and a stage which, if observed, would lead A to wait and see. The presence of stages occurring before, after, or between these two will certainly complicate A's estimates and strategy. But the basic picture, of A waiting until his concern that B has reached a threatening stage of development becomes persuasive before attacking, remains unaltered.

Second, recall that we assumed that A's intelligence would suffer neither from false positive signals of investment nor from false signals of stage. Since, in the equilibria studied in this section, B always invests given the opportunity, and A knows this, the signals of investment A receives are actually irrelevant. If A receives a signal of investment, it only confirms what he already knows; if instead he receives a signal that B did not invest, he disregards it as erroneous. Obviously then, it does not matter whether the signals are true or false. However, A does pay close attention to signals of B's stage. If false negatives (A receives a first-stage signal when in fact B has reached the second) are possible, then the receipt of a first-stage signal will still decrease A's estimate of the probability that B has reached the second stage, but not all the way to zero. Crises of the kind that occur in the third and fourth paths listed above will still defuse if A receives one or more first-stage signals, but more gradually. And if false positives (A receives a second-stage signal when B remains in

the first) are possible, A might need more than one second-stage signal to arouse his suspicions to the point of attacking. Both types of false stage signal would increase the likelihood of mistaken war, because they make attacks without surety that B is in the second-stage more likely.

We will return to these generalizations of our model in the next section, but suffice it to say that all of them undermine the possibility of a non-proliferation deal because they erode the quality of the information that A can use to enforce a deal. They thus make it more likely that the non-deal equilibrium analyzed in this section is the only equilibrium. With these worries laid to rest, we can turn to the role played by the exogenous parameters of our model.

The Effects of the Exogenous Parameters

Although the stochastic elements of the game determine the particular path through the game that occurs, the exogenous parameters determine the probabilities of the different paths. Thus, these parameters can still have substantial effects on the *expected* behavior, and knowledge of these effects can serve two purposes. First, although it is hard to see how the US or a proliferant could affect the stochastic elements of the game, the exogenous parameters are potentially manipulable, at least over the longer term. So, as examples, the US might try to strengthen its intelligence-gathering capabilities (increasing σ), and a proliferant might attempt to secure outside assistance for its nuclear program (increasing ϵ and/or λ). Knowledge of the effects of the exogenous parameters would enable us to predict how changes in the relevant policies might alter the likelihood of different outcomes such as war and proliferation.

Second, these parameters vary substantially from country to country. Some potential proliferants are more technologically sophisticated than others, and the acquisition of nuclear weapons by some would have a larger effect on the balance of power than others, for example. Thus, if we wish to explain why the interaction between the US and Iraq worked out one way, but that between the US and China during the latter's nuclear program another way, then we need to know how differences in exogenous parameters might have contributed to different outcomes.

Unfortunately, the characterization of equilibrium given in Proposition 4.4 and Corollary 4.1 is not precise enough to allow us to derive comparative statics on these parameters. We require exact solutions for the equilibrium at each possible combination of parameter values, which in turn means solving for both the amount of time A is willing to wait before attacking (k) and the set of offers he makes to B on their disputed issues as he waits (labeled $\vec{q_k}$), which itself depends on k. The solution for k is governed by a trade-off. Waiting longer gives A more time to enjoy the surplus from peace, as well as more time for intelligence on B's program to come in, possibly revealing that B remains at the first stage, so that A needn't go to war after all. But it also exposes A to a greater risk that B's program will succeed, forcing A to offer better concessions. The solution for $\vec{q_k}$ is determined by A's desire to give B just enough concessions to avoid war while he waits, subject to the constraint that A can do no better in a period than to concede nothing to B (even if B would accept even less than this).

For technical reasons described in the appendix, there is no closed-form analytical solution for k and $\vec{q_k}$, and so we must instead find numerical solutions for a chosen set

of parameter combinations. A detailed description of the algorithm we use, as well as the statement and proof of some lemmas necessary to demonstrate the algorithm's validity, can be found in the appendix.[4] Here, we describe the particular set of parameter combinations that we chose to study and the results we obtain.

Our algorithm is computationally intensive, so the large number of independent parameters (eight: σ, ϵ, λ, p, p_n, c_A, c_B, and δ) forces us to confront the curse of dimensionality. Our choice of the range of values these parameters could take on was motivated by our judgments about what is empirically plausible, but our choice of the size of the intervals into which each range is divided was driven by the time available for computation. For each of the exogenous parameters listed below, we specify the range that was used and the size of the intervals. We also describe our rationale for the chosen range.

- δ: the players' discount rate, from 0.9 to 0.999 with an interval of 0.03 between values. To know what discount rates are plausible, we need to establish the length of a period in our game. In principle, B could be continuously making decisions about whether to start or stop investment, the mastery of a particular stage could come at any point in time, and A could be continuously negotiating offers with B, monitoring its investment and progress, and deciding whether or not to attack. In practice, there is bound to be some delay between A's receipt of a new intelligence discovery, his decision to attack, and the execution of the attack, even if the option under consideration is less than a full-on invasion. Focusing on the case of the United States, it seems difficult to believe that,

---

[4]The R code for running the algorithm can be found on the authors' websites.

upon a reliable intelligence discovery that a proliferant's program had advanced to a stage that would justify war, options could be considered, a decision made, forces deployed, and an attack executed in less than a few months (in the case of a more limited campaign) to a year (in the case of a full invasion). Given a plausible range of discount rates for a year of 0.9 to 0.99, the range we chose accommodates the possibility that a period could effectively be anywhere from three months to a year in length.

- σ: the per-period probability that A receives a true signal of the stage of B's program, from 0.1 to 0.5 with interval 0.2. Even if a period were a year long, it is difficult to imagine that the US intelligence community would obtain definitive information on the stage of a proliferant's program more than once every two years. And it is also hard to see this happening less than once every ten years.

- ϵ: the per-period probability that B masters the first stage of development, from 0.1 to 0.5 with interval 0.05. If a period is a year, then $\epsilon = 0.1$ means that the proliferant is expected to take 10 years to master the first stage; it is hard to imagine the US paying any attention if proliferation is more than 10 years away. At the other end, it is implausible that a proliferant, once the US *was* watching, could be expected to master the first stage (say, uranium enrichment or plutonium re-processing) in less than six months.

- λ: the per-period probability that B masters the second stage, given that he has mastered the first, and obtains nuclear weapons, from 0.1 to 0.5 with interval 0.05. The same rationale as with the first stage applies here; at the

lower end, the US would not even pay attention, while at the higher, successful weapon design and manufacture would surely not take less than six months in expectation.

- p: the balance of power before B gets nuclear weapons, from 0.7 to 0.999 with interval 0.05. Consistent again with thinking of A as the US, or even any state whose threat to a potential proliferant was sufficient to cause the latter to seek nuclear weapons, we limit the range so that A begins with military superiority over B, whether overwhelming or merely substantial.

- p_n: the balance of power after B gets nuclear weapons, from 0.5 to p with interval 0.05. Given that we assumed A already had nuclear weapons and was conventionally superior, it seems hard to believe that B's acquisition of nuclear weapons would do more than even the odds. And the upper end is restricted to be no more than p, since nuclear weapons won't hurt B's chances in war.

- c_A, c_B: the per-period cost of war for A and B, from 0.01 to 0.05 with interval 0.02. It is hard to see a conflict costing more than 5% of the value of what's at stake in perpetuity—at the highest chosen discount rate, this is equivalent to destroying all of the stakes' value for five years. By contrast, the lower end of the cost range at the lowest discount rate is equivalent to ruining the stakes under contention for a few days.

We focus the discussion on the effects of changes in intelligence-gathering, the speed of weapons development, and the post-proliferation balance of power (i.e., σ, ϵ/λ, and p_n, respectively) on the time A is willing to wait without new stage intelli-

gence (i.e., k) and the probabilities of eventual proliferation, war, and mistaken war. These parameters seem the most easily manipulable or most variable across countries, and thus the most useful for informing policy-making and empirical work. Figure 3 illustrates the results.

First suppose that σ increases. If the change is small, k does not change, but if it is larger, then k also increases. The intuition is that a modest increase in A's ability to gather intelligence about B's stage isn't enough to motivate A to wait another period before attacking, but a larger increase lessens the risk of proliferation enough that A will be more patient. Generally, the probability of war increases and that of proliferation decreases. This is because A simply becomes more likely to catch B in the second stage. But for small values of σ, the opposite can happen (war is less likely, proliferation more likely) because as σ rises A waits longer before attacking, hoping to get a new signal. The probability of mistaken war drops rapidly as σ rises, because the wars that happen are increasingly likely to result from a second-stage signal rather than A's suspicions growing over time.

Next, suppose that ϵ increases. This generally decreases k, although at high levels of ϵ, bigger changes are necessary to alter k. The easier the first stage is to master, the more quickly A will come to believe that B has covertly reached the second stage in the absence of definitive intelligence, and the less time A will wait before attacking. This effect is less pronounced at high levels of ϵ because A is already only willing to wait a few periods. Counterintuitively, the probability of proliferation decreases, and that of war increases, as A more than compensates for the higher sophistication of B's early program by attacking earlier. If k decreases, then the probability of mistaken
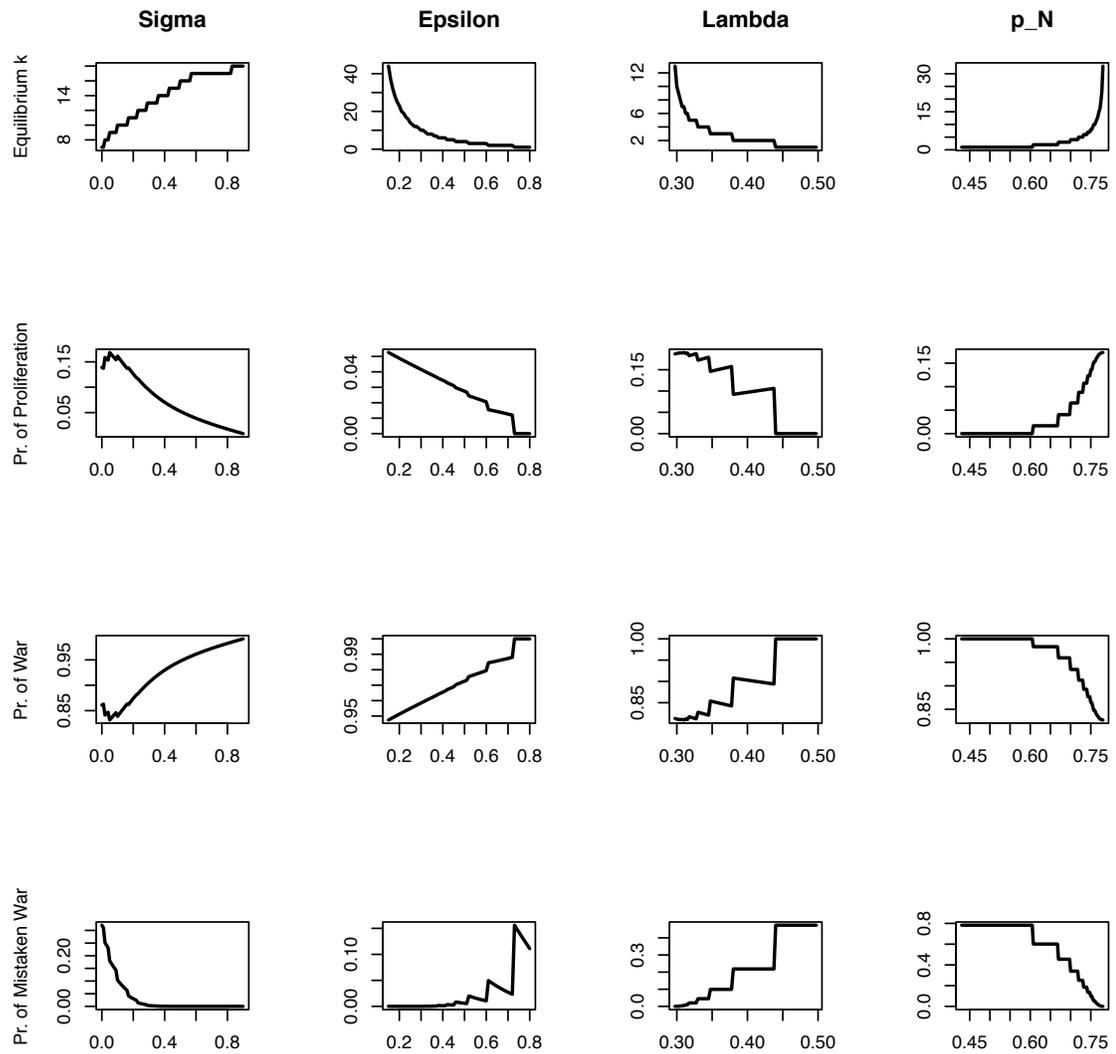
Figure 4.3: These are some representative results from the simulations

war increases substantially, but if k remains the same, this probability goes down. The former is as expected (the less A waits, the more likely he is to be mistaken when he attacks), but the latter is interesting. What happens is that as ϵ rises, but k stays the same, B is more likely to make it to the second stage and thus more likely to be detected there, lessening the proportion of wars that are mistaken.

Now consider an increase in λ. The effect on k is the similar to that of ϵ: k generally decreases, but bigger changes are necessary once λ is high because A is already unwilling to wait more than a few periods. However, the effects on the probabilities of different outcomes are different. If k does not change, then the probability of proliferation increases and that of war decreases: if A is willing to wait just as long, but B's program is speedier, then it is more likely to be successful and thereby avert war. If instead k decreases, then the opposite happens, similar to the effect with ϵ: A more than compensates for the increased speed of B's program by attacking earlier. As expected, if k stays the same, then so does the probability of mistaken war, but the latter rises whenever k decreases.

Finally, consider an increase in p_n. This will generally increase k, though for small levels of p_n, large changes are required. Because larger p_n means the downside of allowing B to get away with proliferation is less severe (i.e., the resulting shift in power is smaller), A is willing to wait longer to try to delay the costs of war. Unsurprisingly, if k stays the same, then the probabilities of proliferation, war, and mistaken war are also constant, but if k increases, then proliferation becomes more likely, and war and mistaken war less likely.

Some surprising implications follow from these results. First, although better in-

telligence capabilities would lessen the likelihood of mistaken wars, they would not actually decrease the probability of proliferation unless intelligence was already reasonably proficient, and if intelligence was poor to start, slight improvements could actually increase the chance of proliferation. Second, greater technological sophistication or outside assistance is not necessarily advantageous for a proliferant, as they may lead its enemy to attack sooner and lower the risk of proliferation, thus leaving the proliferant worse off in expectation. On the other hand, if international efforts to raise the difficulty of the stages of weapons development (e.g., better oversight of nuclear energy firms and tracking of scientists and engineers with relevant expertise) have only modest success, they may have the unintended side effect of increasing the probabilities of both proliferation and mistaken wars. Finally, efforts on the part of the United States to improve its ability to win a war with a nuclear-armed opponent, such as the development of missile defenses and effective preemptive strike capabilities, if successful, may actually increase the risk of proliferation and lessen that of war.

We conclude this section by summarizing the most important results. First, the game is essentially over once B acquires nuclear weapons: A will have to live with this by making bigger concessions, in recognition of the changed balance of power. Second, in the absence of a deal that gives B incentives not to pursue nuclear weapons, he will always invest in a program. Whether A tolerates this or attacks to stop it depends on the anticipated effect of proliferation on the balance of power, the costs of war, and, at the last observable stage of progress of B's program, the speed with which the program is expected to succeed. If, having discovered B in the final stage

of development, A would not attack, then he would never attack and proliferation is inevitable. If A would attack at the last stage, then the two sides play a waiting game until A discovers B to be in this stage, in which B invests and A's suspicions that he has made dangerous progress grow if reassuring intelligence is not forthcoming. This can unfold in many different ways: early toleration can give rise to surprise proliferation; sudden intelligence discoveries can lead to immediate war; mounting apprehension can lead to crises and thence to war or be defused. The occurrence of one path over another is entirely by chance, as B's program may or may not advance, and A's intelligence may or may not detect this. However, deterministic factors like the quality of A's intelligence service or the sophistication of B's program can affect the relative probability of different paths, often in subtle and surprising ways. These effects mean that sometimes policy interventions that seem sound can lead to serious unintended consequences.

Still, in some sense we have told only half the story. When might the two sides be able to avert both proliferation and war with an agreement? We turn to this question next.

4.4 Non-Proliferation Agreements

In this section we will investigate the possibility that the two players could avoid both proliferation and war by agreeing to non-proliferation. In the equilibria analyzed here, B refrains from investment in exchange for incentives from A. The agreements are enforced by the threat of the two sides reverting to the equilibria studied in the previous section. These agreements are not always viable, because it may well be

that even given the most generous incentives A would be willing to offer, B would still prefer to renege on a deal and try to acquire nuclear weapons covertly. Indeed, we will see that the existence of non-proliferation equilibria depends very heavily on what sort of equilibrium would obtain in the absence of an agreement. We first consider deals starting from when A knows that B's program has reached the second stage, and then we will study the possibility of earlier deals that might prevent *any* investment in a weapons program on B's part.

It is worth noting here that the conception of non-proliferation, itself a specific type of arms control, used here is one of cooperation rather than coordination. That is, with one special exception we will discuss, non-proliferation here is not a matter of the players coordinating on an equilibrium (i.e., non-proliferation) that both would prefer to some other equilibrium (i.e., proliferation and war). Instead, arms control here is a matter of the two players cooperating to achieve higher values while one or both face temptations to defect. Our own reading of the empirical record of nuclear non-proliferation suggests that cooperation is the more applicable view.

B Is Known to Be in Second Stage

First we will show that the existence of an enforceable non-proliferation deal depends on whether A can threaten B with war in the absence of a deal. Then we show that, if A can threaten war, the viability of a deal depends on how the severity of this threat compares with the temptation for B to renege on an agreement. We will see that parameters such as the sophistication of B's program, the costs of war to stop it, and the consequences of proliferation can have surprising effects on the viability

of a deal.

Proposition 4.5. *If there is no equilibrium of the overall game, or of the subgame in which B is known to be in the second stage, that features a positive probability of war, then there is also no equilibrium with a deal in either the whole game or the subgame.*

To understand this result, notice that in any non-proliferation deal, two things must be true. First, the value B gets from a deal must be at least a little higher than what he would get without the deal, because otherwise he has no reason not to cheat on the deal—that is, to invest in a weapons program—in the hopes of doing even better once he has nuclear weapons. Second, the value A gets from a deal must be no less than the value he would get without a deal, because otherwise A would rather not accept the deal in the first place. Now, the only way it can simultaneously be true that A gets at least the same value, and B gets a higher value, with a deal than without it, is if going without a deal means doing something that destroys part of the game's total value. In our model, the only thing that destroys value is war; so if no non-deal equilibrium features war, then there is no surplus value with which A can encourage B not to invest while still leaving himself at least as well off.

Remarkably, this does not mean that A could not secure non-proliferation, even while leaving himself better off than he would be if he attacked B to stop the latter's program. In fact, there is a compromise A could offer, that would leave both sides better off than war, and would suffice to ensure B did not develop nuclear weapons. The problem is that this compromise actually leaves A worse off than a lesser offer that would lead to B's eventual proliferation. Thus, sometimes non-proliferation, while feasible, is not worth the cost for A.

Proposition 4.6. *Suppose that B is in the second stage and this is common knowledge. Then there is a non-proliferation equilibrium of this subgame if and only if*

$$p - c_A + \frac{\delta\lambda}{1-\delta}(p - p_n) \geq 1 + \frac{\delta\lambda}{1-\delta}(c_A + c_B) \; and \; [\lambda + \tau(1 - \lambda)](c_A + c_B) \geq \lambda(p - p_n).$$

The first condition in Proposition 4.6 is taken directly from Proposition 4.2, and guarantees that once A knows B has reached the second stage, there is an equilibrium where war occurs. This means that there is potentially something for *both* sides to gain from agreeing on non-proliferation: they will not have to pay the costs of war. Because there is a war equilibrium, A has a credible threat with which to try to enforce a non-proliferation deal—to go to war if B is caught cheating—and so all that matters is whether that threat is severe enough relative to the temptation for B to cheat on the agreement. The second condition formalizes this requirement. If A is to get at least as much value from a deal as from the non-deal equilibrium (i.e., war), then the most he could possibly give B in an attempt to encourage compliance with the deal, and thus the most he could threaten to take away in the event that B is caught cheating, is the surplus value that would be destroyed in the war ($c_A + c_B$, per period). Since A can only punish B if the latter's investment is either detected or successful (in which case A knows B has cheated because B has nuclear weapons), this penalty is weighted by the chance that B's investment either succeeds or is detected ($\lambda + \tau(1 - \lambda)$). For B the temptation to cheat is the shift in the balance of power that proliferation would effect, weighted by the chance that his program would succeed (or $\lambda(p - p_n)$). If the maximum expected penalty outweighs the expected benefit of cheating, then the deal is viable; otherwise, one or the other side will not comply.

As one might expect, the quality of A's monitoring of B's compliance (τ) is critical

to the viability of a deal equilibrium, with better monitoring making a deal more likely to exist. More interestingly, the effects of the other parameters are not monotonic. If the sophistication of B's program (λ) is too low, the costs of war ($c_A + c_B$) are too high, and/or the effect of proliferation ($p - p_n$) is too low, then B's program is not threatening enough to A to cause war, and so A's threat to punish cheating with war is not credible, and thus a deal cannot be enforced. For higher sophistication, lower costs, and/or higher effects of proliferation, A's war threat becomes credible and a deal is easily enforced. If sophistication gets any higher, war costs any lower, or proliferation effects any higher, A's war threat remains credible, but the margin by which the expected penalty exceeds the temptation to cheat will decline, and eventually the temptation to cheat will overwhelm the penalty and the deal will again be unenforceable. Thus program sophistication, war costs, and proliferation effects have competing consequences for the viability of a deal.

This suggests some implications for various policy interventions. Clearly, efforts on the part of the United States to improve the quality of monitoring of non-proliferation agreements unambiguously raise the enforceability of these deals. Efforts to lower the costs of preventive war can strengthen the viability of agreements if they make the US willing to go to war where it previously wasn't, but can undermine agreements if they simply decrease the cost of a war the US was already willing to wage. Efforts to lessen the effects of proliferation, such as the development of missile defenses and preemptive capabilities, and to raise the difficulty of mastering the latter stages of weapons development, have similarly ambiguous consequences.

What about when earlier in the process, when B is not yet know to have mastered

the first stage of weapons development?

B Is not Known to Be in Second Stage

First we will investigate the viability of a non-proliferation deal from the beginning, given that immediate war would result if A caught B in the second stage. In the candidate deal, B would never invest in a program so long as A made sufficiently high offers, and if A caught B cheating, the two sides would revert to the "watch and wait" equilibrium of the previous section, in which B always invests and A waits until he is sufficiently confident that B has reached the second stage to attack. Then we will characterize what would happen if there was the possibility of a later deal once B had been discovered to be in the second stage, and the special issues of coordination and credibility this raises.[5]

To specify the condition under which this deal is viable, we need some additional notation. Let V_{ww}^A and V_{ww}^B be the continuation values of A and B under the watch and wait equilibrium, and S_{ww} be the surplus value that is created by avoiding this equilibrium with its concomitant risk of war (i.e., $S_{ww} \equiv \frac{1}{1-\delta} - \left[V_{ww}^A + V_{ww}^B\right]$). Let $V_1^{B,s1}$ and $V_1^{B,s2}$ be B's continuation values from the subgame of the watch and wait equilibrium in which A has gone one period since the receipt of the last signal that B remained in the first stage, if B is actually in the first and second stages, respectively.

---

[5]Throughout this section, we ignore the possibility of deals based on separating equilibria, in which non-proliferation becomes viable after an initial period of investment. In these equilibria, once there is any uncertainty about B's stage, A offers a deal supported by the threat of immediate war if B is discovered cheating. The deal is designed so that B would comply with it if he remained in the first stage, and cheat on it if he had advanced to the second. This makes A's threat of immediate war credible, even though it would not be if A knew for sure that B remained in the first stage, because if A detects cheating he can infer with certainty that B is in the second stage. While this is perfectly valid within the confines of our model, it strikes us as implausible.

Recall from Proposition 4.1 that $V_n^B \equiv \frac{1-p_n-c_B}{1-\delta}$ is B's continuation value once he has obtained nuclear weapons. Finally, let $V_{np}^{B,s2}$ be the continuation value B would obtain in the subgame in which he had cheated once on the candidate deal and made it to the second stage, all without being detected by A, and recall that B's continuation value from war prior to proliferation is $W^B \equiv \frac{1-p-c_B}{1-\delta}$.

Proposition 4.7. *There is a non-proliferation equilibrium of the game, based on the threat of reverting to the "watch and wait" equilibrium of Proposition 4.4 if A catches B investing, if and only if immediate war is an equilibrium outcome of the subgame in which B is known to be in the second stage and:*

$$V_{ww}^B + S_{ww} \geq \epsilon\lambda V_n^B + \epsilon(1-\lambda)\left[\sigma W^B + (1-\sigma)\tau V_1^{B,s2} + (1-\sigma)(1-\tau)V_{np}^{B,s2}\right]$$
$$+ (1-\epsilon)\left[\sigma\tau V_{ww}^B + (1-\sigma)\tau V_1^{B,s1} + (1-\tau)(V_{ww}^B + S_{ww})\right]$$

The left-hand side of the condition in Proposition 4.7 is the highest value B could possibly receive in any non-proliferation deal that A would be willing to offer. This value is just whatever B would get in the watch and wait equilibrium, plus the entire surplus that is created by avoiding that equilibrium and the possible costs of war it entails. A wouldn't offer any more than this, because this the value of the deal for A would be less than that of the watch and wait equilibrium.

The right-hand side is the value B would prospectively get by cheating on the putative deal. If B's investment leads immediately to nuclear weapons (with probability $\epsilon\lambda$), then B gets the greater concessions from A that proliferation brings (V_n^B). If B's investment leads to mastery of the first stage but not the second (with probability $\epsilon(1-\lambda)$), what happens depends on what A observes. If A detects that B has reached the second stage (with probability σ), then there will be war (yielding

W^B), but if he detects only that B has cheated (with probability $(1 - \sigma)\tau$), then the watch and wait equilibrium will begin with A having gone a period of investment without receiving any signal of stage (giving B a value of $V_1^{B,s2}$). If A detects nothing $((1 - \sigma)(1 - \tau))$, then B has the opportunity to continue cheating on the deal without A knowing ($V_{np}^{B,s2}$). And finally, if B's investment goes nowhere $(1 - \epsilon)$, and A sees both stage and investment $(\sigma\tau)$, then the watch and wait equilibrium begins (giving V_{ww}^B). If instead A detects B's cheating but not his stage $((1 - \sigma)\tau)$, then the watch and wait equilibrium begins with A having gone one period without a stage signal (yielding $V_1^{B,s1}$), and if A does not see B's cheating $((1 - \tau))$, then B can go back to compliance with the deal with A none the wiser.

If the left-hand side exceeds the right, then there is a deal that A is willing to offer and that is generous enough that B would rather comply with it than cheat on it. As with the second-stage-known deal analyzed in the previous subsection, the viability of a deal starting from the beginning of the game turns critically on the quality of A's monitoring of B's investment and progress (σ and τ). Better monitoring makes the relatively good possible outcomes of B cheating (such as making it to the second stage and then continuing to cheat under A's nose) less likely and the relatively bad outcomes (such as war) more likely. It also raises the surplus—as we saw in the previous section, higher σ means a higher risk of war in the watch and wait equilibrium. And it also lowers the value of many of the cheating possibilities, and leaves the rest the same. Thus, better intelligence means that the surplus with which A can encourage compliance is higher, the chance of getting caught cheating is higher, and the value of cheating even if it is not immediately detected is lower.

It is also important to note the role that war in the watch and wait equilibrium plays in the viability of a non-proliferation deal. When war is less likely (or happens later) in the watch and wait equilibrium, the surplus to be gained from avoiding this equilibrium in favor of a deal is smaller, and the punishment of switching to the watch and wait equilibrium is equally or less severe for B. Thus, the less dangerous the watch and wait equilibrium, the harder it will be to find a viable non-proliferation agreement. For some values of the exogenous parameters, there may be a viable non-proliferation deal once B has reached the second stage and this is known, but no such deal earlier in the game when A is pretty sure B remains in the first stage and thus cannot credibly threaten war.

More subtly, the existence of a viable deal once B is known to have reached the second stage can itself undermine the viability of an earlier deal and increase the probability of proliferation. To see why, recall from Proposition 4.6 that whenever there is a second-stage deal, there is also an immediate war equilibrium of that sub-game (it is the threat of reverting to this that supports the deal). Because the war equilibrium is worse for B than the deal, it is easier to support an earlier deal if it is backed by the threat of reverting to the war equilibrium, rather than the threat of reverting to the later deal. The possibility of detected progress leading to a more generous non-proliferation deal might even motivate B to invest in a covert program in the first place, in addition to the prospect of actually obtaining nuclear weapons. This suggests the interesting possibility that some proliferants might pursue nuclear weapons *mainly* out of the hope that it will eventually bring forth a more generous "bribe" from outside powers, in exchange for stopping an advanced program.

So suppose that a deal from the beginning exists under the war threat, but not under the threat of the later deal, and imagine the following situation. The two sides agree to non-proliferation at the beginning of the game, enforced by the threat of the watch and wait equilibrium and eventual war if A estimates that B is likely enough to have reached the second stage. At some point during the game, A receives a signal that B has invested, and the two sides revert to the watch and wait equilibrium. Then at some point A becomes confident enough that B has reached the second stage that the equilibrium strategy calls for him to attack. However, the two players could instead coordinate on the equilibrium of the second-stage deal. B would certainly go along, as this equilibrium is strictly better for him than war, and it is also better for A. Why would A go to war, given that there was an alternative that was better for him and certain to be agreed to?

The problem occurs if B anticipates this scenario and A's failure to carry out his war threat when the equilibrium calls for it. If he does, then it is rational for him to cheat on the initial deal. If in turn A anticipates that B will reason in this way, then the deal would not be viable after all. Thus, the existence of a later deal can fatally undermine an earlier one, even when the earlier one is in equilibrium, unless A can somehow coordinate the two sides' expectations so that both anticipate that A *would* carry out the war threat, even though it is worse for both.

We can see how this might be done by considering whether a similar problem undermines the later deal. In principle, A faces the same choice. Having caught B cheating, two equilibria can be played: immediate war, or simply starting the deal over. In practice, A knows that if he does not carry through with war in the face of

B's cheating, then B will have good reason to cheat again and thus re-starting the deal would only lead to more cheating. Thus, at the point at which the equilibrium strategy calls for it, A will attack, because he expects that not doing so will guarantee that a re-started deal is not viable. Thus, there is no reason for B to doubt the credibility of the threat that underpins the later deal. The key here is that there is reason for A to believe that whether he carries out the required threat now will affect the viability of any future deals—this makes it rational to carry out the threat, and anticipating this, B will comply with the deal.

Within the bounds of our model, this won't work for enforcing an earlier deal. By the above reasoning, B knows that A would enforce a later non-proliferation deal, regardless of whether A enforced an earlier one. Because A's decision on whether to enforce an earlier deal with war does not affect B's expectations that a later deal would be enforced, and thus the viability of a later deal, A has little reason to carry through with a costly punishment now rather than agreeing to a later deal.

In reality, A (e.g., the United States) faces more than one potential proliferant across the world's states. And even if the US goes to war with a proliferant to stop its program, the interaction will not actually end, as our model assumes. The US might lose the war, in which case it will have to continue interacting with the opposing state, which might decide to continue or renew its program. And even if the US wins the war, the defeated state will eventually regain its sovereignty and have to decide anew whether to pursue nuclear weapons. In choosing whether to impose costly punishment to enforce a particular early deal, the US must consider the effect its choice will have on the viability of deals with both other states and later incarnations of the cheating

state. If the US fails to enforce an early deal with one state now, others present and future might conclude that it will not enforce their own deals, with fatal consequences for their viability.

These negative externalities can make it rational for the US to carry out a costly punishment equilibrium, even if a later deal equilibrium that is not itself costly also exists, and thus preserve the viability of the earlier deal. A full analysis of the conditions under which these externalities are large enough to have this effect is beyond the scope of the current paper, but it nonetheless suggests a precise mechanism by which concerns over reputation may influence non-proliferation.

To summarize the most important results of this section, we first found that if war never occurs in any equilibrium, then non-proliferation is never viable, because there is no surplus with which to reward B's compliance. This can happen even when there is a non-proliferation deal with which B would comply and which A would still prefer to war—it's just that A would actually prefer the risk of proliferation and war to offering this deal. As expected, we found that the viability of deals was strongly dependent upon the quality of A's monitoring of B's compliance. However, other factors such as the sophistication of B's program, the costs of war to stop it, and the effects of proliferation on the balance of power can have counter-intuitive effects: changes in these factors can either help or harm the goal of non-proliferation. We further showed that deals starting with B known to be in the first stage are easier to support when the "watch and wait" equilibrium that would otherwise obtain, and to which the players would revert in the event that A caught B cheating, is more war-prone. In particular, it is especially difficult to support non-proliferation from

the get-go if a viable second-stage deal exists, because this gives B another reason to start a program (i.e., to seek a more generous non-proliferation deal) and raises questions about whether A would actually carry out the threat of war given that a less costly (for A) and less severe (for B) later deal exists. Ultimately, the viability of an early-stage non-proliferation deal supported by the threat of eventual war, when a later-stage deal is also viable, may depend on A's need to preserve the viability of other such deals with other potential proliferants.

4.5   Conclusion and Implications

We conclude with a discussion of the implications our results suggest for the empirical study of nuclear proliferation, the theoretical analysis of arming, and policy-making on nuclear issues.

Perhaps our most important result for empirical scholars is the finding that much of the over-time variation, and at least some of the across-country variation, in proliferation interactions is driven by stochastic elements such as when a proliferant's program will make progress and when this will be observed by its opponent. Together with the small number of cases of nuclear weapons programs in the empirical record, this suggests that there are fundamental limits to our ability to make inferences about the role of exogenous factors, such as the effects of proliferation, the costs of war, and the sophistication of a proliferant's program and its outside assistance. These factors affect only the *expected* outcome in a given case; the actual outcome is still highly variable because of the stochastic elements.

Many statistical studies to date have gotten around the small number of weapons

programs by designating the country-year or dyad-year as the unit of analysis, a technique which greatly increases the number of observations and thereby the apparent strength of any patterns in the data. These studies typically assume either that observations are independent over time, or that any dependence over time decays rapidly. Our model implies that this approach is deeply flawed: behavior in the model depends heavily on what has happened in earlier time periods, and for some values of the exogenous parameters, this dependence may be long-lasting. This suggests that statistical analysts of the record must resort either to more detailed, theoretically-informed modeling of the temporal dependence in these interactions, or to the more conservative approach of treating a country or dyad, rather than a country-year or dyad-year, as the modeled unit of analysis.

This does not mean that our model cannot be tested. Though it suggests the exogenous factors that scholars have focused on to date only weakly affect the observed outcomes, the stochastic elements have much stronger effects and are, at least in principle, measurable. An effort to gather data on exactly when historical proliferants' programs mastered various stages of nuclear weapons development, and exactly when this progress was detected by their opponents, would be valuable for testing our theory and for exploring the role of progressive development and intelligence estimates in these interactions more generally. Of course, the difficulties of securing this data from governments with every incentive to conceal it should not be underestimated.

For theorists of arming, our analysis suggests that the neglect of arms control (non-proliferation agreements in our case) in studies to date renders their results suspect. The typical working assumption is that when players have two equilibria

to choose from, if one is Pareto-superior to the other then they play the "better" one. If that is so, then when arms control is viable the players would select it over unconstrained arming, since the latter typically carries the risk of (costly) war. But this selection can invalidate the comparative statics derived from the unconstrained arming equilibrium. So, for example, in our analysis, lower costs of war make war more likely *in the non-deal equilibrium.* But when the possibility of arms control is taken into account, lowering the costs of war may render a deal viable and thus *decrease* the likelihood of war. Determining which of the conclusions surfaced so far in this literature are robust to the consideration of arms control equilibria is an important subject for future research.

The absence of arms control in existing models goes hand-in-hand with the typical restriction to either a finite-period game structure or Markov Perfect Equilibrium in infinite-period games. Though it is possible to consider forms of arms control based on coordination on less-armed equilibria in these settings, they automatically exclude arms control based on cooperation and the threat of future punishment. Each restriction is chosen as a mathematical convenience, and while we do not dismiss the difficulties that are introduced in attempting to model something as complicated as arming in a less restrictive setting, it bears acknowledging that there is simply no substantive justification for either type of restriction. Because it is typically costly, arming is mysterious for exactly the reason war is: why do rational actors employ costly measures (whether war or arming) to manage their disputes? We cannot know the answer if we do not understand why arms control sometimes fails to be viable.

Our results also suggest that there are unappreciated links between the two con-

ceptions of arms control, as coordination onto better equilibria or as cooperation supported by the threat of reversion to a worse equilibrium. In particular, in our analysis of non-proliferation agreements, we found that cooperation on a Pareto-superior deal required coordination on a Pareto-inferior threat equilibrium that might well fail to be credible, even though it was a perfect equilibrium. This turns the usual way of posing the arms coordination problem on its head, and may well apply in arming contexts other than nuclear proliferation.

In our model, we assumed that, with nuclear weapons, the effects on the balance of power are more about whether a state has them than about how many it has. This seems in keeping with much of the literature, but it may be true of other military technologies as well. The focus so far on theoretical models of arming *quantity*, rather than quality, is undoubtedly useful. But in the modern era with which scholars of international conflict are most concerned, the research and development of better weapons may be as important as the conscription of more soldiers and the manufacture of more guns, and raises different theoretical issues. The most salient of these is the fact that development takes time, so that the effect of acquiring the weapons is delayed, and subject to prevention, whether by war or by deal.

For policy-makers, the most important advice to follow from our work is that many of the policy initiatives typically debated in the arena of nuclear issues may have counter-intuitive and unintended consequences. Missile defense, preemptive strike capabilities, and diplomatic efforts to cut off outside assistance to a proliferant's program might well improve the expected outcome for the US, but not always in the way intuition expects—these can sometimes improve the prospects for non-proliferation,

but they also sometimes undermine them.

One initiative that is unambiguously good is to improve intelligence-gathering. Higher-quality intelligence on the existence and progress of a state's nuclear weapons program unambiguously increases the likelihood that a non-proliferation deal becomes and remains viable, but it also unambiguously increases the ability of the US to successfully prosecute interactions with proliferants in the absence of a deal. As far as our model can discern, it is an unalloyed good for the US.

Finally, our model suggests a tension between two goals that arises when the US must respond to a potential proliferant's cheating on a non-proliferation deal. On the one hand, responsibility and the need to preserve the responsible reputation of the US internationally argue for giving a cheater every chance to find a new viable non-proliferation deal. On the other hand, allowing a proliferant to extract additional concessions by virtue of progress in its illegal weapons program can undermine deals with other potential proliferants. Going to war in response to the latter concern might or might not be the right thing for the US to do, but even if it is, the US government is still placed in the unenviable position of explaining to its domestic and international audiences why it undertook a violent, costly resolution of the crisis, when one that would avoid both violence and its costs was available.

Bibliography

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

Bas, Muhammet A. and Andrew J. Coe. 2012*a*. "Arms Diffusion and War." *Journal of Conflict Resolution* . Forthcoming.

Bas, Muhammet A. and Andrew J. Coe. 2012*b*. "A Model of Proliferation and Prevention." Working Paper.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry R. Weingast. 1998. *Analytic Narratives*. Princeton University Press.

Bilmes, Linda and Joseph Stiglitz. 2006. "The Economic Costs of the Iraq War: An Appraisal Three Years after the Beginning of the Conflict." National Bureau of Economic Research Working Paper 12054.

Bush, George and Brent Scowcroft. 1999. *A World Transformed*. Vintage.

Coe, Andrew J. 2012. "The Limits of Arms Control and the Origins of the Iraq War." Working Paper.

Coe, Andrew J. and Jane Vaynman. 2012. "Superpower Collusion and the Nuclear Non-Proliferation Treaty." Working Paper.

Cordesman, Anthony H. 2002. *Iraq's Military Capabilities in 2002: A Dynamic Net Assessment*. Center for Strategic and International Studies.

Craft, Cassady and Joseph Smaldone. 2003. "Arms Imports in Sub-Saharan Africa: Predicting Conflict Involvement." *Defence and Peace Economics* 14(1):37–49.

Davis, Steven J., Kevin M. Murphy and Robert H. Topel. 2006. "War in Iraq versus Containment." National Bureau of Economic Research Working Paper 12092.

de Figueiredo Jr., Rui J. P., Jack Rackove and Barry R. Weingast. 2006. "Rationality, Inaccurate Mental Models, and Self-Confirming Equilibrium." *Journal of Theoretical Politics* 18(4):384–415.

Debs, Alexandre and Nuno Monteiro. 2010. "Nothing to Fear but Fear Itself? Nuclear Proliferation and Preventive War." Working Paper.

Director of Central Intelligence. 2002. "Iraq's Continuing Programs for Weapons of Mass Destruction." National Intelligence Estimate, October.

Duelfer, Charles A. 2002*a*. Testimony before the Foreign Relations Committee of the United States Senate, July 31.

Duelfer, Charles A. 2002*b*. "Weapons of Mass Destruction Programs in Iraq." Testimony before the Subcommittee on Emerging Threats and Capabilities, Armed Services Committee of the United States Senate, February 27.

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

Fearon, James D. 2011*a*. "Arming and Arms Races." Working Paper.

Fearon, James D. 2011*b*. "A Simple Political Economy of Relations among Democracies and Autocracies." Working Paper.

Flibbert, Andrew. 2006. "The Road to Baghdad: Ideas and Intellectuals in Explanations of the Iraq War." *Security Studies* 15(2):310–352.

Gavin, Francis J. 2004. "Blasts from the Past: Proliferation Lessons from the 1960's." *International Security* 29(3):100–135.

Gavin, Francis J. 2009/10. "Same As It Ever Was: Nuclear Alarmism, Proliferation, and the End of the Cold War." *International Security* 34(3):7–37.

Goldberg, Jeffrey. 2010. "The Point of No Return." *The Atlantic* .

Hirshleifer, Jack. 1991. "The Paradox of Power." *Economics & Politics* 3(3):177–200.

Iraq Intelligence Commission. 2005. "Report of the Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction." Unclassified Version, March 31.

Iraq Survey Group. 2004. "Comprehensive Report of the Special Advisor to the DCI on Iraq's WMD." September 30.

Jackson, Matthew O. and Massimo Morelli. 2009*a*. "The Reasons for Wars: An Updated Survey." Forthcoming in The Handbook on the Political Economy of War, edited by Chris Coyne, Elgar Publishing.

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

Kahl, Colin H. 2012. "Not Time to Attack Iran." *Foreign Affairs* 91:166–173.

Krause, Volker. 2004. "Hazardous Weapons? Effects of Arms Transfers and Defense Pacts on Militarized Disputes, 1950–1995." *International Interactions* 30(4):349–371.

Kroenig, Matthew. 2012. "Time to Attack Iran." *Foreign Affairs* 91:76–86.

Krueger, Alan B. 2006. "The Cost of Invading Iraq: Imponderables Meet Uncertainties." *The New York Times* . March 30.

Krueger, Alan B. 2007. "Economics, Iraq, and the War on Terrorism." Summary of remarks at the National Economists Club, September 25, 2007.

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

Mazarr, Michael J. 1995. *North Korea and the Bomb: A Case Study in Nonproliferation.* Palgrave Macmillan.

McBride, Michael, Gary Milante and Stergios Skaperdas. 2011. "Peace and War with Endogenous State Capacity." *Journal of Conflict Resolution* 55(3):446.

McBride, Michael and Stergios Skaperdas. 2007. Explaining Conflict in Low-Income Countries: Incomplete Contracting in the Shadow of the Future. In *Institutions and Norms in Economic Development*, ed. Mark Gradstein and Kai A. Konrad. MIT Press pp. 141–162.

McBride, Michael and Stergios Skaperdas. 2009. "Conflict, Settlement, and the Shadow of the Future." Working Paper.

Mearsheimer, John J. and Stephen M. Walt. 2003. "An Unnecessary War." *Foreign Policy* (134):50–59.

Meirowitz, Adam and Anne E. Sartori. 2008. "Strategic Uncertainty As a Cause of War." *Quarterly Journal of Political Science* 3(4):327–352.

Nordhaus, William D. 2002. The Economic Consequences of a War with Iraq. In *War with Iraq: Costs, Consequences, and Alternatives.* American Academy of Arts and Sciences.

Pollack, Kenneth M. 2002. *The Threatening Storm: The Case for Invading Iraq.* Random House Inc.

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 University Press.

Powell, Robert. 2002. "Bargaining Theory and International Conflict." *Annual Review of Political Science* 5(1):1–30.

Powell, Robert. 2006. "War As a Commitment Problem." *International Organization* 60(1):169–203.

Powell, Robert. 2011. "Bargaining, Fighting, and State Consolidations." Working Paper.

Quester, George H. 2000. *Nuclear Monopoly.* Transaction Publishers.

Rackove, Jack N., Andrew R. Rutten and Barry R. Weingast. 2000. "Ideas, Interests, and Credible Commitments in the American Revolution." Working Paper.

Reiter, Dan. 2003. "Exploring the Bargaining Model of War." *Perspective on Politics* 1(1):27–43.

Silverstone, Scott A. 2007. *Preventive War and American Democracy.* Routledge.

Slantchev, Branislav L. 2010. "Borrowed Power: Debt Finance and the Resort to Arms." Working Paper.

Trachtenberg, Marc. 1988. "A 'Wasting Asset': American Strategy and the Shifting Nuclear Balance, 1949-1954." *International Security* 13(3):5–49.

Wallsten, Scott and Katrina Kosec. 2005. "The Economic Costs of the War in Iraq." AEI-Brookings Joint Center for Regulatory Studies Working Paper 05-19.

Waltz, Kenneth N. 2012. "Why Iran Should Get the Bomb." *Foreign Affairs* 91:2–5.