Title
Resolve, Reputation, and War: Cultures of Honor and Leaders' Time-in-Office

Permalink
https://escholarship.org/uc/item/9v72937b

Author
Dafoe, Allan

Publication Date
2012

Peer reviewed|Thesis/dissertation

Resolve, Reputation, and War:
Cultures of Honor and Leaders' Time-in-Office

by

Allan Dafoe

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Political Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Steven Weber, Chair
Professor Ron Hassner
Professor Jasjeet S. Sekhon
Professor Matthew Rabin

Spring 2012

Resolve, Reputation, and War:
Cultures of Honor and Leaders' Time-in-Office

Abstract

Resolve, Reputation, and War:
Cultures of Honor and Leaders' Time-in-Office

by

Allan Dafoe

Doctor of Philosophy in Political Science

University of California, Berkeley

Professor Steven Weber, Chair

Leaders throughout history have been concerned—often obsessed—with how other leaders perceive them. Historians have argued that many wars have been fought for purposes of reputation, honor, status, and prestige. However, there is little systematic study of these phenomena, and specifically of the effects of *concern for reputation for resolve* on interstate conflict behavior. After precisely defining resolve, and reputation for resolve, this dissertation examines this question by developing a family of formal models of escalation and reputation-engagement. From these models I explicitly deduce four testable implications of variation in concern for reputation for resolve that take selection effects into account and are robust to a variety of assumptions.

To test these implications I search for research designs where concern for reputation is manipulated in a manner that is reasonably well-understood, and in which large unknown biases are unlikely. I find two such designs. The first, analyzed in collaboration with Devin Caughey, compares U.S. conflict behavior depending on whether the president is from the U.S. South. The U.S. South has a "culture of honor" that places greater importance on an individual's reputation for resolve. Using permutation inference and NPC, a technique new to political science, to provide a joint statistical test of the predictions, we find that conflict behavior under Southern presidents is substantially and significantly different from non-Southern presidents in a manner predicted by my theory. Furthermore, this difference remains significant under a large number of matched comparisons. This result is unlikely to be spurious because of its robustness to conditioning and because of theoretical and empirical inconsistencies in the alternative accounts.

The second research design compares the conflict behavior of leaders early in their time-in-office with those same leaders later in their time-in-office. Leaders at the start of their tenure, as compared with themselves later in their career, should care more about their reputations because they have less developed reputations and longer time horizons. Each prediction of my theory is examined and finds statistically significant support with relatively

large effect sizes. Alternative potential explanations are systematically ruled out through critical tests. In summary, the results from a set of clear tests of the effect of concern for reputation are consistent with the strongest claims made in the literature: concern for reputation seems to be an extremely important cause of war.

To my parents.

Contents

List of Figures

List of Tables

Acknowledgments

I have many people that I wish to thank for advice, encouragement, and help. I want to first thank my committee chair: Steven Weber. In addition to many stimulating conversations and essential institutional support, Steve provided the initial encouragement that led me to study the topic of reputation. Steve and I organized a workshop on reputation that led to many productive conversations and contacts, and it was during one of these conversations with Steve when the idea came to me to search for within-country variation in the cultural origins of leaders.

I thank Jas Sekhon for teaching me the art of good research design and the many, many ways that observational causal inference can fail. I am grateful to Ron Hassner for many conversations that clarified my thinking, and for help naming my first paper. I thank Matthew Rabin for many helpful lessons on the aesthetic of formal theory, and for inspiration about scholarly inquiry. I thank Bob Powell for generously spending many hours advising me on different research ideas. I wish to acknowledge my fortune of having been advised by the late David Freedman; David taught me about the many pitfalls as well as the great value of statistical causal inference in social science.

I was fortunate to have a number of scholars from other universities advise me. I am grateful to Mike Tomz for many meetings and helpful advice that shaped the direction of my research. I thank Bruce Russett for the encouragement and many suggestions on my work, and with whom I am pleased to have the pleasure of collaboration. I thank Paul Huth for many productive conversations, some of which I am glad evolved into collaborative research. I thank Devin Caughey for many stimulating conversations and for producing with me the work of scholarship that I am most proud of. I am grateful to the many other people not mentioned by name here who have engaged me in conversations about or otherwise helped my work. And I extend a special thank you to Charlotte Jandér for our many conversations and her immeasurable support.

For fellowship support, I acknowledge the Social Sciences and Humanities Research Council of Canada, the Department of Political Science Formal Theory Fellowship, and the Institute on Global Conflict and Cooperation. For hospitality as a visiting scholar, I thank the Department of Peace and Conflict Research and especially the Chair Magnus Öberg.

Chapter 1

Reputation for Resolve in International Relations

> "Let none of you think that you are going to war over a trifle ... [for it] contains the affirmation and the **test of your resolution**. If you yield to [the Spartans] you will immediately be required to make another concession which will be greater, since you will have made the first concession out of fear."
>
> ——————————————————————————————————
>
> Pericles' argument to the Athenian Assembly, beginning the Peloponnesian War, 431 BCE

Foreign policy elites seem to be obsessed with their state's reputation. Examples abound. Starting with the onset of the first major war for which we have a good history: Pericles famously persuaded his Athenian peers to resist the Spartan ultimatum that involved only a minor demand—a trifle—because of the principle that it represented. This began the Peloponnesian War. Concern for reputation is such a common motive leading to war, I have yet to find an example of a war in which concern for reputation did not appear to play a major role. Other examples include the Punic Wars, War of 1812, the Mexican-American War, World War I, World War II, the Vietnam War, and the Invasion of Iraq.[1] The prominence, even necessity, of reputational considerations in the outbreak of major wars makes sense once we consider the iterated nature of international interactions and the endogeneity of stakes in crises (see section 1.4). Diplomatic historians and classical scholars of international relations generally appreciate the vast importance of concern for reputation. However, modern scholarship has neglected the importance of concern for reputation; there are very few studies of this important phenomenon using systematic data and formal theoretical and empirical analysis. This dissertation will advance our understanding about concern for reputation through a number of innovations:

——————————————————————————————

[1]Ned Lebow (2010) codes 73 out of the 94 wars he analyzes as having as the prime motive either *standing* or *revenge*, as opposed to *security*, *interest*, or *other*.

1. I will articulate theory using mathematics to formally derive a set of testable implications of variation in concern for reputation (see chapter 2). These predictions will explicitly account for selection effects, which otherwise bedevil studies using observational data and especially studies of reputation.

2. I then test these predictions using two complementary empirical strategies, each of which uses the tools and sensibilities of design-based inference to minimize the risk of unknown biases.

3. The first empirical strategy involves comparing the conflict behavior of the U.S. under Southern presidents, who come from a "culture of honor" and should have greater concern for reputation, with non-Southern presidents.

4. The second empirical strategy involves comparing leaders early in their time-in-office, who have incentives to care most about their reputations, with themselves later in their tenure.

In summary, this dissertation reports the results from a set of clear tests of the effect of concern for reputation. The results are consistent with the strongest claims made in the literature: concern for reputation seems to be an extremely important cause of war.

1.1 Chapter Overview

Historical scholarship suggests that, in most historical periods, policy elites have been obsessed with social attributes and behaviors such as credibility, status, honor, prestige, respect, deterrence, deference, humiliation, and revenge (1.2). These all relate to, and many depend on, a specific phenomenon: *concern for reputation for resolve* (1.3). I define *resolve* precisely, and show it to be a trait worth possessing in strategic settings such as the international system (1.4). Studying concern for reputation for resolve, as the study of reputation more generally, has been handicapped by a number of difficulties. This dissertation overcomes these difficulties by quantitatively studying a set of formally derived behavioral implications of concern for reputation on two distinct empirical domains, using tools and sensibilities that are sensitive to the pitfalls of causal inference on observational data. This dissertation makes substantive and methodological contributions: Substantively, it provides powerful quantitative evidence of the importance of social motives in international relations, and in particular of considerations of reputation for resolve. Methodologically, it demonstrates the benefits of deriving hypotheses from a formal model and of employing research strategies more sensitive to design considerations.

1.2 Credibility, Deference, Status and other Reputational Aspirations of States

Thucydides found that people go to war out of "honor, fear, and interest."… If we take honor to mean fame, glory, renown, or splendor, it may appear applicable only to an earlier time. If, however, we understand its significance as deference, esteem, just due, regard, respect, or prestige we will find it an important motive of nations in the modern world as well.. [In many wars] consideration of practical utility and material gain, and even ambition for power itself, play a [small role] in bringing on wars [whereas often] some aspect of honor is decisive.

Donald Kagan, *On the Origins of War and the Preservation of Peace*, p. 8

There is an esteemed lineage of scholarship emphasizing how the concern of leaders about how their reputations, and the reputations of their state, will influence their credibility and status in the international system. Thucydides argued that people waged war because of "honor, fear, and interest." (Kagan, 1995) Hobbes (1996, ch. XIII) considered honor to be one of the "three principall causes of quarrell". Morgenthau writes that a state's "prestige— its reputation for power—is always an important and sometimes a decisive factor" in foreign policy (Morgenthau, 1948, p. 95). Schelling (1966, p. 124) famously wrote that face, or reputation for action, is "one of the few things worth fighting over." Jervis summarizes that "states seem preoccupied with the credibility of their threats" (Jervis and Snyder, 1991, p. 26). Snyder and Diesing (1977, p. 199) observed "threats that engage prestige, honor, and future bargaining reputation...in all [of the 16] cases [that they] studied". Snyder and Borghard (2011, p. 437) observe that "domestic audiences care about their country's reputation for resolve and national honor".

Before proceeding, I will define some key terms and discuss how they relate to each other. *Reputation* is an inference about an actor, based on that actor's past behavior, that informs predictions about their future behavior. *Credibility* is the extent to which an actor's statements or implicit commitments are believed; in our context credibility refers to the extent to which others believe an actor will carry out an explicit or implicit threat.[2,3] *Resolve* is defined here as the probability that an actor is willing to escalate a particular dispute to war; *reputation for resolve* is thus the inference of others about an actor's resolve in a given dispute based on past behavior; these terms will be discussed in greater detail in section 1.3. *Deterrence* refers to the use of threats to dissuade others from doing something. *Honor* is a particularly complicated concept; for our purposes *national honor* refers to the status and reputation for resolve of a country and an individual's *honor* refers to that individual's status

---

[2]McMahon (1991, p. 457) writes: "To be credible means that others will believe your threats and promises and act accordingly."

[3]Credibility can be a property of an actor, or a statement. An actor has credibility if his threats are believed. A threat is credible if the threat is believed, which is a function of the issuer's credibility as well as other aspects of the threat.

and reputation for resolve, though other meanings related to honesty, integrity, and proper behavior are also closely connected. *Prestige* (following O'Neill, 2006) refers to whether there is higher order knowledge that an actor or institution possesses a sought after quality. *Status* is a broad term, but is used here in a manner similar to *standing* to refer to an actor's position in a social hierarchy, and consequently what rights, respect, and patterns of *deference* from others they should expect.

These concepts are all related. States aspire to have credibility as that will help them get their way; a credible threat is a highly effective tool of statecraft since it induces desired behavior in others at no (immediate) cost to the issuer. States have more credible threats if they have a reputation for fulfilling their threats and a reputation for resolve. Deterrence is enhanced by credibility and a reputation for resolve. Concern with honor, national or individual, enhances credibility and deterrence because it implies a greater concern with a reputation for resolve and the status associated with having one's threats and rights respected. Standing and status are upheld in most eras through the willingness to use force to defend one's rights; that is, through being perceived as resolved towards the defence of one's standing.

This section 1.2 will review the extensive literature investigating the role of reputational aspirations in international relations. Specifically, section 1.2.1 will demonstrate that many scholars believe reputational aspirations are important to states and policy elites. The following sections will survey the scholarly literature studying credibility, deterrence, and the pursuit of reputation for power, alliance behavior, the role of honor and revenge, prestige, and the pursuit of status. Section 1.3 will then focus on a common denominator of most of these reputational aspirations: reputation for resolve. Section 1.4 will look at why reputation for resolve is so important in the international system. Section 1.5 will discuss the study of concern for reputation for resolve.

1.2.1   Reputational Aspirations in International Politics

Leaders, policy elites, and national groups have often pursued reputational objectives; they worry about their state's credibility, about the defence of national honor, and about the country's status. Tang (2005) refers to the obsession of policy elites as the "cult of reputation". Kagan (1995) finds considerations of honor prominent in the outbreak of many wars. Lebow argues that many wars are fought from considerations of honor (2008), and in pursuit of standing or revenge (2010). Barry O'Neill (1999) emphasizes the importance of symbolism, honor, and prestige. Even critics of whether reputational inferences operate in international relations are in agreement that many leaders are very concerned about their own reputations (Hopf, 1994; Jervis, 2002; McMahon, 1991; Mercer, 1996; Press, 2005). The list of scholars whose work affirms the salience of reputational considerations to state leaders is long.

1.2.2 Threat Credibility, Deterrence, and Reputation for Power

Some scholars have looked at the correlates of effective deterrence and threat credibility (Clare and Danilovic, 2010; Crescenzi, Enterline, and Long, 2008; Crescenzi, 2007; Crescenzi, Kathman, and Long, 2007; Huth, 1988; Huth, 1997; Huth and Russett, 1993; Lebow and Stein, 1990; Russett, 1963; Sartori, 2005; Sechser, 2007; Sechser, 2010; Walter, 2006; Walter, 2009).

Deterrence and credibility has been explored theoretically, both informally and formally, by political scientists (Ellsberg, 1968; Fearon, 1994b; Morrow, 1989; Nalebuff, 1991; Sartori, 2002; Schelling, 1960; Schelling, 1966; Wagner, 1992), and more generally by economists (Abreu, Dutfa, and Smith, 1994; Abreu, 1981; Fudenberg and Kreps, 1987; Fudenberg and Maskin, 1986; Fudenberg and Tirole, 1995; Greif, 1989; Kreps and Wilson, 1982; Kreps et al., 1982; Milgrom and Roberts, 1982; Selten, 1978; Sorin, 1992; Wen, 1994). Scholars have sought to understand attitudes towards deterrence based on the psychology of individuals (Goldgeier, 1994; Goldgeier, 1997) and the cultures of organizations and nations (Johnston, 1998; Johnston, 1995; Kier, 1995; Van Evera, 1984). Laboratory experiments have also begun to be used (Tingley and Walter, 2011a; Tingley and Walter, 2011b). A number of scholars have questioned whether leaders draw or act upon reputational inferences at all (Hopf, 1994; Jervis, 2002; McMahon, 1991; Mercer, 1996; Press, 2005; Tang, 2005), though there are a number of reasons why we should be wary of putting too much confidence on this skeptical conclusion (see section 5.2). A strand of literature has looked specifically at the role that domestic audiences can play in bolstering the credibility of leader's statements and commitments (Fearon, 1994a; Haynes, 2012; Leventoglu and Tarar, 2005; Partell and Palmer, 1999; Schultz, 1998; Schultz, 1999) and at the role of leadership turnover (Chiozza and Goemans, 2011; Guisinger and Smith, 2002; McGillivray and Smith, 2005; McGillivray and Smith, 2008a; McGillivray and Smith, 2008b; McGillivray and Stam, 2004; Wolford, 2007). Issues of deterrence and credibility have been examined in the context of alliance formation and behavior (Crescenzi et al., 2012; Kydd, 2005; Leeds, 2003; Leeds, Long, and Mitchell, 2000; Leeds, Mattes, and Vogel, 2009; Leeds and Savun, 2007; Leeds et al., 2002; Miller, 2003; Morrow, 2000). More generally, scholars have looked at related phenomena, such as honor and revenge (Dolan, 2010; Kagan, 1995; O'Neill, 1999; Robinson, 2006), prestige (Markey, 1999; Morgenthau, 1948; O'Neill, 2006), status, standing, respect, and recognition (Lebow, 2008; Morrow, 2010; Wolf, 2011).

1.3 Reputation for Resolve

There are many facets of the intertwined phenomena of reputation, honor, prestige, respect, credibility, and deterrence. I focus on one aspect that is common to them and that is, on its own, extremely important to international conflict behavior: *concern for reputation for resolve*. Throughout this dissertation I will often refer to *concern for reputation for resolve*

by the simpler *concern for reputation*; in addition to making the exposition more concise, I believe this short form is theoretically appropriate because throughout most of history the kind of reputation that was of greatest concern to leaders is reputation for resolve. To develop the concept of *concern for reputation for resolve*, I first discuss *resolve*, and then *reputation for resolve.*

Resolve is a famously difficult concept to define. Fearon (1992, pp. 66-73) finds that resolve or "level of interest" is widely regarded as an important determinant of crisis outcomes, though it is otherwise hard to pin down precisely what is meant by these concepts. Resolve is related to a "state's strength and nerve in crisis bargaining" and its "willingness to run the risk of war" (66). Fearon finds in the rationalist literature many "arguments that resolve or level of interest is the critical factor" (66) determining crisis outcomes, though he notes that scholars discuss the concept using many different labels, including "balance of resolve, interests, will, motivation, bargaining power, critical risk, risk propensity, and autonomous risk." (66) Fearon finds that most uses of the concept imply that it has to do with willingness to use force, and that this willingness is not solely determined by material factors.

I define resolve as the probability that a state will use force (or more generally, will escalate a conflict to some high level, such as initiating war), *conditional on being in a crisis* (or more generally, having reached some threshold of escalation) for some particular issue or class of issues.[4] Having high resolve is not equivalent to being belligerent, to having low costs of confl ict, or to being powerful. Each of these may make a state resolved, but they may also be associated with a state having less resolve. The distinction between resolve and any other concept related to willingness to wage war is that resolve is defined *conditional on being in a crisis*. Belligerent powerful states will have greater unconditional willingness to wage war, but they will also select into more crises because they are more powerful and perhaps also because they derive additional benefits from probing the resolve of other powers (Sechser, 2010); inversely, a weak state will be (unconditionally) more reluctant to escalate disputes with others, but conditional on a weak state entering a crisis, they may be more willing to use force (be more resolved). For semantic clarity, a state's *military power* will refer to a state's ability to achieve favorable outcomes from an application of force (not from the bargaining dynamic); *costs of war* will refer to how sensitive a state is to the costs of the application of force; and a state's preferences over an issue in dispute (also sometimes called motivation) by the term *interest*.

The formal model introduced in chapter 2 will be extended in future work to systemati-

---

[4]Formally, then, resolve is a concept that depends on three variables: (l_1) the implicit starting level of conflict and (l_2) the ending level of conflict, and (I) the issue or set of issues in dispute. We could denote resolve as $Re(l_1, l_2, I)$ =probability of escalating to l_2, conditional on having escalated to l_1, for disputes over issues in the set I. Note that l_1 plays an important role, since it will define at what point particular reputations are "at stake", which will then determine how selection effects operate. Snyder and Borghard's (2011) criticism of domestic audience costs, and defense of the importance of national honor to domestic populations, can be represented as a disagreement over how frequently l_1 corresponds to an explicit threat or statement by a leader, as opposed to just general involvement in some dispute.

cally think through the effects of changing power, costs of war, interests, and other determinants of a state's utility on resolve. I conjecture that the models will show that decreases in the costs of war, increases in power or the value of the issue in the dispute, and decreases in the perceived resolve of the opponent *sometimes* increase a state's resolve, but not always. On the other hand, I show in chapter 2 that increases in the costs of backing down, perhaps due to reputational considerations such as honor or domestic audience costs, *always* translate into greater resolve (almost by definition). Thus, resolve is a central determinant of crisis bargaining outcomes; resolve is affected by power and interest, but does not relate to them in a clear unambiguous manner because of strategic selection; resolve is directly and clearly affected by concern for reputation for resolve, which will be greater in actors concerned with national honor, personal honor, or the comparable kind of domestic audience cost. In short, resolve is affected by power, interest, and concern for reputation for resolve, but only unambiguously by concern for reputation for resolve.

Furthermore, based on my reading of historical crises, I believe that concern for reputation is a substantively critical determinant of resolve, and perhaps the most interesting component, analytically and for policy. To see this, it is helpful to look at how the term resolve, and synonyms, are actually used. The Oxford English Dictionary[5] defines *resolve* as follows (I only list the more common usages):

1.a. The fact of having resolved upon a course of action, stance, etc.; a firm intention. Also in *pl.* in same sense (now somewhat *arch.*).
1.b. An act of resolving to do something; a resolution.
2. Firmness or steadfastness of purpose; determination; an instance of this...

And OED defines "resolved" as (I only list usages related to human behavior):
1.a. Of the mind, etc.: freed from doubt or uncertainty; settled. *Obs.*
1.b. Of a person: convinced, satisfied, or certain of something. *Obs.*
1.c. Of doctrine: adopted or accepted after careful deliberation. *Obs.*
2.a. Of a person: that has resolved to do something; having fixed intention; determined, decided.
2.b. Of an action, state of mind, etc.: fully determined upon, deliberate.
2.c. Of a person: staunch, dedicated; committed, confirmed; that is thoroughly committed to the specified or implied course of action, practice, religious belief, doctrine, etc.
3. Of a person, the mind, etc.: characterized by determination or firmness of purpose; resolute.
4. That has been decided or resolved *on* or *upon*.

As is clear from these definitions, the term resolve is often understood as referring to

---

[5]Third edition, March 2010; online version December 2011. http://www.oed.com/view/Entry/163732; accessed 13 March 2012. An entry for this word was first included in New English Dictionary, 1908.

the result of a process of deliberation, decision, and commitment. One *resolves* to do something. In this sense, resolve is not simply a byproduct of a weighing of the pros and cons of escalating, but is instead a decision that one makes and then *commits to.* When Pericles sought to persuade the Athenians to reject the Spartan ultimatum, he argued that the decision over whether to make even a trivial concession under coercion was a "test of your resolution" (*The Peloponnesian War*, p. 1.140.5; under Warner translation, "proof of your determination"). It was common knowledge that the final Spartan demand—rescinding of a minor piece of economic statecraft known as the Megarian Decree—was extremely modest and of little material consequence to the Athenians. Pericles and other Athenians were not concerned that they would get a reputation for having a particular set of *interests*, such as not caring about the Megarian Decree. It is, after all, largely futile for a democracy, such as Classical Athens, to try to build a reputation for having certain *interests* since the preferences of various groups are readily observed through their democratic process of public deliberation. As Pericles pointed out, what was at stake was Athens' reputation for committing to its position, which in this case was the principle that Athens and Sparta should treat each other as equals and overcome disputes through arbitration. A reputation for resolve is not a reputation for wanting to fight over many issues, but a reputation for being willing to fight for those issues that one commits to.

It is thus my contention that much of what constitutes resolve is the degree to which an agent (individual or group) *commits* to a course of action that it chooses to pursue; for semantic clarity we might call this *determination*.[6] Two groups may have made the same decision facing identical preferences, but one has greater determination and thus greater resolve. Determination will tend to be greater for groups with stronger and more persistent leadership (since the preferences of the group are more constant), that undergo a more careful process of deliberating and resolving which course of action to pursue[7], and that perceive greater costs for backing down on a decision (specifically, greater reputational costs).

Reputation for resolve refers to the beliefs of others about how resolved an actor is likely to be in any given crisis.[8] While the persistence of characteristics such as low costs

---

[6]OED provides two definitions of determination that pertain to this meaning: "The mental action of coming to a decision; the fixing or settling of a purpose; the result of this; a fixed purpose or intention" and "The quality of being determined or resolute; determinedness, resoluteness."

[7]One could create a model where different agents deliberate to different degrees. For example, a thoughtless agent could be modeled as having greater uncertainty about the costs of escalating. In deciding whether to escalate initially, that agent would make some mistakes. If the opponent fought back, this agent would then acquire more information about the true costs (because they would then deliberate more extensively about the path ahead), at which point those who had made a large optimistic error initially would back down. A "thoughtful" agent, on the other hand, who had more precise information initially would make fewer such reversals in their course of action. This model could also involve endogeneous determination of one's preferences, where an agent pays a cost to gain a more precise estimate of their utility for different actions; agents with different costs of deliberation would then behave in the manner described. I conjecture that agents who care more about reputation for resolve will pay more deliberation costs to reduce the risk that they will want to back down in the future.

[8]Mercer (1996, 15) and Tang (2005, 37) define resolve as "the extent to which a state will risk war to

of war, high military capability, or high valuation of a particular set of issues (such as territories in dispute) could generate a reputation for resolve, as the above section articulated, the connection between these is ambiguous; and hence the connection between these and persistent resolve and thus also inferences about persistent resolve are less clear. On the other hand, persistence in those characteristics that are more clearly associated with resolve—such as persistence in leadership, more careful deliberation, and greater costs to backing down— will probably be more reliably associated with greater resolve. As such, inferences about resolve will be informed by these characteristics and behaviors, and will inform beliefs about them. Wherever the inference comes from, a belief by others that one is likely to be resolved in subsequent crises is what is meant by saying that one has a reputation for resolve.

1.4 The Fundamental Importance of Reputation for Resolve

This section will take a step back to think about the concept of reputation for resolve more abstractly.

Consider first what is the function of reputation. A reputation is a belief about an agent's characteristics or commitments, based on the agent's past behavior, that informs predictions about the agent's future behavior. Reputations can have profound impacts on strategic dynamics; by influencing the beliefs of others about how an agent is likely to behave in any given circumstance, reputations can influence the actions of others. Consider the generic reputation game represented in figure 1.1. Reputation dynamics requires a game with at least two players, where one player would have incentives to acquire a reputation for behaving a certain way to persuade the other player to pursue a different action. The game in figure 1.1 is thus the simplest possible stage game that can illuminate incentives to acquire a reputation.[9]

In figure 1.1 we have two agents: A and B. I assume common knowledge of the utility payoffs, depicted in parentheses. B will then choose C if $z \geq y$, and NC otherwise. A knows this, and will make a similar utility calculation to decide whether to choose C or NC. A solves the game using backward induction. Both players know with certainty what will happen.

Without loss of generality we can assume that $y > z$. There are then $6 \times 3 = 18$ possible games, for all possible permutations of the preference rankings (omitting the possibility of indifference over preferences). It turns out, however, that all of these but three are trivial because they involve either complete harmony of interest, or a constellation of interests such

---

keep its promises and uphold its threats". In this definition of resolve, the baseline level of escalation (the l_1) is the issuance of a promise or threat, and the upper level of escalation (the l_2) is war. The set I is left unspecified, though it perhaps refer to all issues.

[9]A simultaneous move version of this game would add strategic uncertainty, and thus the extensive form version isolates better the incentives behind reputation building.
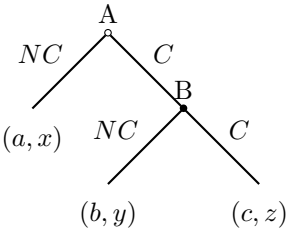
Figure 1.1: **Generic Reputation Game**

that providing either player with a commitment device could not alter the equilibrium. Of the remaining three, one of them is also trivial given the assumption of rational expectations.[10] The only two non-trivial games involve the following payoff combinations: (1) $b > a > c$ and $x > y > z$; (2) $c > a > b$ and $y > z > x$. I will consider each of these non-trivial games in turn.

1.4.1 Coercion Game: $b > a > c$ and $x > y > z$

Suppose that B is not happy with the outcome. That is, suppose the equilibrium has player A choosing C and player B choosing NC, denoted as $s = (C, NC)$, and $x > y > z$. This implies that B is "not happy with the outcome" and would rather A choose NC. If A would choose C no matter what (that is, $b > a$ and $c > a$) then there is nothing B could do. Suppose, however, that $c < a < b$, so that if B could convince A that B intended to play C then A would choose NC. At this point we have what I call a coercion game (or deterrence game), and in the literature was introduced as the chain-store game (Davis, 1985; Kreps and Wilson, 1982; Kreps et al., 1982; Milgrom and Roberts, 1982; Selten, 1978); this is also an extensive form version of chicken. It is the simplest way of representing the role of reputation in coercive encounters. We can put labels on the actions: for player A let C refer to "challenge", and NC refer to "no challenge"; for player B, C can refer to "conflict", NC "no conflict". The game can then be narrated as follows: A wants to challenge B for some territory, but will only do so if A thinks B will back down. A knows, however, that B will back down once A commits to an invasion, and so A mobilizes the military ("challenges") and B rationally chooses to not have a conflict. The dilemma facing B is that if B could only commit to entering a conflict (choosing C) then A would be deterred, B would be better off, and B would never have to exercise its commitment. This characterizes the essence of why being able to endogenously generate commitments can be so beneficial in coercive encounters (Schelling, 1960; 1966).

One means to generate such a commitment is through having the ability to generate a

---

[10]This game, defined by $c > a > b$ and $y > x > z$, gives player B incentives to trick A about what B will do. However, rational expectations requires that A cannot be tricked, so this is not possible. Here again, a commitment device can not help either player.

reputation for choosing conflict in such situations; once one has a reputation the value of preserving it can make it rational for one to choose conflict in any given situation, and the possession of such a reputation will then persuade potential challengers (A) to not challenge. Formally, if this game is iterated an infinite number of times, and B's discount factor is sufficiently large, then one equilibrium consists of A always choosing NC, and B always choosing C if ever given the chance. A is deterred by B's credible threat to engage in conflict over any challenge.

To summarize, when the payoffs are ranked so that $x > y > z$, $b > a > c$ then B would like to persuade A that B will choose C so as to persuade A to choose NC. B can benefit from a commitment device, such as being able to establish a reputation. This is a coercion reputation game because if B can establish a reputation then B will be able to coerce A into choosing NC, making B better off and A worse off from the equilibrium of the one-shot game.

1.4.2   Cooperation Game: $c > a > b$ and $y > z > x$

There is another situation in which reputation is beneficial: suppose that $c > a > b$ and $y > z > x$. The sub-game perfect equilibrium is $s = (NC, NC)$. Given the choice, B will choose NC; however, A strongly dislikes the outcome of $s = (C, NC)$, so A will avoid it by choosing NC. We can provide a narrative to the game: A would like to cooperate with B as long as B will cooperate back (C refers to cooperating). However, B has incentives to defect (not cooperate) given the choice. A, knowing this, will not trust B. Even though they would both be better off cooperating ($u_i = (C, C) > u_i(NC, NC), \forall i \in \{A, B\}$), B cannot persuade A that B will follow through on her promises. Once again, B is in need of a commitment device, and one such device is having a reputation for cooperating. If we iterate this game infinite times, then for a sufficiently large discount factor for B, it will be an equilibrium of the game that B chooses C, and A always chooses C so long as B has cooperated in the past. I refer to this game as the Cooperation Game; it is also an extensive form version of the Prisoner's Dilemma.

1.4.3   Importance of Commitment Devices, and hence Reputation

To summarize, we began with a very basic game form (illustrated in figure 1.1) involving two players, each with two possible moves, one player moving first (the sequencing is done to simplify the game further by eliminating strategic uncertainty). Of **all** possible rankings of payoffs for this game structure, only two involved strategic dynamics that could be modified by a commitment device, since the remainder involved sufficient harmony of interest. These two games look like the classic games of social science, Prisoner's Dilemma and Chicken, and, interestingly, encapsulate the two possible uses of a commitment device: (1) to enhance cooperation by making a promise credible, and (2) enhancing one player's coercion by making his threat credible.

Of the "games" that states engage in that can be characterized by this simple game structure, there could be many possible arrangements of the payoffs. However, the above analysis illustrates that only two of these have non-trivial strategic dynamics (given our assumption of perfect information). Thus, in all those games with trivial dynamics, agents would peacefully and easily solve the game, and thus these games would be historically and analytically uninteresting. It is the remaining two classes of games that are interesting, and it is these two classes that always provide agents with incentives to acquire commitment devices. Reputation is one such commitment device, and perhaps the most commonly used since whenever discount factors are sufficiently high (and play is repeated, and information can be transmitted so that reputations can form), reputations provide an endogenous and flexible means to commit to whatever behavior an actor would like. However, reputation as a commitment device is only possible when others believe that one values preserving one's reputation, and thus *concern for reputation* is what makes reputation as a commitment device possible. Agents thus should care more about their reputations to the extent that having a reputation is valuable, and should even cultivate concern for reputation (such as through generating domestic audience costs, the evolution of cultures of honor, etc...) to make this commitment device more powerful. In short, even in the idealized world with perfect information and no strategic uncertainty, there are strategic dynamics associated with commitment, and reputation provides a highly flexible and general means to make credible commitments. Concern for reputation should thus be universal whenever political entities have sufficiently long time horizons and exist in a community in which reputations can form.[11]

1.5 The Study of Concern for Reputation for Resolve (CRR)

Many scholars have studied phenomena related to concern for reputation for resolve, such as concern for honor (Kagan, 1995), prestige (Markey, 1999; Morgenthau, 1948), standing (Lebow, 2008; Lebow, 2010), status, and credibility (Press, 2005). The study of domestic audience costs in particular is the study of how a democratic electorate, or autocratic selectorate (Weeks, 2008), can punish a leader that breaks his word or otherwise harms the nation's honor (Snyder and Borghard, 2011).

---

[11]While in theory coercive reputations and cooperative reputations could co-exist, in practice they might be in tension since they lean on different kinds of institutions and preferences. As such, an interesting practical issue concerns the balance between coercive and cooperative reputation games in any given system. This dissertation focuses on coercive reputations, which is justified since throughout most of history leaders seem to be more concerned about their coercive reputations.

1.5.1 Sechser 2007

One scholar whose work comes most closely to the goals of this dissertation is Todd Sechser. Sechser (2007, p. 103) looks at whether "reputational incentives can help explain why compellent threats succeed or fail."[12] Sechser (2007, p. 189) writes, also providing a motivation for the contribution of this dissertation: "What has been missing [from scholarship] is a theory that helps explain when leaders will worry about their own reputations when facing external threats. This study provides such a theory. I argue that states have greater reputational incentives when their rivals are powerful, nearby, and have a history of making threats.... This is a substantial improvement over existing research, which asserts only that 'all states worry about their reputation for resolve to some degree' (Mercer 1996: 21)." In section 5.3 I discuss a number of other potential sources of variation in concern for reputation, including living in a hostile neighborhood, having many actual or potential rivals, being a colonial power, being a great power, having one's conflict behavior observed, not having a well formed reputation, requiring support of allies, and more generally having many issues (such as territory) that could be challenged. Barbara Walter's (2006) work on when governments violently repress demands by ethnic groups for territorial autonomy also looks at variation in concern for reputation, where her central source of variation is the number of potential rebel challengers.

1.5.2 Research Design Issues in the Study of CRR

The problem with the quantitative research strategies employed to date is that there are many potential confounds for the observed correlations; just because we can (perhaps correctly) identify a source of variation doesn't mean we can productively study it in a statistical analysis on observational data.

Consider each of Sechser's primary treatment variables. First, Sechser examines whether states are more likely to resist a threat, defined as involving complete compliance with the threat and less than 100 fatalities (2007, p. 120) issued by a more powerful state than by a weak state. But a powerful challenger is more likely than a weak state to impose 100 fatalities on another state in the act of using force to get its way; thus, any association between the power of a challenger and Sechser's measure of compellence success will be biased upwards. We should be cautious, then, of interpreting a positive partial correlation as evidence that the target was more resistant because of reputational incentives arising from the challenger's greater commitment problems. Similarly, a challenger may be more willing to use force that leads to 100 fatalities against a weaker target; again, the expected association arises but not from reputational incentives. Last, as Sechser also partly argues, powerful states may issue more threats, reducing their selectivity in their threats. Again, the observable implication of Sechser's theory confounds with another potential explanation.

---

[12]One of Sechser's major contributions is to refine the empirical referent to "compellent threats", avoiding what Sechser and Downes (2010) argue is a mistaken use of the Militarized Interstate Dispute dataset.

Sechser theorizes that a challenger that has issued more threats against a target in the past will suffer from a greater assurance problem: a target of such a threat will believe it likely that this challenger will make another demand in the future. As such, they will have greater incentives to resist this initial threat, rather than make the demanded concession. Again, though, there may be other reasons for this empirical association. States that issue many threats against a target may be less selective about their threats; a state that doesn't value its reputation for credibility will issue more threats and be less likely to follow through on them. This induces the same empirical implication as Sechser's theory.[13]

Finally, Sechser argues that geographic proximity will reduce the probability of successful compellent threats. But states that are close together may also find it easier to use force to try to achieve their objectives, and in the process to generate the 100 fatalities that will make this event coded as a failure of compellence.

I don't know if any of the above concerns are driving Sechser's results, and further analysis may be able to rule out these potential confounds. However, more generally, research based on available observational data that tries to overcome confounding through statistical models—also called model-based inference—will often suffer from the possibility of unobserved sources of bias. There are likely many other potential sources of bias, requiring us to be very cautious in our inferences. To overcome these substantial limitations, this dissertation will pursue "design-based inference" (Dunning, 2010; Sekhon, 2009; Sekhon and Titiunik, 2012), which is statistical causal inference on empirical domains that were carefully selected because unobserved biases are less likely to be a problem. I will look for sources of variation in concern for reputation for resolve that occur within a country or leader (so as to control for country or leader specific effects), that occurs due to a manipulation, and for which we understand the *treatment-assignment process* (that is, the reasons why our key independent variable takes on different values).

1.5.3 Selection Effects and Testable Implications of CRR

Another major challenge facing the empirical evaluation of concern for reputation for resolve, and reputation more generally, are selection effects. Agents who care more about reputation will respond to the possibility of conflict differently than those who care less. They may be more reluctant to engage in disputes over issues that are not that important to them; they may seek out disputes in which they can score reputational victories; others may be more reluctant to challenge them. Sechser derives his testable implications from a formal model that accounts for some selection processes: great powers are more likely to engage smaller powers in disputes to evaluate their resolve. Barbara Walter, similarly, theorizes that any given ethnic group should be less likely to challenge a government the more that government cares about its reputation for resolve, which is increasing in the number of ethnic groups

---

[13]Another, related, possibility is that states that have credible threats need not issue them to get what they want, since they can imply them. States that do not have credible implicit threats, then, must make greater efforts to communicate them, thus leading to more observed threats and less success in each.

in the country and in the value of territory that those ethnic groups might claim. Fearon (2002) more generally points out how studies of deterrence need to take into account the processes by which states select into a dispute, and how the effects of variables may change at different levels of escalation because of the process by which states rationally take into account these factors when deciding whether to select into a dispute.

1.6 Research Strategy

To summarize, there are two major challenges to studying concern for reputation. The first is that concern for reputation, like reputational phenomena more generally, often involves processes of strategic selection that complicate the observable implications of any given theory. The second is that testing even clean observable implications on the behavior of states is problematic because variation in the factors of interest do not generally arise as if from a randomized controlled experiment. My dissertation works to overcome these challenges through two strategies. First, I formally deduce observable implications from a theoretical model that explicitly takes into account processes of selection. Second, I search for opportunities in history in which relatively well-understood processes manipulate leaders' concern for reputation in a manner unrelated to most other phenomena; this allows me to better assess and address issues of confounding.

This dissertation also contributes to the literature by investigating an important question from a new methodological perspective. While many historians and political scientists have employed qualitative methods to evaluate the role of concern for reputation in international relations, there is little quantitative study of concern for reputation, and no such studies that build their research design so explicitly on the lessons of design-based inference. Donald Kinder (2011, p. 527), citing Donald Campbell, argues for the benefits of studying a research question using different methods: "pursue research questions from a variety of methodological angles, each one fallible, but fallible in different ways. Dependable knowledge is grounded in no single method, but rather in convergent results across complementary methods. Hypotheses prove their mettle by surviving confrontation with a series of equally prudent but distinct tests." This dissertation studies a phenomenon regarded as important by historians and some political scientists (especially classical theorists), but employing a new set of tools designed to deduce observable implications that take selection into account and to systematically evaluate causal effects.

To theorize about how concern for reputation relates to conflict behavior I construct a family of formal models based on the simplest extensive form game structure possible that allows for selection effects (which also happens to be very similar to the canonical crisis escalation models used in the literature). I make a number of theoretical contributions. I formally deduce a number of testable implications. I deduce as many testable implications as possible, ensuring that they are robust to modest modification of the game form. I explicitly articulate the assumptions necessary to link the formal model to the data, and I justify the

key assumptions. I derive as many testable implications as possible because that increases the opportunities for empirical evaluation of my theory and for testing it against alternative explanations (King, Owen Keohane, and Verba, 1994, p. X; this technique is also known as pattern matching, Campbell, 1975); empirical richness also has the advantage of enabling more powerful statistical tests of the theory, through the introduction of the Non-Parametric Combination (NPC) methodology (Pesarin and Salmaso, 2010).

I will examine the testable implications of my theory on two very distinct empirical domains: a comparison of conflict behavior under Southern and non-Southern US presidents, and a comparison of conflict behavior under leaders at different points in their time-in-office. Both of these empirical domains were selected because they offered a rare opportunity to examine variation in concern for reputation most likely to be independent of other confounding processes. I select two empirical domains, and I search for potential additional well-identified empirical domains (see section 5.3) because it is always possible that the results from any one (non-experimental) research design could be driven by some source of unconsidered bias. However, if we find the same rich pattern of results on multiple distinct empirical domains, we should have much greater confidence in the inferences that follow from these tests. In addition, these two empirical strategies are especially complementary because they leverage different sources of variation in concern for reputation. The study of US presidents leverages variation in the cultural origins of leaders; as such, the source of different levels of concern for reputation may be interpreted as more constructivist and sociological (though it need not be). On the other hand, variation in concern for reputation associated with leaders' time-in-office arises from a rationalist appraisal of the returns to behavior that build a reputation. In this sense these two empirical strategies are complementary because they show how theories of concern for reputation need not rely on any particular micro-foundation, rationalist or constructivist, but are in fact consistent with both.

Chapter 2

Model of Escalation and Reputation-Engagement

2.1 Chapter Overview

How should a leader who cares more about their reputation for resolve behave? Some scholars have intuitively argued that a leader with greater concern for a reputation for resolve should be more likely to experience militarized disputes and wars. The logic seems straightforward: leaders who care more about having a reputation for resolve will be less likely to back down and they may want to signal their resolve to others. Unfortunately, the strategic dynamics are not so straightforward. A greater cost for backing down will also make agents more cautious about entering disputes and less likely to bluff about their resolve. This will reduce the tendency for leaders concerned with reputation to initiate disputes or otherwise get into conflicts involving peripheral issues. To account for these countervailing effects, it is valuable to employ the theoretical apparatus of formal theory to carefully and systematically think through the strategic logic, effects, and implications of variation in concern for reputation.

I construct the simplest extensive form game structure possible that allows for selection effects. There will be two agents who take turns choosing their action. Each has the option of escalating the dispute or conceding. One player (P_1) will go twice, so that we can model selection effects. Players pay a material cost of escalating if the other player also escalates (if their opponent "fights back"). If P_1 escalates in the last round, the dispute is resolved through an exercise of force. Players pay a reputation cost for backing down after the conflict reaches a certain point. What characterizes the moment when reputation *becomes engaged* turns out to be a crucial feature of the model, and is something that scholars have not devoted sufficient attention to. I make the provisional assumption, with some justification, that reputation is engaged in disputes that are militarized. I allow for uncertainty over the costs of conflict, and in separate models not reported here I alternatively allow for uncertainty over the balance of power and the value of the prize. I take comparative statistics with

respect to concern for reputation for resolve, allowing that CRR is commonly known or private information. I also see if the observable implications are robust under models in which agents are "surprised" by the reputation engaging dispute, as well as under models where the agents can select into a dispute in which their reputation is on the line.

While scholars have examined crisis-bargaining models such as this from a number of perspectives, as far as I know the specific set-up and the testable implications derived here are novel. In addition to examining the effects of changes in CRR under a variety of different assumptions, I explicitly introduce the necessary auxiliary assumptions so as to formally link comparative statics to hypotheses that are testable on actual conflict data. This last contribution demonstrates the viability of the methodological approach advocated by scholars under the banner "Empirical Implications of Theoretical Models" (EITM); unlike much EITM work, however, the hypotheses I derive are based on a simple formal model and are shown to be robust to a breadth of functional forms, game structures, and other aspects of the model. This modeling exercise provides a means to discipline thinking and check intuitions through relatively straightforward models, rather than an opaque theoretical machinery giving rise to predictions for reasons not readily understood.

The models tell us the following. Conditional on reputation being engaged in a dispute, agents who care more about their reputation will be more likely to escalate the dispute further. This is true regardless of which side was responsible for the action that engaged reputation. Before reputation is engaged, predictions are ambiguous. Agents who care more about their reputation may be less likely, equally likely, or more likely to initiate a dispute, depending on details of the model; these ambiguous results run contrary to the initial intuition of many, including me, though once we walk through the strategic logic the reasons for it should become clear.

I will try to convey the intuition behind this result. (1) The direct effect of an increase in concern for reputation is to make an agent less likely to back down once reputation is engaged. (2) Knowing this, an opponent will be less likely to initiate a dispute that engages the agent's reputation[1], and (3) will be more likely to back down if challenged. (4) An agent who cares more about reputation will have conflicting incentives in regards whether to initiate a dispute: this agent will be less likely because they are less willing to bluff, but they are more likely because their opponent will be more likely to back down. (5) Since only tougher types of opponents will choose to challenge an agent who is more concerned with his reputation, if an opponent selects into a dispute then that opponent will be less likely to back down.

One of the central findings of the model is that the results depend critically on when and how reputation becomes engaged to a dispute. Comparative statics are robust above the reputation-engagement threshold, but are ambiguous below the threshold. To test the implications of this model, I hypothesize that reputation for resolve is engaged once a dispute

---

[1]Though, conditional on initiating a dispute, an opponent will be more resolved because weaker types are more likely to select out.

is militarized. From this it follows that leaders who care more about their reputation will, conditional on being in a conflict that is militarized, experience more subsequent escalation (fatalities, duration, uses of force) than leaders who care less about reputation and will be more likely to have positive outcomes. The benefit of drawing out as many predictions as possible is that it allows me to subject my theory to as stringent a test as possible by testing each prediction; alternatively, when the data is not sufficiently abundant or does not have sufficient variation to provide a powerful test on a single dependent variable, I am able to combine the multiple predictions into a single, much more powerful, statistical test of the theory using the Non-Parametric Combination technique.

2.2 Summary and Analysis of the Formal Models

2.2.1 Summary of the Models

I develop and solve four members of a family of models. These four models all build from the same basic game, but vary on key assumptions. I examine this family of models, rather than an individual model, to ensure that my predictions are robust to modification of certain key assumptions. The first two models, referred to as models *with private CRR*, examine differences in CRR that are private information; these models are consistent with the view that others do not realize that there is a systematic difference in concern for reputation for resolve for different kinds of leaders (e.g. Southerners and non-Southerners; or leaders early in their time-in-office vs later in their time-in-office). The second two models, referred to as models *with observable CRR*, assume that the differences in concern for reputation is observed and commonly known.

Within each of these groups, I consider a model in which the focal-agent (denoted **Model F** for focal-agent) and a model in which the rival-agent (denoted **Model R** for rival-agent) makes the first move. The actor who makes the first move (denoted P_1) is the actor who decides whether to carry out the action that engages the reputation of both parties. I assume that current theory and datasets are unable to empirically identify which country is responsible for the reputation-engaging event. Therefore, the population of Militarized Interstate Disputes (MIDs) can be regarded as a mixture of conflicts generated by Models F and R. I thus look for comparative statics that are present in both Models F and R.

Two agents, P_1 and P_2, both want to resolve some issue, commonly valued at $v > 1$, in their favor (e.g., claim territory, restructure the interstate system, renegotiate their strategic relationship). The game is three rounds long, as illustrated in Figure 2.2.1. P_1 first decides whether to escalate the conflict ($\omega_1^1 = E_1$, where ω_i^j denotes the strategy of player i in round j), such as through a verbal commitment or threat, or to concede the issue ($\omega_1^1 = C_1$). If P_1 escalates then both parties' reputations become engaged to the dispute. P_2 then similarly decides whether to concede ($\omega_2^2 = C_2$) or to further escalate the issue through a use of force ($\omega_2^2 = E_2$). If P_2 concedes, then neither side pays any material cost of conflict, but P_2 pays a

reputational cost $R_2 + r_2$, where $R_2 \geq 0$, $r_2 \geq 0$. If P_2 uses force, P_1 pays an immediate cost of c_1/k, where $k > 1$. Following escalation by P_2, P_1 decides whether to concede ($\omega_1^3 = C_3$), and absorb the reputational cost of $R_1 + r_1$ ($R_1 \geq 0, r_1 \geq 0$), or escalate one last time by using force ($\omega_1^3 = E_3$). If P_1 uses force, this portion of the game ends and the issue in dispute is resolved by some contest function that yields the prize to P_1 with probability π and P_2 with probability $(1 - \pi)$, P_1 pays an additional cost of c_1, and P_2 pays a cost of c_2. The material costs of conflict c_1 and c_2 are private information drawn from continuous cumulative distribution functions $F_1(x)$ and $F_2(x)$, where $\forall \epsilon > 0, i \in \{1, 2\} : F_i(\epsilon) > 0$. These have corresponding density functions $f_1(x)$ and $f_2(x)$, defined so that their support is a convex set.[2] A component of the reputational cost is common knowledge (R_1 and R_2), and a component is private information (r_1 and r_2) drawn from continuous cumulative distribution functions $G_1(x)$ and $G_2(x)$ with density functions $g_1(x)$ and $g_2(x)$. In models with observable honor, I set $r_1 = r_2 = 0$ and examine comparative statics with respect to observable honor: R_1 or R_2. In models with private CRR, I consider comparative statics with respect to private CRR: r_1 or r_2. All other aspects of the game are common knowledge.

Let p_1, q_2, and p_3 denote the probability that the relevant agent escalates in Round 1, 2, or 3, respectively, in equilibrium from the perspective of an observer who is unable to see the player's private costs (c_1 or c_2) but can see the player's private reputational concern (r_1 or r_2). These parameters capture the key observable implications of the model for an analyst who can detect differences in r, as I presume to do. Let $\hat{p}_3$ denote P_2's beliefs about the probability that P_1 will escalate in Round 3 (and similarly for $\hat{q}_2$). Note that, from Assumption 1, bargaining behavior prior to the dispute becoming militarized occurs in Round 1. It is in this sense that I model, and hence account for, selection into a MID. Since extant datasets do not provide systematic data on bargaining behavior before disputes have become militarized, predictions on p_1 can not be tested on the MID dataset.

2.3 Equilibrium Strategies and Beliefs

I restrict my attention to portions of the parameter space that result in "interior" equilibrium solutions, which best reflect the logic of crisis bargaining.[3] Given these conditions, in any

---

[2]That is, $\forall y, x, z, i \in \{1, 2\}$, where $y < x < z$, $f_i(y) > 0$, and $f_i(z) > 0$, then $f_i(x) > 0$

[3] Specifically, I impose the following two conditions:

Condition C: $\hat{q}_2 < q_2^C \equiv \frac{vk}{(R_1 + r_1)(1+k) + v(p+k)}$. Condition C guarantees that I only consider conflicts in which there is some probability that P_1 will back down in Round 3. In a model with rational expectations, $\hat{q}_2 = q_2^*$, therefore Condition C requires $q_2^* < q_2^C$ (where q_2^* denotes the equilibrium level of q_2 from the perspective of an observer who doesn't see r_2).

Condition I: The upper bound on the support of the distribution for c_1 is greater than $c_1^C \equiv \frac{k(v(1-\hat{q}_2) - (R_1 + r_1)\hat{q}_2)}{\hat{q}_2}$. Condition I ensures that some P_1 types will not escalate in the first round.
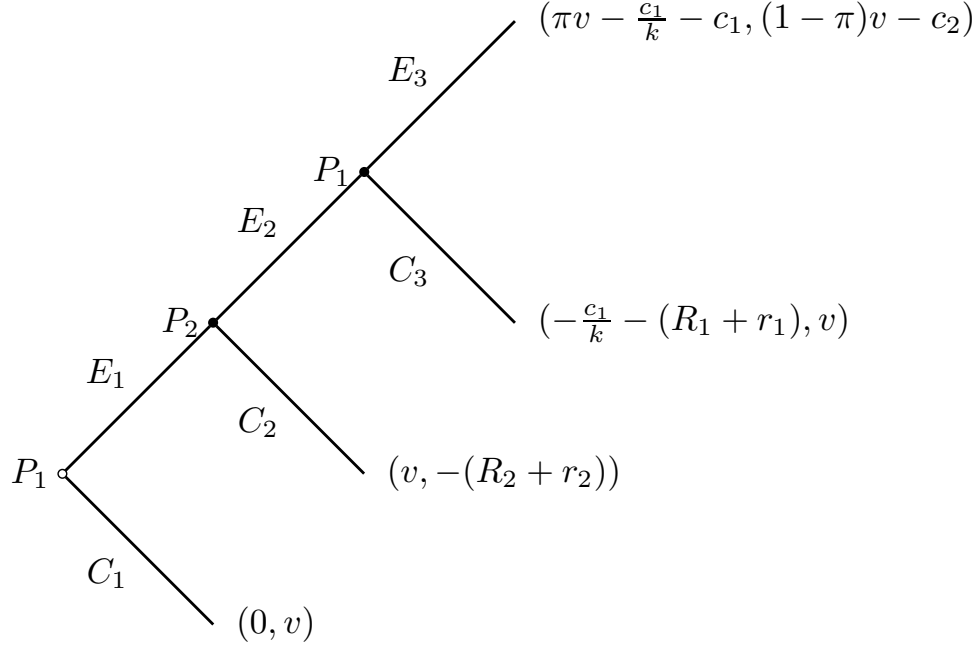
Figure 2.1: **Model of Escalation and Reputation Engagement.**
E_1 denotes a threat or commitment, E_2 and E_3 a use of force, C_i a concession on the issue
in dispute in Round i. $p_1 = Pr(\omega_1^1 = E_1|r_1)$; $q_2 = Pr(\omega_2^2 = E_2|r_2)$; $p_3 = Pr(\omega_1^3 = E_3|r_1)$

---

perfect Bayesian equilibrium the strategies of P_1 and P_2 are, respectively,

$$\omega_1 = \begin{cases} \{C_1, \cdot\} & \text{if } c_1 > c_1^C \equiv \frac{k(v(1-\hat{q}_2)-(R_1+r_1)\hat{q}_2)}{\hat{q}_2} \\ \{E_1, C_3\} & \text{if } \pi v + R_1 + r_1 < c_1 < c_1^C \\ \{E_1, E_3\} & \text{if } c_1 < \pi v + R_1 + r_1 \end{cases} \qquad \text{and} \qquad \omega_2 = \begin{cases} \{C_2\} & \text{if } c_2 > \frac{v(1-\pi\hat{p}_3)+(R_2+r_2)}{\hat{p}_3} \\ \{E_2\} & \text{otherwise} \end{cases}$$

It follows that

$$p_1 = Pr(c_1 < c_1^C(r_1)|r_1) = Pr\left(c_1 < \frac{k(v(1-\hat{q}_2)-(R_1+r_1)\hat{q}_2)}{\hat{q}_2}\middle|r_1\right)$$

$$q_2 = Pr\left(c_2 < \frac{v(1-\pi\hat{p}_3)+(R_2+r_2)}{\hat{p}_3}\middle|r_2\right)$$

$$p_3 = \frac{Pr(c_1 - (R_1+r_1) < pv|r_1)}{Pr(c_1 + (R_1+r_1) < k(v(1-\hat{q}_2))/\hat{q}_2|r_1)}$$

where $\hat{p}_1$, $\hat{q}_2$, and $\hat{p}_3$ are the players' beliefs about their opponents' probability of escalating
the conflict at the relevant stage.

Given rational expectations, $\hat{q}_2(\hat{p}_3) = \bar{q}_2(\bar{p}_3)$, and $\hat{p}_3(\hat{q}_2) = \bar{p}_3(\bar{q}_2)$, where $\bar{q}_2$ and $\bar{p}_3$ denote the expected probability of escalation for a given set of beliefs from the perspective of an observer who cannot see r_1 or r_2. Since $\bar{q}_2(\bar{p}_3)$ and $\bar{p}_3(\bar{q}_2)$ cross each other once and only once, there is a unique equilibrium.

2.4 Analysis of the Model

This section works through the model to demonstrate that the above strategies and beliefs represent the unique perfect Bayesian equilibrium.

Round 3 (Final Round)

P_1's best response function is:

$$BR = \begin{cases} E_3 & \text{if } c_1 - (R_1 + r_1) \leq \pi v \\ C_3 & \text{if } c_1 - (R_1 + r_1) > \pi v \end{cases} \tag{2.1}$$

P_1 escalates with probability:

$$p_3 = Pr(c_1 < \pi v + R_1 + r_1 | r_1)$$

This probability statement is written conditional on r_1 to indicate that p_3 is conditional on the realization of r_1.

Round 2

P_2 escalates in Round 2 with probability:

$$q_2 = Pr\left(c_2 < \frac{v(1 - \pi \hat{p}_3) + (R_2 + r_2)}{\hat{p}_3} \bigg| r_2\right)$$

Round 1

$EU_1(C_1) = 0$ (since no reputation costs)
$EU_1(E_1, C_3) = (1 - \hat{q}_2)v + \hat{q}_2(-(R_1 + r_1) - c_1/k$
$EU_1(E_1, E_3) = (1 - \hat{q}_2)v + \hat{q}_2(\pi v - c_1 - c_1/k)$
    We know from Round 3 that
    $EU_1(E_1, E_3) > EU_1(E_1, C_3) \iff c_1 - (R_1 + r_1) < \pi v$. Denote **Case E** (for escalate) when $c_1 - (R_1 + r_1) < \pi v$ and **Case C** (for concede) when $c_1 - (R_1 + r_1) > \pi v$. Denote the value of c_1 for which P_1 will be indifferent between E_1 and C_1 under Case E for a given r_1 as $c_1^E \equiv \frac{kv(1 - \hat{q}_2 + \hat{q}_2\pi)}{\hat{q}_2(k+1)}$, and the value of c_1 for which P_1 will be indifferent between E_1 and C_1 under Case C as $c_1^C(R_1 + r_1) \equiv \frac{k(v(1 - \hat{q}_2) - (R_1 + r_1)\hat{q}_2)}{\hat{q}_2}$.
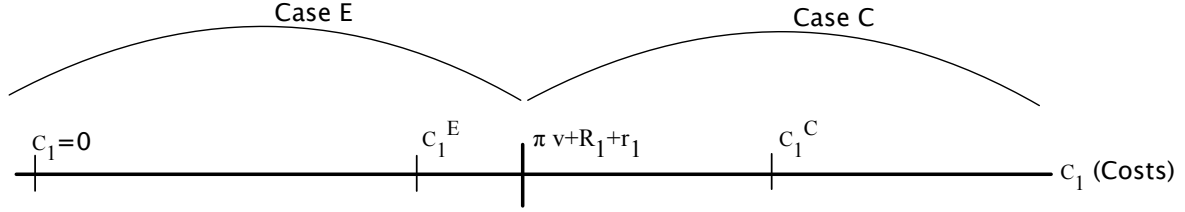
Figure 2.2: **Cases and Indifference Points for P_1's Costs**

Within case E, P_1 will prefer E_1 when $c_1 < c_1^E$ and not otherwise. Within case C, P_1 will prefer E_1 when $c_1 < c_1^C$.

Therefore, the probability that P_1 will escalate initially is:

$$p_1 = Pr(c_1 < \pi v + (R_1 + r_1))Pr(c_1 < c_1^E) + Pr(c_1 > \pi v + (R_1 + r_1))Pr(c_1 < c_1^C(R_1 + r_1))$$

2.4.1 Limiting the Parameter Space Under Consideration

There are three possibilities to consider: (1) **Situation EC**: c_1^E is interior to Case E and c_1^C is interior to Case C; (2) **Situation E**: c_1^E is interior to Case E and c_1^C is not interior to Case C (so that for all values in case C, P_1 prefers to concede); (3) **Situation C**: c_1^E is not interior to Case E and c_1^C is interior to Case C. Situation E is problematic because a change in R_1 or r_1 will have no effect on p_1, q_2, or p_3 since in this equilibrium Player 1 never backs down after initially escalating. Situation E is more representative of the dynamic of pure force, rather than the logic of crisis bargaining. Situation EC consists of a combination of the results from Situation E and Situation C. To simplify the analysis, I exclude formal consideration of Situation E and Situation EC.

2.4.2 Only C Indifference Point Interior? (Denoted Situation C)

In order for just c_1^C to be interior to Case C I need that $c_1^C > \pi v + (R_1 + r_1)$ and $\pi v + (R_1 + r_1) < c_1^E$. This implies:

$$k(v(1 - \hat{q}_2) - (R_1 + r_1)\hat{q}_2)/\hat{q}_2 > \pi v + (R_1 + r_1)$$

and

$$\pi v + (R_1 + r_1) < \frac{1 - \hat{q}_2 + \pi \hat{q}_2}{\hat{q}_2} \frac{vk}{k + 1}$$

Which imply:

$$\hat{q}_2 < \frac{vk}{(R_1 + r_1)(1 + k) + v(\pi + k)} = q_2^C$$

2.4.3 Solving the Model under Situation C

Situation C requires that:

Condition C: $\quad \hat{q}_2 < q_2^C \equiv \frac{vk}{(R_1+r_1)(1+k)+v(\pi+k)}$

In a model with rational expectations, $\hat{q}_2 = q_2^*$, therefore Condition C requires that $q_2^* < q_2^C$ (where q_2^* denotes the equilibrium level of q_2 from the perspective of an observer who doesn't see r_2).

And in order to have an "interior" solution so that some P_1 types don't escalate in the first round, I need that:

Condition I: The upper bound on the support of the distribution for c is greater than

$$c_1^C \equiv \frac{k(v(1-\hat{q}_2) - (R_1+r_1)\hat{q}_2)}{\hat{q}_2}$$

Probability P_1 Escalates in First Round (p_1)

$$p_1 = Pr(c_1 < c_1^C(r_1)|r_1) = Pr\left(c_1 < \frac{k(v(1-\hat{q}_2) - (R_1+r_1)\hat{q}_2)}{\hat{q}_2}\bigg|r_1\right) \qquad (2.2)$$

Probability of P_2 Escalating (q_2)

$$q_2 = Pr\left(c_2 < \frac{v(1-\pi\hat{p}_3) + (R_2+r_2)}{\hat{p}_3}\bigg|r_2\right) \qquad (2.3)$$

Probability of P_1 Escalating in Final Round (p_3)

$$p_3 = \frac{Pr(c_1 - (R_1+r_1) < \pi v|r_1)}{Pr(c_1 + (R_1+r_1) < k(v(1-\hat{q}_2))/\hat{q}_2|r_1)} \qquad (2.4)$$

2.4.4 Existence and Uniqueness of Equilibrium

Given Conditions C and I, there is a unique perfect Bayesian equilibrium. To see this, let $\bar{q}_2$ denote the probability of P_2 escalating from the perspective of an observer who cannot observe r_2, and $\bar{p}_3$ the probability of P_1 escalating in Round 3 from the perspective of an observer who cannot observe r_1. These can be thought of as the "population" best response function since it maps beliefs about the other's strategy into an expected response.

That is,

$$\bar{q}_2(\hat{p}_3) = \int_{r_2=0}^{\infty} q_2(r_2, \hat{p}_3)g_2(r_2)dr_2$$

$$\bar{p}_3(\hat{q}_2) = \int_{r_1=0}^{\infty} p_3(r_1, \hat{q}_2) g_1(r_1) dr_1$$

In equilibrium the beliefs of the agents must be correct: $\hat{q}_2 = \bar{q}_2$, and $\hat{p}_3 = \bar{p}_3$. From 2.3 and 2.4 We see that, in equilibrium, $\frac{d\bar{q}_2(\bar{p}_3)}{d\bar{p}_3} < 0$ and $\frac{d\bar{p}_3(\bar{q}_2)}{d\bar{q}_2} > 0$ (and therefore $\frac{d\bar{q}_2(\bar{p}_3)}{d\bar{p}_3} < 0$ and $\frac{d\bar{p}_3(\bar{q}_2)}{d\bar{q}_2} > 0$). These expressions have the natural interpretation that P_2 will be less likely to escalate as P_1 is more likely to escalate in Round 3, and P_1 will be more likely to escalate in Round 3 as P_2 is more likely to escalate in Round 2 (because the types of P_1 that select into Round 2 will be more likely to be resolved). I am also able to deduce where the curves $\bar{p}_3(\bar{q}_2)$ and $\bar{q}_2(\bar{p}_3)$ touch their extreme points, which allows us to show that the curves cross each other once and, by monotonicity, only once (see figure 2.3). Specifically, I know that when $\bar{q}_2 = 0$, there will be a proportion of P_1 types that escalate in Round 3 greater than 0 and less than 1; as $\bar{q}_2$ approaches 1, $\bar{p}_3$ will also approach 1 (and will hit it when $\bar{q}_2 = q_2^C$); as $\bar{p}_3$ approaches 0, $\bar{q}_2$ approaches 1 (and reaches 1 if the distribution on c_2 has an upper bound); when $\bar{p}_3 = 1$, $\bar{q}_2$ equals some positive value (since there are always cost types sufficiently low, that is $c_2 < v(1 - \pi) + (R_2 + r_2)$). Thus, given continuity in $\bar{q}_2$ and $\bar{p}_3$ and **Condition C** being satisfied, there will be a unique equilibrium.

2.4.5   Private CRR

Model F with Private CRR

I now consider specific models. Model H with Private CRR refers to the setup in which P_1 is the CRR-agent, so that some types of P_1 have high realized values of r_1 ("Southerners") and others have low realized values. Thus, I am looking for comparative statics in values of r_1.

First I note that since P_2 doesn't observe the realized value of r_1, $\frac{d\hat{p}_3}{dr_1} = 0$. By equation 2.3 this implies

$$\frac{dq_2}{dr_1} = 0$$

This then implies (by equations 2.2 and 2.4) that:

$$\frac{dp_1}{dr_1} < 0$$

and

$$\frac{dp_3}{dr_1} > 0$$

In words, these results state that for conflicts in which the focal-agent performs the reputation engaging action the focal-agent will be more likely to escalate above the reputation-engagement threshold, and less likely to escalate below the reputation-engagement threshold. The behavior of the rival-agent will not vary since, by assumption, they are unable to perceive a difference between focal-agents with greater or less concern for reputation.
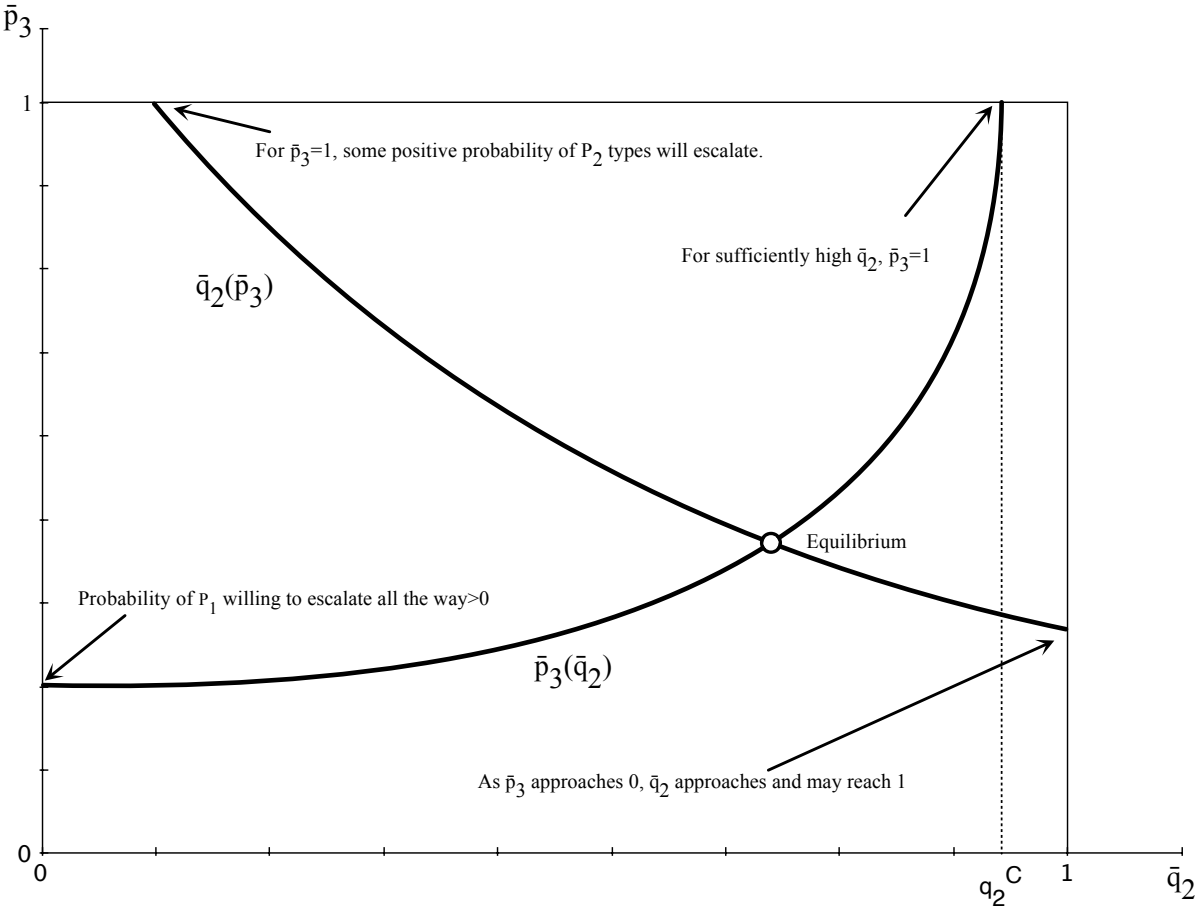
Figure 2.3: **Unique Equilibrium**

Model R with Private CRR

Model R refers to the setup in which P_2 is the focal-agent, so that I am looking for comparative statics in realized values of r_2.

Since P_1 doesn't observe the realized value of r_2, we have (by 2.2 and 2.4):

$$\frac{dp_1}{dr_2} = 0$$

$$\frac{dp_3}{dr_2} = 0$$

This then implies that (by 2.3):

$$\frac{dq_2}{dr_2} > 0$$

In words, this result states that for conflicts in which the rival-agent performs the action that crosses the focal-agent's reputation-engagement threshold, the focal-agent will be more likely to escalate. As before, the rival-agent's behavior will not vary since, by assumption, they are unable to perceive a difference between focal-agents with greater or less concern for reputation.

Conclusion

In summary, the selection dynamic in models with private CRR is straightforward: agents with higher CRR pay higher costs for making a failed bluff and so are less willing to engage their reputation to a dispute. Both the higher costs of backing down and the biased selection of tough types into conflicts implies that agents with higher CRR are more likely to escalate once their reputation is engaged. These comparative statics form the basis for the hypotheses tested in this paper.

Consideration of models with observable CRR complicates these predictions because now the rival-agent will tend to select out of disputes in which the CRR-agent is likely to be resolved. In these models, it matters far more which agent is the one responsible for leading the conflict over the reputation-engagement threshold since their equilibrium behavior will then be shaped by selection effects. With suitable auxiliary assumptions, however, I show that these core comparative statics are probably reasonable even in situations with observable CRR.

2.4.6   Model F and R with Observable CRR

Under these models I set $r_1 = 0$ and $r_2 = 0$, so that given rational expectations we have $\hat{p}_3 = p_3$ and $\hat{q}_2 = q_2$. As above, let the cumulative distribution function of cost types for P_1 be $F_1(x)$ and for P_2 be $F_2(x)$. Let $E_1(x) = F_1(x)/F_1(c_1^C)$ for $x < c_1^C$ and $E_1(x) = 1$ for $x \geq c_1^C$. This, then, is the rescaled cumulative distribution function for the portion of $F_1(x)$ below c_1^C. We can then rewrite equations $2.2, 2.3, 2.4$ as:

Probability of P1 Escalating in Last Round (p_1)

$$p_1 = F_1(c_1^C) = F_1\left(\frac{k(v(1-q_2) - R_1 q_2)}{q_2}\right) \tag{2.5}$$

Probability of P2 Escalating (q_2)

$$q_2 = F_2\left(\frac{v(1 - \pi p_3) + R_2}{p_3}\right) \tag{2.6}$$

Probability of P1 Escalating in Final Round (p_3)

$$p_3 = E_1(\pi v + R_1) = \frac{F_1(\pi v + R_1)}{F_1\left(\frac{k(v(1-q_2) - R_1 q_2)}{q_2}\right)} \tag{2.7}$$

Note that by Condition C, $\pi v + R_1 < c_1^C$, which implies $p_3 < 1$.

Comparative Statics for Models with Observable CRR

$$\frac{dp_3(R_1)}{dR_1} = \frac{dE_1(\pi v + R_1)}{dR_1} = \frac{F_1(c_1^C(R_1))\frac{\partial F_1(\pi v + R_1)}{\partial(\pi v + R_1)} - F_1(\pi v + R_1)\frac{\partial F_1(c_1^C(R_1))}{\partial c_1^C(R_1)}\frac{d(c_1^C(R_1))}{dR_1}}{(F_1(c_1^C(R_1)))^2} \tag{2.8}$$

$$\frac{dq_2(R_1)}{R_1} = -\frac{\partial F_2\left(\frac{v(1 - \pi p_3) + R_2}{p_3}\right)}{\partial\left(\frac{v(1 - \pi p_3) + R_2}{p_3}\right)}\frac{\frac{dp_3(R_1)}{dR_1}(R_2 + v)}{(p_3(R_1))^2} \tag{2.9}$$

$$\frac{dp_1(R_1)}{dR_1} = \frac{\partial F_1(c_1^C(R_1))}{\partial(c_1^C(R_1))}\frac{dc_1^C(R_1)}{dR_1} = \frac{\partial F_1(c_1^C(R_1))}{\partial(c_1^C(R_1))}(-k)\left(1 + \frac{v\frac{dq_2(R_1)}{dR_1}}{(q_2(R_1))^2}\right) \tag{2.10}$$

$$\frac{dp_3(R_2)}{dR_2} = \frac{dE_1(\pi v + R_1)}{dR_2} = \frac{kvF_1(\pi v + R_1)\frac{\partial F_1(c_1^C(R_2))}{\partial c_1^C(R_2)}\frac{dq_2(R_2)}{dR_2}}{(F_1(c_1^C(R_2)))^2(q_2(R_2))^2} \tag{2.11}$$

$$\frac{dq_2(R_2)}{R_2} = \frac{\partial F_2(\frac{v(1 - \pi p_3(R_2)) + R_2}{p_3(R_2)})}{\partial\left(\frac{v(1 - \pi p_3(R_2)) + R_2}{p_3(R_2)}\right)}\left(\frac{p_3(R_2) - \frac{dp_3(R_2)}{dR_2}(R_2 + v)}{(p_3(R_2))^2}\right) \tag{2.12}$$

$$\frac{dp_1(R_2)}{R_2} = -\frac{kv\frac{\partial(F_1(c_1^C(R_2)))}{\partial(c_1^C(R_2))}\frac{d(q_2(R_2))}{dR_2}}{(q_2(R_2))^2} \tag{2.13}$$

Proposition 1.

$$\frac{dp_3}{dR_1} > 0$$

Proposition 2.
$$\frac{dq_2}{dR_2} > 0$$

Propositions 1 and 2 state that as an agent's concern about his reputation increases, the agent will be more likely to escalate conditional on being above the reputation-engagement threshold. That is, agents that are more concerned about reputation that have either selected into a conflict or have been surprised by a reputation-engaging challenge will be more likely to escalate.

Proof of $\frac{dp_3}{dR_1} > 0$ by contradiction. Suppose that $\frac{dp_3}{dR_1} \leq 0$. But then equation 2.9 and the fact that all the terms therein, except $\frac{dp_3}{dR_1}$ and the negative sign, are necessarily positive, implies:

$$\frac{dq_2}{dR_1} \geq 0$$

$$\Rightarrow \left(1 + \frac{v \frac{dq_2(R_1)}{dR_1}}{(q_2(R_1))^2} \right) \geq 0$$

This implies, by equation 2.10, that

$$\frac{\partial c_1^C(R_1)}{\partial R_1} < 0 \text{ and } \frac{dp_1(R_1)}{dR_1} < 0$$

But this implies by equation 2.8 that

$$\frac{dp_3(R_1)}{dR_1} > 0$$

contradicting my starting assumption.

$\square$

Proof of $\frac{dq_2}{dR_2} > 0$ by contradiction. Suppose that $\frac{dq_2}{dR_2} \leq 0$. Then, by equation 2.13, $\frac{dp_1}{dR_2} \geq 0$ and by equation 2.11 $\frac{dp_3}{dR_2} \leq 0$. But then, by equation 2.12, $\frac{dq_2}{dR_2} > 0$ (so long as $p_3 > 0$ which will be true for Situation C so long as the lower bound on the support of $F(\cdot)$ is lower than $\pi v + R_1$ which we assume to be the case). Contradiction. $\square$

Proposition 3.
$$\frac{dq_2}{dR_1} < 0$$

Proposition 4.
$$\frac{dp_1}{dR_2} < 0$$

Propositions 3 and 4 state that rival-agents are less likely to escalate over the reputation-engagement threshold when facing an CRR-agent.

Proof of $\frac{dq_2}{dR_1} < 0$. Equation 2.9 and Proposition 1 ($\frac{dp_3}{dR_1} > 0$) implies $\frac{dq_2}{dR_1} < 0$. $\square$

Proof of $\frac{dp_1}{dR_2} < 0$. Equations 2.13 and Proposition 2 imply that $\frac{dp_1}{dR_2} < 0$

$\square$

Proposition 5. $\frac{dp_3}{dR_2} > 0$

Proof of $\frac{dp_3}{dR_2} > 0$. Equations 2.11 and Proposition 2 imply that $\frac{dp_3}{dR_2} > 0$

$\square$

Proposition 5 states that rival-agents who have selected into the conflict will be more likely to escalate against agents with higher CRR. The logic here is that as P_2 becomes more concerned about CRR, P_2 is less likely to back down in Round 2. Consequently, fewer P_1's are willing to gamble that P_2 will back down, whereas a greater proportion of P_1's who escalate in Round 1 intend to escalate again in Round 3.

Proposition 6.

$$\frac{dp_1}{dR_1} > 0 \iff \left(1 + \frac{v\frac{dq_2(R_1)}{dR_1}}{(q_2(R_1))^2}\right) < 0$$

$$\iff \frac{dq_2(R_1)}{dR_1} < -\frac{(q_2(R_1))^2}{v}$$

This follows immediately from equation 2.10.

Proposition 6 states that an increase in concern for reputation may lead an CRR-agent, below the reputation-engagement threshold, to escalate more or less. The CRR-agent will escalate more at the pre-engagement stage if the rival is sufficiently sensitive to an increase in the CRR-agent's concern for reputation, if v is sufficiently large, and/or if q_2 is sufficiently low.

2.4.7 Signing $\frac{dp_1}{dR_1}$ for Specific Distributions

To sign $\frac{dp_1}{dR_1}$ I need to determine the actual equilibrium solutions. For uniform uncertainty this is done analytically. For logistic uncertainty this is done numerically.

Uniformly Distributed Types

I now assume that the costs are drawn from a uniform distribution. $c_1 \sim U[L_1, H_1]$, $c_2 \sim U[L_2, H_2]$, where $H_1 > 0$, $H_2 > 0$. In order to make this model tractable, an additional simplifying assumption is later made that $L_1 = L_2 = 0$. For convenience I also assumed that $\pi = 1/2$.

$$F_1(x) = \int_{L_1}^{x} \frac{1}{H_1 - L_1} = \frac{x - L_1}{H_1 - L_1}$$

$$F(c_1^C) = \frac{c_1^C - L_1}{H_1 - L_1}$$

$$E_1(x) = \begin{cases} \frac{x - L_1}{H_1 - L_1} \frac{H_1 - L_1}{c_1^C - L_1} & \text{if } x < c_1^C \\ 1 & \text{if } x \geq c_1^C \end{cases}$$

$$\Rightarrow E_1(x) = \begin{cases} \frac{x - L_1}{c_1^C - L_1} & \text{if } x < c_1^C \\ 1 & \text{if } x \geq c_1^C \end{cases}$$

$$F_2(x) = \frac{x - L_2}{H_2 - L_2}$$

Therefore, setting $L_1 = L_2 = 0$ and remembering that in equilibrium $c_1^C = \frac{k(v(1-q_2) - R_1 q_2)}{q_2}$ I have that:

$$p_3 = \frac{q_2(v/2 + R_1)}{k(v(1 - q_2) - R_1 q_2)}$$

$$q_2 = \frac{v(2 - p_3) + 2R_2}{H_2 2 p_3}$$

Let $Z = -v(2R_1 + v) + 4k(R_1 + v)(R_2 + v)$. $Z > 1$ for the permitted values of the parameters. Solving for p_3 gives.

$$p_3 = \frac{Z - \sqrt{16kv(2H_2 + R_1 + v)(2R_1 + v)(R_2 + v) + (-Z)^2}}{4kv(2H_2 + R_1 + v)}$$

or

$$p_3 = \frac{Z + \sqrt{16kv(2H_2 + R_1 + v)(2R_1 + v)(R_2 + v) + (-Z)^2}}{4kv(2H_2 + R_1 + v)}$$

This shows that, within the numerator, the term under the radical sign is larger than the term outside the radical. Therefore, the first solution is always negative and should be discarded. The second solution is always positive. It is now possible to solve for q_2 and p_1:

$$q_2 = \frac{-4k(R_1 + v)(R_2 + v) - v(2R_1 + v)}{4H_2(2R_1 + v)}$$

$$+ \frac{\sqrt{16kv(2H_2 + R_1 + v)(2R_1 + v)(R_2 + v) + (v(2R_1 + v) - 4k(R_1 + v)(R_2 + v))^2}}{4H_2(2R_1 + v)}$$

$$p_1 = \frac{k(v(1-q_2) - R_1 q_2)}{H_1 q_2}$$

$$= -\frac{1}{16 H_1 H_2{}^2 (2R_1 + v)^2} k$$

$$\left(v(2R_1 + v) + 4k(R_1 + v)(R_2 + v)\right.$$

$$\left. -\sqrt{16kv(2H_2 + R_1 + v)(2R_1 + v)(R_2 + v) + (v(2R_1 + v) - 4k(R_1 + v)(R_2 + v))^2}\right) \times$$

$$\left(4H_2 v(2R_1 + v) + 4k(R_1 + v)^2(R_2 + v) + (R_1 + v)\right.$$

$$\left(2R_1 v + v^2 - \sqrt{v^2(2R_1 + v)^2 + 8kv(4H_2 + R_1 + v)(2R_1 + v)(R_2 + v) + 16k^2(R_1 + v)^2(R_2 + v)^2}\right)\right)$$

Proposition 7. *For Uniform Uncertainty, $\frac{dp_1}{dR_1} > 0$*

The proof of proposition 7 involves signing the derivative of p_1, which is too cumbersome to report here. Please contact the author for more information.

Logistically Distributed Types

The model was not analytically tractable for logistically distributed uncertainty, but through numerical methods it was possible to establish that $\frac{dp_1}{dR_1} < 0$ for many parameter values with logistic uncertainty (and $\frac{dp_1}{dR_1} > 0$ for some). Let μ_1 and μ_2 be the means of the logistic distribution for player 1 and player 2, respectively, and σ_1 and σ_2 the corresponding standard deviation of the distributions. Specifically, the probability of escalation given a best response to a correct conjecture about the probability of the other player escalating was calculated ($p_3(q_2)$ and $q_2(p_3)$). I then started with $q_2 = 0.5$ and iterated through both functions until I found the fixed point. The values tended to converge, so long as σ_1 and σ_2 were not too small (which reflects how the conflict equilibrium is unstable when there is low uncertainty). I then plotted q_2, $p_3(q_2)$, and $p_1(q_2)$ as a function of R_1 and R_2 and calculated the change in each as a function of small changes in r_1 and r_2. I selected particular realizations of the parameter values, varied R_1 through 40 values between 0 and 2, examined the proportion of those in which $dp_1/dR_1 < 0$. I report these in table 2.4.7. To summarize, the comparative statics in the model with logistically distributed costs for the parameter values examined were identical to those for the uniform case (as they should be), with one exception: for most parameter values, $\frac{dp_1}{dR_1} < 0$.

2.4.8 Summary of Comparative Statics

Table 2.2 summarizes the comparative statics of the four models. The results are most straightforward for the models with private CRR: agents with higher CRR are less likely to escalate before their reputation is engaged ($dp_1/dr_1 < 0$) because the costs of a failed bluff

Table 2.1: **Numerical Examination of $\frac{dp_1}{dR_1}$ for Logistic Distribution**

μ_1	μ_2	σ_1	σ_2	v	k	r_2	Proportion with $\frac{dp_1}{dR_1} < 0$
4	4	4	4	1	4	0	1.0
4	4	4	4	1.5	6	0	0.78
4	4	5	5	1	6	0	1.0
4	4	4	4	1	6	1.0	1.0
4	4	4	5	2	3	0	0.9

Table 2.2: **Comparative Statics for Models F and R**

Model	Round	Agent	Action	Private CRR	Observable CRR
F	1	Focal	Engage reputation	$dp_1/dr_1 < 0$	Depends on uncertainty: Uniform $\implies dp_1/dR_1 > 0$ Logistic $\implies$ (mostly) $dp_1/dR_1 < 0$
F	2	Rival	Use force	$dq_2/dr_1 = 0$	$dq_2/dR_1 < 0$
F	3	Focal	Use force	$dp_3/dr_1 > 0$	$dp_3/dR_1 > 0$
R	1	Rival	Engage reputation	$dp_1/dr_2 = 0$	$dp_1/dR_2 < 0$
R	2	Focal	Use force	$dq_2/dr_2 > 0$	$dq_2/dR_2 > 0$
R	3	Rival	Use force	$dp_3/dr_2 = 0$	$dp_3/dR_2 > 0$

are greater. The increased reputational costs of a concession and the greater reluctance of agents with higher CRR to commit their reputation when costs are high implies that agents with higher CRR whose reputation does become engaged tend to be more willing to escalate further ($dp_3/dr_1 > 0$ and $dq_2/dr_2 > 0$).

The models with observable CRR involve more complicated dynamics because others anticipate agents the behavior of agents with higher CRR and adapt accordingly. As is the case in models with private CRR, when CRR is observable agents with higher CRR are more likely to escalate once their reputation is engaged ($dp_3/dR_1 > 0$ and $dq_2/dR_2 > 0$). In addition, so long as they have not selected into a conflict in which reputation is engaged, players facing an CRR-agent are less likely to escalate the dispute ($dq_2/dR_1 < 0$ and $dp_1/dR_2 < 0$). However, rival-agents who have selected into a conflict in which reputation is engaged will, unlike the models of private CRR, be more likely to escalate against an CRR-agent ($dp_3/dR_2 > 0$). Moreover, the comparative statics of agents with higher CRR' probability of escalation before their reputation is engaged are indeterminate ($dp_1/dR_1 = ?$).

2.5 Empirical Implications of the Model

The models yield the following predictions:

1. **Use of Force:** The clearest prediction that emerges from my models concerns the use of force. Following from the fact that $dp_3/dr_1 > 0$, $dp_3/dR_1 > 0$, $dq_2/dr_2 > 0$ and $dq_2/dR_2 > 0$, I can conclude that for models with either private and observable CRR, for all possible distributions of types that satisfy Conditions C and I, agents with higher CRR will be more likely to use force in a militarized dispute. Furthermore, this result is present for both Models F and R, which is to say that the result holds irrespective of which party is responsible for the action that engages reputation: selection effects do not muddy this prediction.

> **Hypothesis 1**: *Conditional on a MID occurring, leaders with a greater concern for reputation for resolve will be more likely to use force.*

2. **Duration of Disputes:** Let ϕ_2 and ϕ_3 be the length of a MID that ends, respectively, with a concession in the second or third round. Let ϕ_4 be the duration of a MID that involves escalation to the highest level (escalation in Round 3). I assume that $\phi_4 > \phi_3 > \phi_2 > 0$. Then, the average length of a MID is

$$\phi = \phi_2 Pr(\text{conflict ends at Round 2|Round 2 has been reached})$$
$$+\phi_3 Pr(\text{conflict ends at Round 3|Round 2 has been reached})$$
$$+\phi_4 Pr(\text{conflict escalates in Round 3|Round 2 has been reached})$$
$$= \phi_2(1 - q_2) + \phi_3 q_2(1 - p_3) + \phi_4 q_2 p_3$$
$$\phi = q_2 \left(p_3(\phi_4 - \phi_3) + \phi_3 - \phi_2\right) + \phi_2$$

Differences in the reputational concern of a leader will be associated with differences in the average MID length according to the following equations:

$$\frac{d\phi}{dr_i} = \frac{dq_2}{dr_i}(\phi_3 - \phi_2 + p_3(\phi_4 - \phi_3)) + \frac{dp_3}{dr_i}q_2(\phi_4 - \phi_3) \tag{2.14}$$

$$\frac{d\phi}{dR_i} = \frac{dq_2}{dR_i}(\phi_3 - \phi_2 + p_3(\phi_4 - \phi_3)) + \frac{dp_3}{dR_i}q_2(\phi_4 - \phi_3) \tag{2.15}$$

The comparative statics in 2.14 and 2.15 are positive for all type distributions, with both private and observable CRR, under Model R and Model F, with one exception: under Model F with observable CRR, the sign of equation 2.15 could be positive or negative. In Section 2.5.1 I show that for reasonable parameter values these results hold also for Model F with observable CRR.

> **Hypothesis 2**: *The durations of MIDs experienced by leaders with a greater concern for reputation for resolve will tend to be greater.*

3. **Fatalities:** If we assume that expected fatalities are also non-decreasing in escalation, then average fatalities should respond to changes in CRR in the same way that duration responds. That is, let ϕ_2^f and ϕ_3^f be the expected fatalities of a MID that ends, respectively, with a concession in the second or third round, and ϕ_4^f the expected fatalities of a MID that involves escalation to the highest level (escalation in Round 3), and $\phi_4^f > \phi_3^f > \phi_2^f > 0$. Then, using the same argument as above, we can conclude

> **Hypothesis 3**: *The fatalities of MIDs experienced by leaders with a greater concern for reputation for resolve will tend to be greater.*

The argument for believing that expected fatalities should be non-decreasing is straightforward: a conflict that escalates to a higher level is one that involves more rounds of costly signaling and more efforts to impose costs on the other. As such, there is generally more opportunity for fatalities to occur and accumulate, and more risky actions adopted.

4. **Winning Disputes:** Could variation in concern for reputation for resolve account for variation in the probability of winning a MID? Winning can be conceptualized in multiple ways: (1) winning could be the probability that the other player backs-down (coercion), or (2) it could be the probability of successful fighting over and winning the prize (force).[4] Consistent with the general approach of this research, I will consider victories arising from both coercion and force as plausible, and look for comparative statics that are robust under both.[5]

Let w_F^1 and w_R^1 denote the proportion of MIDs which the focal-agent wins through coercion (conceptualization 1) in Model F and Model R, which we interpret as occurring when the rival agent concedes: $w_F^1 = 1 - q_2$; $w_R^1 = q_2(1 - p_3)$.

$$\frac{dw_F^1}{dr_1} = -\frac{dq_2}{dr_1} = 0; \quad \frac{dw_F^1}{dR_1} = -\frac{dq_2}{dR_1} > 0; \quad \frac{dw_R^1}{dr_2} = \frac{dq_2}{dr_2}(1 - p_3) - \frac{dp_3}{dr_2}q_2 > 0; \quad \frac{dw_R^1}{dR_2} \gtreqless 0 \tag{2.16}$$

---

[4]Or could be some weighted combination of (1) and (2)

[5]Note that the coding rules on MIDs do make a distinction that corresponds to these two conceptualizations (Jones, Bremer, and Singer, 1996, p. 180): "A **victory** is defined by the favorable alternation of the status quo by one state through the use of militarized action which imposes defeat upon the opponent... A **yield** is defined by the coerced submission by one state to the demands made by another state but short of any clear alteration of the status quo directly attributable to the threat, display, or use of military force." *Victory* is closer to the first conceptualization of winning, whereas *yield* is closer to the second. However, *victory* seems to also be able to include conceptualization 2.

Let w_F^2 and w_R^2 denote the proportion of MIDs which the focal-agent wins through force (conceptualization 2) in Model F and Model R, which I interpret as occurring when both agents escalate to the highest level and the focal-agent wins the contest: $w_F^2 = q_2 p_3 \pi$; $w_R^2 = q_2 p_3 (1 - \pi)$.[6]

$$\frac{dw_F^2}{dr_1} = \frac{dq_2}{dr_1}(p_3 \pi) + \frac{dp_3}{dr_1}(q_2 \pi) + \frac{d\pi}{dr_1}(q_2 p_3) > 0$$

$$\frac{dw_F^2}{dR_1} = \frac{dq_2}{dR_1}(p_3 \pi) + \frac{dp_3}{dR_1}(q_2 \pi) + \frac{d\pi}{dR_1}(q_2 p_3) \gtreqqless 0$$

$$\frac{dw_R^2}{dr_2} = \frac{dq_2}{dr_2}(p_3(1 - \pi)) + \frac{dp_3}{dr_2}(q_2(1 - \pi)) - \frac{d\pi}{dr_2}(q_2 p_3) > 0$$

$$\frac{dw_R^2}{dR_2} = \frac{dq_2}{dR_2}(p_3(1 - \pi)) + \frac{dp_3}{dR_2}(q_2(1 - \pi)) - \frac{d\pi}{dR_2}(q_2 p_3) > 0$$

The comparative statics for winning are fairly, but not completely, robust. Under models F and R, with private and observable CRR, for all type distributions, agents with higher CRRs will (weakly) win more, with two exceptions ($\frac{dw_R^1}{dR_2}$ and $\frac{dw_F^2}{dR_1}$) in which case the result is ambiguous. In Section 2.5.1 I show that $\frac{dw_R^1}{dR_2} > 0$ for reasonable parameter values. This provides theoretical justification for my next hypothesis, given the caveat that this hypothesis is less robust than the previous hypotheses.

> **Hypothesis 4**: *Leaders with a greater concern for reputation will be more likely to win their MIDs.*

5. **Losing Disputes:** Could variation in concern for reputation for resolve account for variation in the probability of losing a MID? Like winning, losing can be conceptualized as (1) the probability that the focal-agent backs down, or (2) the probability of fighting and losing the contest for the prize. Let l_F^1 and l_R^1 be the proportion of MIDs in which the focal-agent backs down (losing according to conceptualization 1): $l_F^1 = q_2(1 - p_3)$ and $l_R^1 = 1 - q_2$

---

[6]At this point I am going to allow that π may be a function of r_i or R_i. I assume that a greater CRR will make one more likely to win a contest of force: $\frac{d\pi}{dr_1} \geq 0$; $\frac{d\pi}{dR_1} \geq 0$; $\frac{d\pi}{dr_2} \leq 0$; $\frac{d\pi}{dR_2} \leq 0$. However, note that the above analysis of these models did not allow for non-zero effects of CRR on π. It seems intuitive to me that non-zero effects will not alter the comparative statics for private CRR (since it just exacerbates the prior comparative statics), and probably not for observable CRR; the direct effect of non-zero effects of CRR on π is to make the CRR agent more likely to escalate at every stage. In a future analysis of this model I will formally include non-zero effects of CRR on π from the beginning.

$$\frac{dl_F^1}{dr_1} = \frac{dq_2}{dr_1}(1-p_3) - \frac{dp_3}{dr_1}q_2 < 0; \quad \frac{dl_F^1}{dR_1} < 0; \quad \frac{dl_R^1}{dr_2} = -\frac{dq_2}{dr_2} < 0; \quad \frac{dl_R^1}{dR_2} < 0 \quad (2.17)$$

Let l_F^2 and l_R^2 denote the proportion of MIDs which the focal-agent loses (according to conceptualization 2) in Model F and Model R, which I interpret as occurring when both agents escalate to the highest level and the focal-agent loses the contest: $l_F^2 = q_2 p_3 (1-\pi)$ and $l_R^2 = q_2 p_3 \pi$.

$$\frac{dl_F^2}{dr_1} = \frac{dq_2}{dr_1}p_3(1-\pi) + \frac{dp_3}{dr_1}q_2(1-\pi) - \frac{d\pi}{dr_1}q_2 p_3 \gtreqqless 0$$

$$\frac{dl_F^2}{dR_1} = \frac{dq_2}{dR_1}p_3(1-\pi) + \frac{dp_3}{dR_1}q_2(1-\pi) - \frac{d\pi}{dR_1}q_2 p_3 \gtreqqless 0$$

$$\frac{dl_R^2}{dr_2} = \frac{dq_2}{dr_2}p_3\pi + \frac{dp_3}{dr_2}q_2\pi + \frac{d\pi}{dr_2}q_2 p_3 \gtreqqless 0$$

$$\frac{dl_R^2}{dR_2} = \frac{dq_2}{dR_2}p_3\pi + \frac{dp_3}{dR_2}q_2\pi + \frac{d\pi}{dR_2}q_2 p_3 \gtreqqless 0$$

Comparative statics with respect to losing are thus ambiguous. Greater concern for reputation is robustly associated with less backing down (conceptualization 1 of losing); however, it has ambiguous associations with losing a contest of force (conceptualization 2 of losing). The reason for the latter ambiguity is that greater concern for reputation will make the focal agent more likely to escalate (leading to more losses), will have ambiguous effects on the conflict behavior of the other, and will lead to fewer loses in any given contest. Thus, I don't draw any testable hypotheses about losses.

2.5.1 Numerical Examination of Hypotheses

Comparative statics from the family of models on duration, fatalities, and victory are consistent in all of the models, with a few exceptions. The comparative statics on duration and fatalities are consistent for 3 out of 4 of the models, and are ambiguous for the fourth. The comparative statics on winning are consistent for 6 out of 8 of the models, and ambiguous for the last two. This section will examine these ambiguous exceptions to confirm that there exist parameter values where the comparative statics are consistent with my hypotheses, as well as to see if there are parameter values where the comparative statics run contrary to my hypotheses. The formal theoretic study of interstate conflict is not at the point where I could credibly calibrate my model on real world data in order to estimate reasonable values

of parameters. Instead, I select values of the parameters so that Condition I and Condition C are satisfied, and the slopes of the response functions are such that the computational methods I use can discover the solution readily.[7] Absent any guiding principle about how best to search the parameter space, I simply select a number of parameter sets and demonstrate that the prediction could be signed positive or negative.

2.5.2 Duration/Fatalities in Model F with Observable CRR: $\frac{d\phi}{dR_1}$

$$\frac{d\phi}{dR_1} = \frac{dq_2}{dR_1}(\phi_3 - \phi_2 + p_3(\phi_4 - \phi_3)) + \frac{dp_3}{dR_1}q_2(\phi_4 - \phi_3)$$

where $\frac{dq_2}{dR_1} < 0$ and $\frac{dp_3}{dR_2} > 0$. Therefore $\frac{d\phi}{dR_1}$ is more likely to be positive when ϕ_4 is large relative to ϕ_3 and ϕ_2. In table 2.3, I report numerical analyses in which I solve for the equilibrium for particular values of the parameters, and report the results by stating how large ϕ_4 would have to be, relative to ϕ_3 and ϕ_2.

Table 2.3: **Numerical Examination of** $\frac{d\phi}{dR_1}$

μ_1	μ_2	σ_1	σ_2	v	k	R_1	R_2	Condition for $\frac{d\phi}{dR_1} > 0$
4	4	3	3	1	6	0	0	$\phi_4 < 7.6\phi_2 - 6.7\phi_3$
4	4	4	4	1	6	0	0	$\phi_4 > 2.5\phi_3 - 1.5\phi_2$
4	4	5	5	1	6	0	0	$\phi_4 > 1.9\phi_3 - 0.9\phi_2$
4	4	4	4	1.5	6	0	0	$\phi_4 > 3.3\phi_3 - 2.3\phi_2$
4	4	4	4	1.5	1.1	0	0	$\phi_4 > 1.8\phi_3 - 0.8\phi_2$
4	4	4	4	1.5	2	0	0	$\phi_4 > 2.2\phi_3 - 1.2\phi_2$
4	4	4	4	1.5	4	0	0	$\phi_4 > 4\phi_3 - 3\phi_2$
4	4	4	4	1.5	6	0	0.5	$\phi_4 > 0.3\phi_3 - 8.3\phi_2$
4	4	4	4	1.5	6	0	1.5	$\phi_4 > 3.7\phi_3 - 2.7\phi_2$
4	4	4	4	1.5	6	0.5	0	$\phi_4 > 2.4\phi_3 - 1.4\phi_2$

2.5.3 Victory in Model R with Observable CRR: $\frac{dw_R^1}{dR_2}$

Table 2.4 reports the direction of the prediction on victory for agents with higher CRR in Model R with observable CRR for different values of the parameters.

---

[7]For some distributions on costs the equilibrium is unstable in the sense that an iterated mapping through the response functions from a point close to the equilibrium diverges.

Table 2.4: **Numerical Examination of** $\frac{dw_R^1}{dR_2}$

μ_1	μ_2	σ_1	σ_2	v	k	R_1	R_2	$\frac{w_R^1}{dR_2}$
4	4	3	3	1	6	0	0	+
4	4	4	4	1	6	0	0	+
4	4	5	5	1	6	0	0	+
4	4	4	4	1.5	6	0	0	+
4	4	4	4	1.5	1.1	0	0	−
4	4	4	4	1.5	2	0	0	−
4	4	4	4	1.5	4	0	0	+
4	4	4	4	1.5	6	0	0.5	+
4	4	4	4	1.5	6	0	1.5	−

Chapter 3

Honor and War: Southern U.S. Presidents and the Effects of Concern for Reputation (with Devin Caughey)

Abstract[1]

Reputation has long been considered central to international relations, but unobservability, strategic selection, and endogeneity have handicapped quantitative research. We exploit a rare source of variation—the fact that a third of US presidents were raised in the American South, a well-studied example of a "culture of honor"—to identify the effect of heightened concern for reputation for resolve. Using matching, permutation inference, and the nonparametric combination of tests we find strong support for our theory. Disputes under Southern presidents are approximately twice as likely to involve US uses of force, last on average twice as long, and are three times more likely to end in victory for the United States. Other characteristics of Southern presidencies do not seem able to account for this pattern of results.

"Submission[,] whether as regards individuals or nations[,] provokes insult and aggression."

The Yazoo (Mississippi) *Democrat*, on the Mexican–American War, 1848

Shortly after his inauguration, President Thomas Jefferson broke with his predecessors' policy and rejected the Barbary Pirates' humiliating annual demand for tribute. "An insult unpunished," he explained, "is the parent of many others." Citing the many "honorable" rationales "pleading for war," Jefferson sent the United States' first overseas military expedition to the Barbary States to "chastise their insolence" (Herring, 2008, pp. 98–9; Wheelan, 2003, p. 3). Woodrow Wilson's forceful response to German attacks on American shipping leading up to World War I evinces a similar logic: "We covet peace and shall preserve it at any cost but the loss of honour," wrote Wilson. "[If we] accept a single abatement of right...many other humiliations would certainly follow" (Donelan, 2007, p. 106). Fifty years later, on the eve of his massive escalation of US involvement in Vietnam, Lyndon Johnson declared, "Our national honor is at stake in Southeast Asia, and we are going to protect it." Even after the war descended into a bloody stalemate, Johnson still considered withdrawal unimaginable. "If America's commitment is dishonored in South Vietnam," he averred, "it is dishonored in forty other alliances or more....We do what we must" (Wyatt-Brown, 2005, pp. 441–2).

The foregoing anecdotes illustrate an ancient theme: honor as a motive for war (Donelan, 2007; Kagan, 1995; Lebow, 2008; O'Neill, 1999). These statements were not mere rhetoric, nor did they simply express what the president believed to be virtuous for its own sake. Rather, they articulate a compelling instrumental logic: submission to challenges and insults leads others to believe that one is more likely to acquiesce in the future. Allies will lose confidence in the benefits of affiliation; enemies will gain confidence in the benefits of confrontation. The logic of honor is thus closely tied to reputation, as well as to other concepts central to international relations, such as prestige, face, standing, deterrence, and credibility.

Our paper leverages this connection between honor and reputation to shed new light on the role of reputation in international relations. While IR scholars using diverse approaches have greatly contributed to our understanding of reputation,[2] the study of reputation faces several challenges: reputational beliefs are not directly observable, policymakers may strategically misrepresent their beliefs, and selection effects render many observational studies indeterminate (see, e.g. Fearon, 2002).

Reputation is the "dark matter" of international relations: numerous theories rely on it, but clear evidence of its effects remains elusive.

Identifying reputational effects requires a research design that can isolate exogenous variation in a reputation-related factor, such as the content of actors' reputations, their beliefs

---

[2]Most prominently qualitative and historical approaches (Hopf, 1994; Jervis and Snyder, 1991; Kagan, 1995; Lebow, 2008; Mercer, 1996; Morgenthau, 1948; Press, 2005), but also formal theory (O'Neill, 1999; Schelling, 1966; Sechser, 2010; Wolford, 2007), large-n statistical studies (Crescenzi, 2007; Crescenzi, Kathman, and Long, 2007; Huth, 1988; Huth and Russett, 1984; Walter, 2006), experiments (Tingley and Walter, 2011b; Tomz, 2008), and combinations thereof (McGillivray and Smith, 2008b; Sartori, 2005; Walter, 2009).

about reputation, or the importance they place on reputation. As the opening quotations suggest, a defining feature of honor is *intense concern for reputation for resolve*, which is related to reputations for toughness and credibility, for being willing to suffer harm and use violence to defend one's property, one's status, and one's honor itself (O'Neill, 1999, pp. 85–101). Because honor's importance differs dramatically across societies, this cross-cultural variation can be exploited to identify the effects of concern for reputation for resolve. A major challenge to studying cultural effects in international relations, however, is separating the effects of culture from those of other determinants of foreign policy that may be correlated with culture, such as geography, political regime, and economic interests. To overcome this challenge, we exploit within-country variation in national leaders' attachment to honor—and thus their concern for reputation for resolve—induced by differences in their cultural background.

Specifically, we exploit the fact that one-third of all US presidents—including the three featured in the opening paragraph—were socialized in the American South, a well-studied example of a "culture of honor" (Nisbett and Cohen, 1996).[3] We compare US international conflict behavior under Southern and non-Southern presidents, matching presidents to control for time-varying confounders. In addition to holding constant all persistent national characteristics, this design exploits a transparent and well-understood manipulation of the factor of interest, increasing the credibility of our causal inferences (Angrist and Pischke, 2010). We test a set of predictions that were deduced from a family of formal models of interstate conflict escalation, with concern for reputation for resolve being parameterized as a cost to acquiescing once reputation is engaged in the dispute (see chapter 2). Consistent with these predictions, we find that militarized disputes under Southern presidents are twice as likely to involve a US use of force, last on average twice as long, and are three times more likely to end in a favorable outcome for the United States. We introduce a powerful statistical method to political science, the nonparametric combination of dependent partial tests (Pesarin and Salmaso, 2010), to conduct joint permutation tests of our three hypotheses, thus maximizing statistical leverage. In addition to providing some of the clearest evidence to date of the effects of concern for reputation in international relations, these findings also have implications for culture and politics, the role of leaders, and other debates.

---

[3]We considered several other potential cases, including Great Britain, Israel, Italy, and Nigeria. We selected the United States because its subnational culture of honor has been well studied, it has experienced a large number of conflicts, there has been sufficient variation in presidents' cultural background, and the US president's predominance in foreign provides a relatively simple and transparent treatment assignment process, the selection of presidents.

3.1 Honor, Interstate Conflict, and the American South

Honor's essential feature is its reputational character: one does not possess honor unless others *believe* that one does. Further, an honorable man[4] cares intensely about being seen as honorable and demands recognition of his status (Peristiany, 1966, pp. 21–2; Pitt-Rivers, 1968, p. 503; Stewart, 1994, pp. 12–13). One's reputation for resolve is particularly important. Honorable men must be prepared to risk harm and usually to commit violence in order to defend their honor, even against minor insults and challenges (O'Neill, 1999, pp. 91–2; Moritz, 2008, p. 101). Honor also requires fulfilling ("honoring") one's promises and threats; once made, commitments themselves become points of honor that must be defended for honor to be preserved (Pitt-Rivers, 1968, p. 505; O'Neill, 1999, pp. 127–35). Although some systems of honor require additional virtues, nearly all share the characteristics we have highlighted here: intense concern with reputation for resolve, violent defense of honor, and commitments as points of honor.

As Campbell (1964) and many others have noted, honor tends to be most prominent in settings where possessions are easily expropriated, political authority is weak, and reputations are well known. In anarchic environments like rural herding societies or the international state system, a reputation for tenacious defense of oneself and one's property may be the only effective means of deterring predation. Honor thus exhibits a self-reinforcing logic: it is valuable because it makes one's threats credible, and one's threats are credible because of the value of preserving one's honor.[5] In some societies, however, such as the contemporary US South, honor has achieved "functional autonomy," persisting long after the social conditions that gave rise to it have faded away (Vandello, Cohen, and Ransom, 2008).

3.1.1 Honor in International Relations

Honor plays a relatively minor role in most modern societies, but in international relations honor still retains its potency, though often in the guise of related reputational terms like *prestige*, *face*, *standing*, and *credibility* (Donelan, 2007, p. 1; O'Neill, 1999, *xii*). Scholars since Thucydides have cited honor as a cause of interstate conflict (Kagan, 1995, p. 8). As Lebow (2008, Ch. 3) emphasizes, the extent to which societies place intrinsic importance on honor has important consequences for their foreign policies. Even scholars who downplay the intrinsic motivation to build a reputation affirm its instrumental value (Markey, 1999). Morgenthau (1948, p. 95) argues that a state's "prestige—its reputation for power" is crucial to its foreign policy success; Schelling (1966, p. 124) asserts that "face" ("reputation for action") is "one of the few things worth fighting over." Reputation is central to theories

---

[4]We use the word *man* because the specific set of expectations and values associated with the version of honor that we investigate applies predominately to men (O'Neill, 1999, p. 87).

[5]The emergence of honor in such contexts can be understood either as a rational strategy (O'Neill, 1999, pp. 85–192) or as an evolutionarily adaptive trait (McElreath, 2003).

of international (Crescenzi, 2007; Huth, 1997; Huth and Russett, 1984; Jervis and Snyder, 1991; Mercer, 1996; Press, 2005; Sartori, 2005) and domestic (Walter, 2009) deterrence, international cooperation (Axelrod, 1984; Keohane, 1984), international law (Guzman, 2008), dispute settlement (Gelpi, 1997), credit-worthiness (Tomz, 2007b), compellence (Sechser, 2010), and credible communication (Fearon, 1994b; McGillivray and Smith, 2008b; Weeks, 2008).

Academics are not alone in emphasizing such themes: the public rhetoric and private conversations of policy makers are replete with references to honor and related concepts (e.g., Jervis and Snyder, 1991, pp. 20–22). As Kissinger (1979, p. 228) notes, "no serious policy-maker could allow himself to succumb to the fashionable debunking of 'prestige,' or 'honor' or 'credibility.'" President George W. Bush expressed a similar sentiment in his 2004 State of the Union Address: "For diplomacy to be effective, words must be credible, and [after the invasion of Iraq] no one can now doubt the word of America" (Woolley and Peters, 2004; see Lebow, 2008, pp. 473–93 on the role of honor and standing in Bush's decision to invade Iraq).

3.1.2   The Culture of Honor in the Southern United States

Bush's linking of credibility and military force probably had special resonance for listeners from the US South (Lind, 2003, pp. 142–3). Honor has long been one of the ordering principles of white Southern society, with "tremendous importance [for] regulating and determining the conduct of the individual" (Franklin, 1970, p. 34). Honor was a prominent aspect of both major cultural influences on the white South: the Scotch-Irish and the English "Cavaliers" (McWhiney, 1988, pp. 169–70; Ayers, 1989, p. 1483; Fischer, 1989). Southern whites in all subregions and classes displayed a similar devotion to honor: the upper-class duel and lower-class feud were distinguished by "the relative crudeness of the violence" but were "expressions of the same desire" (Franklin, 1970, pp. 33–7; Wyatt-Brown, 1982, pp. 355–6). By contrast, "honor never sank deep roots" in the more orderly, commercial, and egalitarian North (Ayers, 1989, p. 1483; Krause, 2002, pp. 97–131). While honor did have a place in the North through the early 19th-century, non-Southerners tended to conceive of it in terms of Christian virtue and were less liable to resort to violence in its defense (Fischer, 1989, pp. 188, 582–3; Freeman, 2001, p. 168).

As the psychologists Dov Cohen and Richard Nisbett have shown in a series of innovative studies, the culture of honor persists in the present-day South (see Nisbett and Cohen, 1996 for a book-length summary of their work). Using evidence from national surveys, they demonstrate that white Southerners are more likely to endorse violence, but only when honor is implicated (Cohen and Nisbett, 1994). Based on a series of experiments they find that in response to insults Southerners become more upset than non-Southerners, are more likely to believe that the insult "damaged their appearance of strength and masculinity," become more cognitively primed for aggression, and do in fact begin to act in a more physically aggressive and domineering manner (Cohen et al., 1996; Nisbett and Cohen, 1996, pp. 41–55).

Cohen and Nisbett's research, and that of others (e.g., Lee et al., 2007), has convincingly demonstrated that the culture of honor in the South is an enduring phenomenon, leading Southerners to care more about their reputation for resolve than non-Southerners (Shackelford, 2005).

3.1.3 Southern Honor and Foreign Affairs

The connection between honor and foreign affairs is a prominent theme in the historiography on the American South.[6] "Whenever a Southerner fought another," the Southern historian Charles Hope Franklin has observed, "he was, in a very real sense, engaged in war. The honor and dignity at stake were no less important to the individual than they would be to an embattled nation" (1970, p. 36). Bertram Wyatt-Brown, the foremost scholar of Southern honor, argues that the ethic of honor has led Southerners to follow a distinctive "code of conduct" in international affairs, one characterized by an intense concern with the nation's status in the world and "a compulsion for revenge when, in both personal and collective terms, repute for one or another virtue is repudiated" (2005, p. 433). In his comprehensive survey of the South and foreign affairs, Fry (2002) identifies honor as one of the fundamental principles structuring the region's relations with the rest of the world. Honor powerfully shaped Southern attitudes towards the War of 1812 (Risjord, 1961), the acquisition of Florida (Stevenson, 2004), the Mexican–American War (Hospodor, 1999), the Civil War (Cooper, 1983, pp. 180–1; Franklin, 1970), and US imperialism (Fry, 2002; Quirk, 1967), as well as the major conflicts of the 20th and early 21st centuries (Fischer, 1989, p. 843; Wyatt-Brown, 2005).

The Vietnam War illustrates some of the complex ways that honor has shaped Southerners' attitudes towards international conflict. Southern political leaders were initially wary of US involvement in Vietnam, anticipating the costs of defending that commitment. As Georgia's Richard Russell, leader of the Southern bloc in the Senate, put it: "Once you've committed the flag...there's no turning back" (Fite, 1991, p. 359; see also Caro, 2002, pp. 521–5). Despite their initial skepticism, Southern congressional leaders became the firmest opponents of withdrawal once the conflict was under way, insisting that having put America's "honor...at stake," the only question was, "How can we win?" (quoting Russell's Senate protégé, John Stennis of Mississippi; Fry, 2006, pp. 1–2). President Lyndon Johnson shared both his fellow Southerners' initial ambivalence and their belief that leaving Vietnam would lead to dishonor and humiliation. Seeing "the war as a test of his own manliness" as well as of the nation's honor, Johnson chose to escalate the war rather than pursue plau-

---

[6]In addition to the historical scholarship, several other works have also explored this connection. For example, a working paper by Joshi (2008) examines whether the highest US hostility level is greater in militarized disputes that occur under Southern presidents. He, however, examines only MIDs in which the United States is the "target" state, with the rationale that in these MIDs the US has been "insulted," and obtains a null result.

sible opportunities for withdrawal. Despite the war's personal and political toll, Johnson continued to insist, "We will not be defeated. We will not grow tired. We will not withdraw" (Logevall, 1999, p. 393; Herring, 2008, pp. 736–45; see also Berman and Routh, 2003; Wyatt-Brown, 2005).[7]

Similar regional distinctiveness is evident at the mass level as well. Southern survey respondents are more likely to support "militant internationalism," which refers to support for international involvement on the basis of conflictual, but not cooperative, strategies (Wittkopf, 1990, pp. 39–49). They believe that war is more likely to occur and are more likely to tolerate the escalation of limited conflicts into general wars. Southerners generally consider deterrence based on military might to be the only reliable way to prevent an attack and guarantee national security. If war does break out, they tend to believe that "we should fight to 'lick' our opponents" so that others "will know what to expect from us and act accordingly" (Hero, 1965, pp. 81–6, 111–26). In 2008, white Southerners reported more favorable attitudes towards the military than whites elsewhere, and were more likely to believe that the Iraq War was "worth the cost" and effective at deterring terrorists (ANES 2008).

3.2   Testable Implications

As developed above, an essential aspect of honor is that once honor is at stake in a situation, it must be defended, even at great cost. Backing down diminishes one's standing in the eyes of others as well as the loss of the material prize being contested. This suggests that a natural way to model honor is as a utility cost to acquiescing in a conflict in which reputation is engaged. This formal operationalization subsumes a variety of possible reasons leaders from an honor culture are more averse to conceding in disputes: they might place more intrinsic value on their honor, have different beliefs about what their actions signal to others, employ different decision-making heuristics, or have different domestic political incentives. In short, people socialized in a culture of honor, and especially the U.S. South, should behave as if they have a greater *concern for reputation for resolve* (CRR). As chapter 2 formally theorized, greater CRR should be associated with a number of behaviors. Given that the model in chapter 2 provides a reasonable metaphor for how crises escalate, and the key empirical assumption that reputation for resolve is engaged in militarized disputes, we have the following testable hypotheses.

Hypothesis 1: *Conditional on their having become involved in a MID, leaders from a culture of honor will be more likely to use force in that MID.*

---

[7]Logevall (1999, pp. 395–400) argues strongly that had the non-Southerner John F. Kennedy not been assassinated, he would not have escalated the war as Johnson did, in part because unlike Johnson, Kennedy did not view the conflict as a test of personal manliness.

Hypothesis 2: *The durations of MIDs experienced by leaders from a culture of honor will tend to be greater.*[8]

Hypothesis 3: *Leaders from cultures of honor will be more likely to win and less likely to lose their MIDs.*

3.3 Empirical Analysis

We now evaluate the evidence for these predictions in a specific empirical domain: US interstate disputes under Southern and non-Southern presidents. We test these hypotheses using a nonparametric framework, controlling for potential confounders by matching presidents on important covariates. The individual hypotheses are nonparametrically combined into a single joint test of our theory. We demonstrate that the null hypothesis of no difference in conflict behavior between Southern and non-Southern presidents is implausible under a variety of coding schemes, conditioning strategies, sample definitions, and test statistics. Other factors correlated with Southernness are unlikely to account for our results. This strongly supports our proposed causal mechanism: because Southerners care more about honor, reputation for resolve weighs more heavily in US foreign policy when a Southerner is president, leading to the differences in interstate dispute behavior that we find.

3.3.1 Data and Variables

Our dataset consists of militarized interstate disputes (MIDs) involving the United States in the years 1816–2001.[9] We focus on "bilateral" MIDs (those involving the United States and only one other country) because they correspond most closely to the two-player set-up of our formal model, but we also report the results of analyses that include multiparty disputes. A total of 320 MIDs appear in our dataset, of which 192 are bilateral disputes.

Thirty-six presidents experienced at least one MID, thirty-four of whom presided over at least one bilateral MID. We categorized every president as being culturally Southern or non-Southern according to the following rule: a president is labeled "Southern" if and only if (a) he was born and spent his childhood in the South, or (b) he was either born or raised in the South and he spent his pre-presidential political career there. Following Gastil (1971) and Nisbett and Cohen (1996), we define "the South" as the states of the former Confederacy plus Arizona, Kentucky, New Mexico, Oklahoma, and West Virginia. Eleven presidents are coded as "Southern" under this scheme. As demonstrated below, our results are robust to alternative codings of presidents' Southernness.

---

[8]We don't test the predictions about fatalities because there was insufficient power in the MID sample for U.S. presidents on this dependent variable.

[9]Derived from the MID3 dataset using the EUGene computer program (Bennett and Stam, 2010; Ghosn, Palmer, and Bennett, 2004). Fishing disputes are excluded from our dataset (see Weeks and Cohen, 2009).

MIDs are assigned to the president under which US involvement in the MID began. Three response variables were constructed to test our hypotheses. *US Use of Force* is coded as 0 if the highest hostility level reached by the US was "none," "threat of force," or "show of force" and 1 if the US reached "use of force" or "war." *Duration* is defined as the number of days the dispute lasted, truncated at the last day of the original president's term. *Outcome* is an ordinal variable with three levels, "US loss" (-1), "draw" (0),[10] and "US win" $(+1)$, where the first and third categories are disputes that ended in a victory or yield by one party.

3.3.2 Descriptive Statistics and Nonparametric Approach

A key feature of our dataset is that the response variables are measured at the level of the MID, but the causal variable of interest varies only at the level of the president. As a result of the clustering in the data, the effective n in our study is much closer to the number of presidents than to the number of MIDs. Nearly all standard regression techniques for analyzing clustered data lean heavily on parametric assumptions and/or behave poorly in small samples (Angrist and Pischke, 2009, Ch. 8). For these reasons, we instead employ a nonparametric approach that avoids the stronger assumptions required by regression and is better suited to analyzing a small number of clusters with an unknown structure of intra-cluster dependence. This approach treats each president as a fixed cluster of MIDs, making no assumptions regarding the covariance of MIDs within presidencies (Small, Ten Have, and Rosenbaum, 2008). The statistical tests we use are permutation-based: the labels "Southern" and "non-Southern" are randomly shuffled across presidents and a p-value is calculated based on the proportion of permutations with a value of the test statistic as or more supportive of the alternative hypothesis than the one observed. The clustering in the data is respected by permuting at the level of the president rather than the MID. In each permutation, we calculated the value of the test statistic in the pooled set of MIDs in each treatment group.

Our primary interest is not the individual response variables. Rather, in the spirit of R. A. Fisher's famous advice to "make your theories elaborate," we wish to evaluate the overall support the data provide for our theory's multiple empirical implications (Cochran and Chambers, 1965, p. 252). To do so, we exploit a technique new to political science, the nonparametric combination (NPC) of dependent partial tests (Pesarin and Salmaso, 2010; see Good, 2005, pp. 170–71 for a concise overview). In essence, the NPC method calculates the probability of observing *all* the observed differences if the treatment actually had no effects, taking into account the empirical dependence across tests under the null hypothesis. This is done by combining the p-values of the partial tests into a single statistic (e.g., Fisher's omnibus statistic) whose observed value is compared to its permutation distribution.

In contrast to multiple comparison methods, the focus of NPC is global inference rather than the individual tests. NPC is particularly valuable in cases like this, where theory

---

[10] Disputes not resolved by the end of the original president's term are also coded as draws.

yields several empirical predictions but a test of any one hypothesis is handicapped by low power. In addition, unlike parametric multivariate techniques, NPC "frees the researcher from the need to model the dependence relations among responses" (Pesarin and Salmaso, 2010, p. 119). The test statistics need not be independent—indeed, in our case they are almost certainly highly dependent—nor are they assumed to follow a particular parametric distribution (Good, 2005, p. 171). To highlight the overall degree of support for our theory, the analyses that follow highlight the NPC p-values rather than the results for the individual response variables.

Figure 3.3.2 contrasts the characteristics of bilateral MIDs under Southern and non-Southern presidents. MIDs under Southerners last almost twice as long, and the United States is nearly twice as likely to use force and nearly three times as likely to win the dispute (also, four bilateral MIDs ended in a US defeat, one of which occurred under a Southerner). This pattern of results is highly supportive of our theoretical predictions; the cluster-wise joint NPC p-value of the differences is 0.030. Although Southerners and non-Southerners behave very differently once a dispute has become militarized, they become involved in MIDs at the same rate, about 1.6 disputes per year in office. This too is consistent with our formal models, which do not predict systematic differences in pre-militarization escalation.

3.3.3   Statistical Tests and Causal Inference

The raw differences between MIDs under Southern and non-Southern presidents are certainly striking, but can they support the inference that the differences are *caused* by a Southerner being president? That is, would the interstate conflict behavior of the United States have been any different had, all else equal, a viable non-Southern presidential candidate won the presidency, rather than a Southerner? We believe such an inference is plausible, for the following reasons.

First, the processes by which US presidents are selected into office—national elections and vice-presidential succession—are not strongly related to the potential outcomes of interest. In fact, three of the Southerners in our dataset became president upon the unexpected death of their non-Southern predecessor. Moreover, as a general rule presidential "elections are not decided on foreign policy issues" (Larson, 1985, p. 317; see also Almond, 1950). Even if foreign affairs do matter to voters (see Aldrich, Sullivan, and Borgida, 1989), the international situation on Election Day is often a poor guide of what will unfold over a president's term. On the eve of his inauguration, for example, Woodrow Wilson famously commented, "It would be the irony of fate if my administration had to deal chiefly with foreign affairs" (Notter, 1965, p. 217). A similarly ironic lack of foresight of what would become a defining issue of a presidency was evident in the 2000 election, in which foreign policy played little role aside from George W. Bush's criticism of military interventionism and "nation-building" (Herring, 2008, p. 938). It is therefore unlikely that Southerners are systematically selected
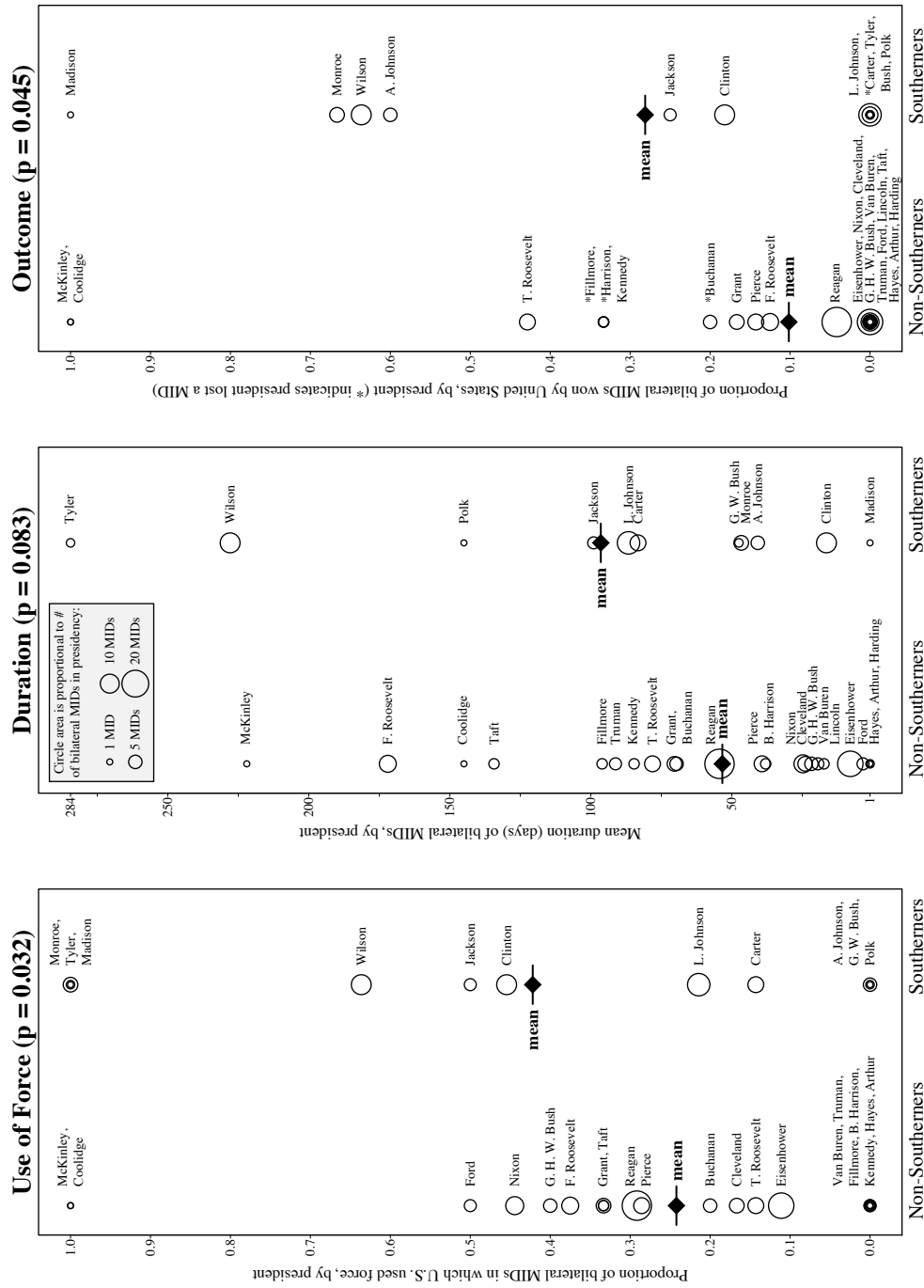
Figure 3.1: **Conflict behavior under Southern and non-Southern presidents.**
US Use of Force, Duration, and *Outcome* in bilateral MIDs under Southern ($n = 64$) and non-Southern ($n = 128$) presidents. President-level means are plotted, with circle size proportional to their number of bilateral MIDs. Diamonds indicate the cluster-wise permutation p-values of the difference of means of each variable (including losses in the case of *Outcome*). The NPC joint p-value of the marginal tests is 0.030.

into office at times when the United States is particularly prone to conflict or its foreign affairs are otherwise unusual. However, to rule out the possibility of chance confounding, we control for the international situation when presidents take office (see the "Structural" set of control variables below) and find the same pattern of results.

3.3.4 Details of the Statistical Analyses

To comprehensively evaluate the robustness of these findings, we varied the statistical analyses in four key respects: (a) the set of control variables used to match presidents (Sekhon, 2009); (b) the presidents coded as Southern; (c) the sample of MIDs examined; and (d) the test statistics used in the individual permutation tests. The R function `GenMatch` was used to search for the optimal set of matches for each combination of coding scheme, variable set, and MID sample (see Table 3.3.5).[11] As with the unmatched data, the NPC method was used to combine the results of the partial tests, the sole difference being that permutations were made only within matched pairs (Pesarin and Salmaso, 2010, p. 38).

A survey of the literature on US uses of force abroad yielded nearly twenty potential confounders.[12] In general, matching on all available control variables is not necessarily optimal. Each additional control may reduce balance on other covariates, possibly biasing the estimate. Bias can also be exacerbated by controlling for variables that are affected by treatment, are only weakly related to the outcome, or are "colliders" on a backdoor path (Rosenbaum, 2002, p. 78; Pearl, 2010; Morgan and Winship, 2007, pp. 64–73). In light of these considerations, we do not base our conclusions on a single "correct" matched dataset, but rather demonstrate the robustness of the results to controlling for different sets of confounders.

Our baseline set of control variables includes only pre-treatment covariates and is designed to capture the structural conditions in place when the relevant president entered office. Controlling for these "Structural" covariates addresses the concern that Southerners happened to hold office during periods of heightened international conflict. Four of the structural covariates are temporal controls: the year the president assumed office (*Year President's Term Began*), plus indicators for the power status of the United States (*Great Power* post-1896, *Superpower* post-1945) and the *Cold War* (1945–91). We also include six measures of the United States' level of international engagement and "war-weariness": *Log*

---

[11]We used one-to-one ATT matching with replacement: each Southern president was matched to one otherwise similar non-Southerner, and a given non-Southerner was allowed to match multiple times. Presidents were given equal weight for the purposes of calculating covariate balance (that is, they were not weighted by the number of MIDs they experienced in office). The "optimal" set of matches was the set that maximized the minimum balance test *p*-value across all variables used to match. See Sekhon (2011) for details on `GenMatch`.

[12]We relied especially on Fordham (1998), Gelpi and Feaver (2002), Howell and Pevehouse (2005), James and Oneal (1991), Meernik (1994), Ostrom and Job (1986), and Park (2010).

Years Since Last War, *Log Deaths Per Capita in Last War*, *Number of MIDs in Previous 10 Years*, and *Uses-of-Force per MID in Previous 10 Years*, plus indicators for *MID Ongoing When President Entered Office* and *War Ongoing When President Entered Office*. We also controlled for the percentage of the US political elite who are veterans (*Elite Veteran*; see Gelpi and Feaver, 2002) and for the lag of treatment, *Previous President Southern*.

We created four additional matched datasets, each successive one of which includes all variables used to create the preceding dataset plus at least one more. Variables were added to the set of controls in reverse order of their likelihood of being affected by the president's attachment to honor. The second set of matching variables, labeled "Party," includes all structural covariates plus indicators for whether the president is a *Whig*, *Republican*, or *Democrat/Democratic-Republican*. This set is designed to control for durable differences in the parties' approach to foreign policy. Next, the "Length" control set adds a control for the number of days the president served in office. The "In-Term" set includes the proportion of the president's term spent in an economic recession (*Proportion in Recession*) and the proportion with unified party government (*Proportion Unified*). The state of the economy and the partisan composition of Congress have both been identified as important influences on US uses of force abroad (e.g., Howell and Pevehouse, 2005), but both are also potentially affected by US conflict behavior. Finally, the "Complete" variable set includes a dummy for whether the president had ever served in the military, which may affect presidents' personal attitudes towards the use of force (Gelpi and Feaver, 2002) but is also likely to be a product of presidents' cultural attitudes towards honor (Cohen and Leung, 2010).

In addition to varying the set of controls used to match presidents, we also checked the results' sensitivity to the coding scheme used to classify presidents as Southern and non-Southern. We tried alternative definitions of the South (e.g., including Missouri, as in Cohen et al., 1996) and also considered the ethnocultural ancestry of each president (see Fischer, 1989, pp. 834–9). Only two presidents' regional classifications were sensitive to these considerations: George W. Bush (Southern under our original coding scheme) and Harry Truman (non-Southern; see SI for further details). We therefore replicated all the analyses using the four possible regional classifications of these culturally ambiguous presidents.

The third aspect of our analyses that we vary is the sample of MIDs. As noted above, we focus on bilateral MIDs because they most closely resemble our two-agent model of interstate conflict, and therefore our predictions are probably most applicable to this class of disputes. It is possible, however, that restricting our sample based on characteristics of the disputes themselves might induce bias if those characteristics are a consequence of treatment. In addition, restricting our sample to bilateral MIDs requires that we drop two presidents, John Quincy Adams and Herbert Hoover, who experienced only multiparty MIDs. By expanding the pool of potential non-Southern presidents to match with Southerners, the inclusion of multiparty disputes modestly improves covariate balance. It also increases the number of MIDs we examine from 192 to 320, potentially making our inferences more reliable (though the effective sample size remains constrained by the number of presidents).

Finally, we examine the robustness of our results to different choices of test statistics

for the univariate permutation tests, all of which are one-sided. The first test statistic, the difference of means between treated and control, can be thought of as the permutation equivalent of Student's t-test without the assumption that the responses are normally distributed (Pesarin and Salmaso, 2010, pp. 25–6). The mean has the advantage of familiarity and simplicity, but other test statistics are superior in the face of outliers, heteroskedasticity, and other departures from a model of fixed additive effects. Tests based on the ranks of the responses rather than the values themselves are robust to such departures, though at the cost of a possible loss of power that can be substantial in small samples (Good, 2005, p. 47).

The most common rank-based test statistic is the Wilcoxon rank sum statistic, which can be used for continuous and, with an appropriate correction for ties, ordinal data. The rank sum tests the null hypothesis against the alternative that responses tend to be larger (or smaller) under treatment (Lehmann, 1975, pp. 1–31; Agresti, 2010, pp. 199–201). This test statistic is used for the response variables *US Use of Force* and *Outcome*. The logrank statistic, a censoring-adjusted variant of the rank sum appropriate for survival analysis, is used for *Duration* (Harrington, 2005). Checking for robustness using ranks is especially important for *Duration* because it is highly skewed and contains several high-leverage observations.

3.3.5   Results

Table 3.3.5 lists the matched pairs of Southern and non-Southern presidents created with each combination of control variables for our original coding of Southernness, among presidents who experienced at least one bilateral MID (see SI for the other matched sets). For each matched dataset, the table lists the covariate most imbalanced between treated and control, as measured by the minimum p-value of Kolmogorov–Smirnov and t-tests. The standardized mean difference between treated and control (the mean difference normalized by the treated standard deviation), a common measure of covariate balance, is given as well.

Figure 3.2 plots the NPC joint p-values corresponding to the 96 possible combinations of the four Southernness coding schemes, six conditioning strategies, two samples of MIDs, and two categories of test statistics (mean-based and rank-based). Each NPC p-value combines the results of three partial tests, one for each hypothesis. Overall, the data provide strong support for our theory. The median NPC p-value in the plot is 0.067, and the mean is 0.078. The results are strongest for disputes that correspond most closely to our two-player model: the 48 p-values based on bilateral MIDs range between 0.015 and 0.153. The unmatched p-values are generally smaller than the matched ones, which is expected given that the unmatched results are based on a sample of presidents 150% larger than the matched sample. Otherwise there is no clear pattern of increasing or decreasing significance depending on the variables controlled for. Given the small sample size of the matched data (20 to 24), the results are remarkably consistent. The reason we obtain statistically significant results with such a small sample is that the differences are substantively quite large, and using NPC to combine the results of the partial tests increases our power against the null hypothesis.

Southern President	Control Variable Set				
	"Structural"	**"Party"**	**"Length"**	**"In–Term"**	**"Complete"**
Madison	Fillmore	Harding	Van Buren	Van Buren	Van Buren
Monroe	Fillmore	Van Buren	Van Buren	Van Buren	Van Buren
Jackson	Buchanan	Buchanan	F. Roosevelt	Buchanan	Buchanan
Tyler	Lincoln	Fillmore	Fillmore	Fillmore	Fillmore
Polk	Van Buren	Van Buren	Van Buren	Van Buren	Van Buren
A. Johnson	Fillmore	Buchanan	Buchanan	Buchanan	Buchanan
Wilson	Taft	F. Roosevelt	T. Roosevelt	T. Roosevelt	T. Roosevelt
L. Johnson	Kennedy	Kennedy	Kennedy	Kennedy	Kennedy
Carter	Ford	Kennedy	Kennedy	Kennedy	Kennedy
Clinton	T. Roosevelt	T. Roosevelt	F. Roosevelt	F. Roosevelt	F. Roosevelt
G. W. Bush	Reagan	Reagan	G. H. W. Bush	G. H. W. Bush	G. H. W. Bush
Worst-balanced Covariate(s) (std. mean diff.)	*Log Deaths per Capita in Last War* (−0.24)	*Republican, Dem/Dem-Rep* (−0.60, +0.45)	*# of MIDs in Last 10 Years* (+0.25)	*Log Years Since Last War* (−0.42)	*Log Years Since Last War* (−0.42)

Table 3.1: **Matched sets of presidents.**

Southern presidents and the non-Southerner matched to them using various sets of control variables. Only the matches based on our original Southernness coding scheme and the sample of bilateral MIDs are shown. The bottom row of the table lists the worst-balanced covariate(s) for each set of matches, as measured by the minimum p-value of Kolmogorov–Smirnov tests and paired t-tests (Sekhon, 2011). The standardized mean difference of the covariate(s) is given in parentheses. See SI for the matched datasets for other codings of Southernness and for complete balance information on all matched datasets created using our original coding scheme.

3.3.6 Ruling Out Alternative Causal Mechanisms

The previous section provides strong evidence that US dispute behavior differs markedly when a Southerner is president. We show the results to be robust to alternative coding choices and to controlling for factors—recent dispute history, era, political party—possibly correlated with both conflict behavior and Southern presidencies. The preceding analyses do not, however, prove definitively that the effect of Southern presidencies is mediated through their greater attachment to honor and concern for reputation. In this section we consider whether other aspects of Southernness could be the causal mechanism at work, ultimately concluding that Southern honor is the only plausible mechanism.

In his comprehensive historical survey of the South and foreign affairs, Fry (2002, pp. 5–6) identifies five factors that have shaped the region's orientation towards foreign policy: attachment to honor and military prowess; commitment to white supremacy; agrarian economic interests; fear of centralized state authority; and loyalty to the Democratic Party. Bensel's *Sectionalism and American Political Development* yields a similar list of alternatives to honor, though with greater emphasis on Southern support for free trade and its hostility to a permanent military establishment until World War II. The South's rural agrarianism, Democratic loyalty, and system of racial apartheid are also emphasized in Chester (1975, pp. 274–85). Hero's book-length treatment of foreign-policy attitudes in the South highlights the region's "cautious, conservative view of international relations," its pessimism about the prospects for international harmony, and its esteem for military virtues (1965, 81 and passim). More generally, the scholarship on the South emphasizes that through the mid-20th century, the region was poorer, more rural, more hierarchical, and less egalitarian than the rest of the nation, with a much larger black population and an ethnically more homogenous white population (Bartley, 1995, pp. 1–2; see also Mickey, forthcoming; Wright, 1986).

For one of the regional differences listed above to offer a plausible alternative explanation to honor, it would have to satisfy three conditions. First, Southern and non-Southern presidents themselves would have to plausibly differ with respect to the alternative factor, as they do in their attachment to honor. Second, it must be plausible that the rival factor causally accounts for the pattern of differences we find. Third, the factor cannot be an aspect or consequence of the culture of honor.

The foreign-policy effects of the South's system of racial hierarchy, which persisted a century after the abolition of slavery, are not straightforward.[13] On one hand, Southern elites' desire to insulate their racial system from external interference rendered them hostile to a powerful central government with a standing army (Bensel, 1984, pp. 404–05). Southerners also feared that US imperialism would lead to the incorporation of non-whites into the polity (Fry, 2002, p. 109). On the other hand, perhaps Southern presidents were more likely to underestimate non-white foreign adversaries and thus to enter into conflicts with them. To

---

[13]Nor was racism confined entirely to the South; racial hierarchy was a core assumption of American foreign policy ideology for most of US history; see Hunt (1987, pp. 46–91).

test this possibility, we examined whether Southerners were more likely to get into disputes with non-white opponents (non-European countries other than Canada and Australia). We found that non-whites actually constituted a modestly larger percentage of non-Southerners' opponents. It is thus unlikely that the racism of Southern presidents can explain our results.

Other explanations also fail at least one of the conditions outlined above. The rural and agrarian nature of Southern society may have been one reason why cultures of honor that were imported from elsewhere continued to flourish there (Nisbett and Cohen, 1996). In addition, the South's status as an economically "peripheral" region dependent on export-oriented agriculture caused it to oppose protectionist tariffs and contributed to its resistance to US acquisition of colonies that would compete with Southern staples (Bensel, 1984). Other than by sustaining a culture of honor, then, it is difficult to see how agrarianism would lead to the interstate dispute patterns we observe. The same can be said of Southern fears of centralized state authority and a standing military, the effects of which would seem to run counter to the effects of honor. Loyalty to the Democratic Party is also an unlikely alternative explanation. For most of US history, the Democrats have been less enthusiastic about foreign interventionism than their partisan rivals. In any case, matching presidents to control for their party affiliation yields the same pattern of results. As for wealth and race, there are few differences between Southern and non-Southern presidents on these factors: nearly every president in our time period was a rich white male.

Finally, there is Southerners' oft-noted valorization of the military and martial virtues. Separating Southern militarism from the culture of honor is difficult, as the former could easily be an aspect or consequence of the latter. Nevertheless, militarism does offer a potential alternative mechanism, if not to honor, then to reputation. That is, if Southerners simply like fighting more (or perceive lower costs to doing so), this might cause behavior similar to that predicted by our reputational model. We do match presidents on military service, finding the same results, but this may not fully account for attitudinal differences between presidents. The militarism hypothesis, however, seems contrary to our finding that Southerners and non-Southerners become involved in disputes at the same rate. Nor is it consistent with the qualitative historical evidence that Southerners have often been ambivalent about a standing army and military adventurism before a conflict is actually underway.

In short, other than the culture of honor, none of the major regional differences mentioned in the literature seem able to account for the pattern of results we document. While alternative explanations cannot be ruled out for certain, the evidence strongly suggests that US interstate disputes differ under Southern and non-Southern presidents because of differences in concern for reputation for resolve stemming from the Southern culture of honor.

3.4 Conclusion

Building on the work of qualitative and historical scholars, our formal and statistical approach sheds new light on the role of honor and reputation in international affairs. We find

compelling evidence that US presidents who care more about reputation for resolve, due to their socialization in the South's culture of honor, behave substantially and significantly different in interstate militarized disputes: they are twice as likely to use force, experience disputes twice as long, and are three times more likely to achieve victory. These results are consistent with the predictions of the formal models of concern for reputation for resolve and escalation developed in chapter 2. Further, these results are unlikely to be caused by other factors correlated with Southern presidencies because theories associated with these other factors do not predict the same pattern of behavior.

While interesting in themselves, our findings also have important implications for understanding international politics. First, our results provide a rare source of credible systematic evidence in support of the view that concern for reputation is critically important in international relations. Consistent with our theory, a greater concern for reputation for resolve appears to yield benefits in the form of a greater likelihood of victory, but at the cost of longer and more violent disputes. It does not, however, necessarily cause leaders to become more involved in militarized disputes in the first place. It is important to keep in mind that the large estimated effect sizes are those from comparing Southern and non-Southern US presidents. Historical scholarship suggests that even our "control" group of non-Southern presidents place high value on reputation. As a result the full effect of concern for reputation for resolve is likely much larger than the differences we estimate.

These findings have implications for other literatures as well. They offer new evidence for the political effects of culture and suggest that designs based on within-country comparisons offer a promising alternative to existing approaches in the study of culture and IR (cf. Johnston, 1995). Our work contributes to the study of the importance of individual leaders (e.g., Hermann and Milburn, 1977) and specifically to the extensive but largely qualitative literature on the US presidency (e.g., Greenstein, 2000). It also provides a fresh perspective on the enduring theme of sectionalism in American politics (Bensel, 1984). These insights are made possible by an innovative multimethod approach that closely integrates qualitative evidence, formal theory, research design, and nonparametric statistical methods. In particular, we introduce a powerful method to political science—the non-parametric combination of tests—which allows the testing of elaborate theories under conditions, such as small sample sizes, in which statistical power is otherwise limited.

Figure 3.2: **NPC p-values for (96) joint hypothesis tests.**
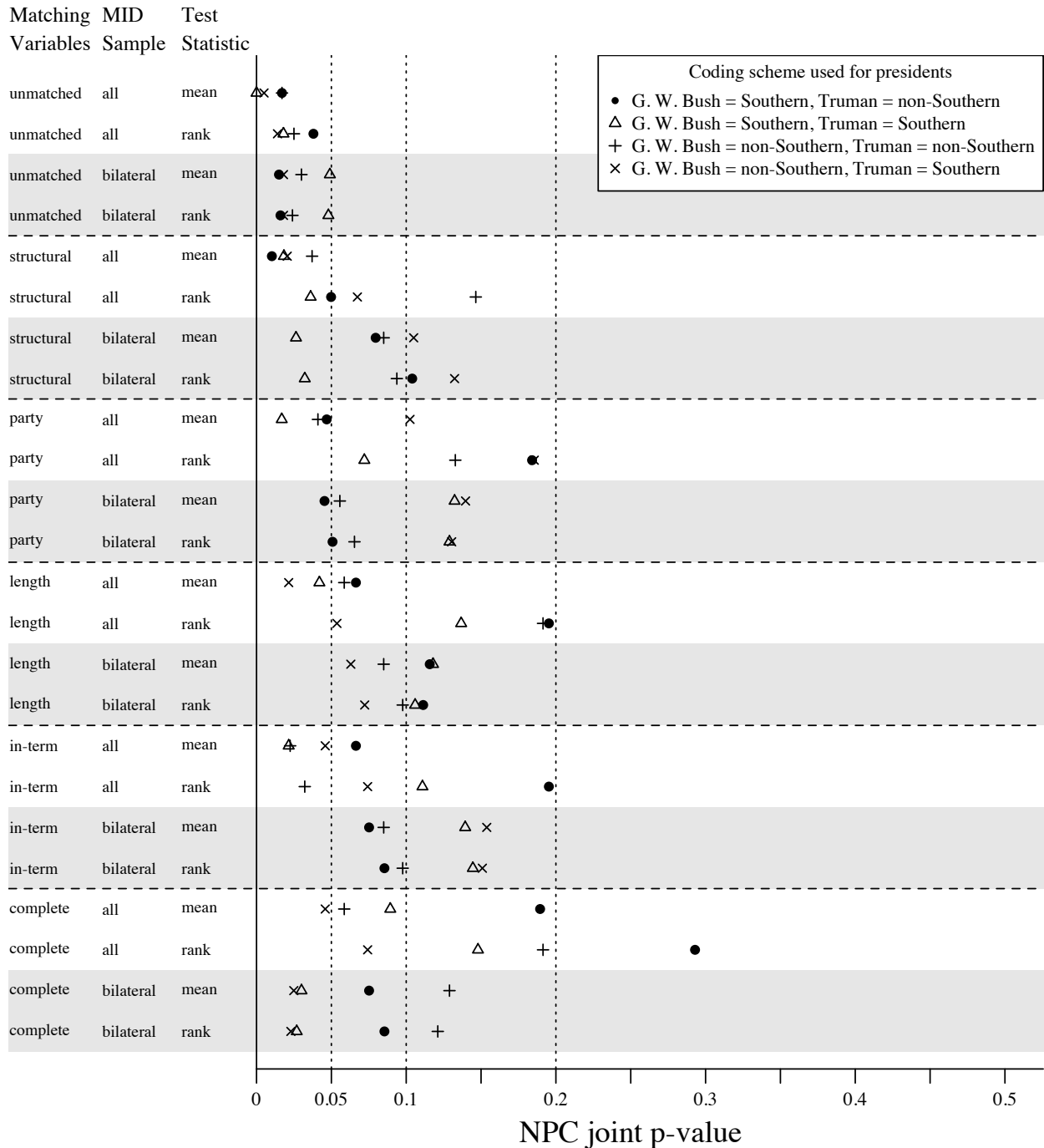
NPC joint p-values based on the 96 possible combinations of coding scheme, matching variables, MID sample, and test statistic. Shading indicates the results for the sample of bilateral MIDs. All p-values are based on cluster-wise permutation tests. The unmatched results are based on a sample of 34 presidents in the case of bilateral MIDs and 36 in the case of all MIDs. The matched samples contain 10, 11, or 12 pairs of presidents, depending on the coding scheme.

Chapter 4

Making a Strong First Impression: Leader Time-in-Office and Concern for Reputation

Fortune, especially when she wishes to increase the reputation of a new prince, who has a greater need to acquire prestige than a hereditary prince does, creates enemies for him and has them take action against him so that he will have the chance to overcome them... .

Machiavelli, Ch. XX, The Prince

Reputation is the dark matter of international relations: it is believed by many scholars to be critically important, but it is very hard to study. Reputation can be defined as the formation of a belief about another agent that can be used to predict that agent's future behavior. Reputations enable the leveraging of the stakes of future interactions to support preferred behavior in the present. From this leveraging of future stakes, cooperative behavior like alliances (Morrow, 2000), debt borrowing and repayment (Tomz, 2007b), and honest diplomacy (Sartori, 2005) becomes possible (Axelrod, 1984), despite the absence of enforceable contracts and the often substantial temptation to renege on promises. Similarly, deterrent behavior is made possible by this harnessing of reputation: countries, and even strategic rivals, for the most part understand and respect their respective spheres of influence (George and Smoke, 1974; Huth, 1997; Schelling, 1966). Reputation is hard to study, however, because it is not directly observable, its implications are often obfuscated through selection and measurement biases, it is theoretically complex and culturally specific, and its ubiquity makes it hard to study for want of variation.

A strategy for overcoming this methodological impasse is to search for situations in which some aspect of reputation varies in a manner that is not likely to correlated with other relevant processes (is exogenous). This paper proposes that as a leader's time-in-office increases that leader's incentives to build a reputation decrease. Unlike reputations themselves which

are a function of analyst-unobserved factors and are subject to strategic manipulation, variation in incentives to build a reputation arising from time-in-office emerge more directly from objectively measurable and exogenously determined structural circumstances. As such, a research design focussing on variation in reputation incentives arising from time-in-office overcomes the two most persistent sources of unknown bias in the study of reputation: strategic manipulation of reputation and the presence of analyst-unobserved factors that confound with reputation.

In addition, a focus on time-in-office enables a powerful research design looking at within-leader variation in behavior, on the domain of all countries and time periods for which data is available. Such a design can eliminate *all* sources of cross-sectional and between-leader confounding[1] by focusing on variation within a leader's term using a leader fixed-effects regression or exact matching on the leader.

Leaders' incentives to build a reputation decrease with time-in-office for two reasons:

1. **Hardening Beliefs**: The informativeness of actions decreases as one's reputation becomes more firmly established. Observers will have weak prior beliefs about the traits and intentions of a new leader; over time the leader's actions will reveal information which will strengthen observers' beliefs. This will harden the leader's reputation, reducing the effect of any subsequent action on their reputation.

2. **Time Horizon**: The instrumental value of reputation decreases as the set of future opportunities to benefit from reputation decreases. Many democracies have term limits; a leader approaching his term limit will have fewer future opportunities to benefit from any given reputation. More generally, leaders' expected future tenure time tends to decrease with time-in-office.[2]

This chapter focuses on variation in concern for reputation for resolve. Specifically, following chapter 2 I theorize about concern for reputation for resolve as a utility cost to backing down once a reputation-engaging action has occurred. I assume that the first militarized threat, display, or use of force is the most significant reputation-engaging action. From the family of formal models built on these assumptions developed in chapter 2, I deduce that leaders more concerned about reputation should be more likely to:

1. use force in a militarized dispute (MID)

2. have longer MIDs

3. have more fatal MIDs

4. win their MIDs

---

[1]Heterogeneous causal effects are still possible across leaders.

[2]A notable exception would be leaders who initially have an insecure hold on power, but who will solidify their position the longer that they hold on to power.

5. fulfill their security commitments[3]

   I test each of these predictions using a fixed effects regression on four subsets of the data to address potential selection issues and causal heterogeneity. I find consistent and mostly statistically significant support for these five predictions. I also examine whether the patterns observed interact with other factors such as age, era, and regime-type, the results of which provide greater insight into the causes of these patterns. I consider alternative potential accounts for these findings. None of the most plausible or widely discussed alternative accounts predict the pattern of outcomes observed with these five dependent variables (see Table 4.1).

---

[3]This last was not derived from the formal models.

4.1 Theory

This section will develop my argument that leaders have incentives to invest more in their reputation early in their time-in-office and that this gives rise to a number of testable implications. Section 4.1.1 introduces the formal theory developed in chapter 2 to derive predictions about the effects of a leader's concern for reputation. Section 4.1.2 introduces the notion that reputations may be leader specific. Section 4.1.3 argues that a leader's incentives to build and maintain a reputation should diminish with time-in-office. Finally, section 4.1.4 discusses alternative causal processes thought to be associated with time-in-office.

A reputation is an inference about an actor, based on their past actions, that has implications for predictions about their future behavior. Reputational inferences help to resolve some of the uncertainty about the target actor's preferences, the situation facing the actor, or the beliefs held by the actor. Reputations may be inferences about a relatively-fixed characteristic of an agent—about their "type", such as a leadership being tolerant of fatalities—and it is often convenient to think about and model reputation in this way. However, reputations can also be inferences about expectations, beliefs, and strategies, such as a reputation for retaliation, that inform predictions about future behavior.[4]

In many studies a prominent strategic function of reputations is that they enable actors *to commit to courses of actions that they otherwise might not be able to.* States that have reputations for resolve, fulfilling their threats, or retaliation are better able to credibly deter and coerce others (Schelling, 1966; Morgenthau, 1948, p. 95; Crescenzi, 2007; Huth, 1997; Kagan, 1995; Sartori, 2005; Sechser, 2010; Walter, 2009). Similarly, states that have a reputation for being cooperative and fulfilling their promises are better able to achieve mutually beneficial arrangements (Axelrod, 1984; Greif, 1989; Guzman, 2008; Keohane, 1984; McGillivray and Smith, 2008b).

There are (infinitely) many kinds of possible reputations, though in practice most reputational inferences tend to follow a simple, all be it culturally specific, logic. Some kinds of reputation are more fundamental, such as reputations for cost tolerance, for prudence, for military effectiveness, for honesty, and for retaliation against affronts. Others are more abstract and general, such as a reputation for resolve or for fulfilling commitments—which combines some of the previous reputations. Even in a world with many kinds of fundamental reputations, it would often be practical to theorize in terms of more general composite reputations. This is because any given action, such as a concession in a dispute, typically could be caused by a number of factors, and thus shapes a number of reputational inferences. Furthermore, since most actions that one undertakes that improves one's reputations reveals

---

[4]The formal literature reflects this distinction in its two modeling approaches: the first (e.g. Fudenberg and Kreps, 1987; Kreps and Wilson, 1982; Kreps et al., 1982; Milgrom and Roberts, 1982) regards reputation as an inference about an actor's *type*; the second (e.g. Axelrod, 1984; Greif, 1989; Sartori, 2002) regards reputation as an inference about the equilibrium of play, typically in the context of an (effectively) infinitely repeated game (for reputation as equilibria in finite games, see Neyman, 1999; Radner, 1980; Selten, 1978). While these approaches do have similarities, they also have important subtle differences (see Dafoe, 2009).

information about one's intertemporal preferences, it has implications for all of one's reputations: an actor who stands firm in a dispute may reveal high valuation for the future, and thus is also more likely to be honest and fulfill his other commitments. This tendency for reputations to "comove" is apparent in the "unity of honor" (O'Neill, 1999, p. 90), which is the perception in many cultures that the various traits associated with honor (e.g. honesty, bravery, and willingness to use violence to defend status) are regarded as either all present or all absent in an individual.

4.1.1 Reputation Engagement and Escalation

This chapter focuses on a fundamental question that is insufficiently answered in the extant literature: how should conflict behavior change for a leader that is more concerned about his reputation? The literature on deterrence and reputation has predominantly either assumed an answer to this question or resisted drawing predictions due to problems of selection effects (Danilovic, 2001; Fearon, 1994b; Fearon, 2002; Lebow and Stein, 1990; two examples of scholars analyzing variation in concern for reputation are Sechser, 2007; Walter, 2009). In chapter 2 I propose a theoretical structure for resolving this question. I build a family of formal models of conflict escalation and reputation engagement that allows the identification of observable implications of concern for reputation that are robust to selection and different assumptions about the strategic environment.

I conceptualize *concern for reputation for resolve* in a reduced-form way as a utility cost to conceding in disputes in which reputation has become engaged. My conceptualization of reputation for resolve abstracts from a number of others found in the literature. A reputation for honesty, for commitment, or for cost tolerance, for example, could each be regarded as a reputation in which the set of *reputation-engaging events* consist of, respectively, statements of intent to stand firm, the initiation of a public contest, and the initiation of situations in which costs are born. To focus my analysis on the strategic dynamics surrounding reputation-engagement, my models adopt a simple two-step reputation-engagement function: I assume that reputation becomes fully engaged in a dispute following any reputation-engaging event. Specifically, as in chapter 2 I make the following assumption to link the model to the data

Assumption 1. *The militarization of a dispute (i.e., the first threat, show, or use of military force) indicates that reputation has become engaged.*

The theory summarized in chapter 2 then allows me to deduce[5] the following four testable predictions for leaders earlier in their time-in-office.

Prediction 1. *Conditional on a MID occurring, they will be more likely to use force.*

Prediction 2. *The average duration of a MID will be longer.*

---

[5]Given some additional assumptions about the mapping between the game and the empirical domain, and a minor qualification, that were omitted due to space constraints.

Prediction 3. *They will experience more fatalities per MID.*

Prediction 4. *They will be more likely to win their MIDs.*

In addition, all else equal, leaders who are more concerned with their reputations are more likely to maintain their security commitments and are more likely to have their commitments trusted. Treaty commitments are particularly well suited to testing theories of reputation because expectations of behavior and hence reputational inferences are most clear when concerning legally articulated commitments (Simmons, 2010).

Prediction 5. *Leaders who are more concerned about their reputation will be more likely to fulfill the terms of a security commitment.*

4.1.2 Leader-Specific Reputation

Reputations may adhere to any kind of agent, group of agents, or type of organization: a leader, a clan, an army, a people, a type of political system, a country. Scholars and leaders have throughout history referred to all of these. For example, as recounted by Thucydides, speakers made regular reference to the reputations of individuals as well as to peoples.[6] Similarly, Machiavelli places extensive emphasis on the importance of reputation about leaders, armies, and states. Contemporary scholarship has typically theorized about reputation as inferences about a state, though recently there has been a resurgence of interest in reputations about leaders (Goemans, Gleditsch, and Chiozza, 2009; Guisinger and Smith, 2002; McGillivray and Smith, 2005; McGillivray and Smith, 2008a; McGillivray and Smith, 2008b; McGillivray and Stam, 2004; Wolford, 2007). Suffice it to say that reputations, as inferences with behavioral implications, may occur at multiple levels; the decision of a country to abandon a military commitment likely says something about the leader, about the country's military, and about the interests and cohesion of the country's politically influential population. Dafoe and Huth (2011) propose a theoretical generalization of these different perspectives in a theory of *Influence-Specific Reputation* which states that reputations will adhere more to agents with more decision making influence in the relevant policy domain. Thus, in political systems where leaders have more influence on the decision to use force, reputations about resolve will be more leader-specific and less state-specific.

This paper tests behavioral implications of leader-specific reputation. To the extent that reputations are not leader-specific it should attenuate the observed effects of concern for reputation and bias against my finding supportive results. This paper will also test the theory of influence-specific reputation by examining whether my time-in-office predictions are stronger in regimes where leaders have more military authority (see table 4.5).

---

[6]"consider the enormous differences between [the Spartans] and the Athenians... An Athenian is always... But [the Spartan] nature is always..." "We must realize, too, that, both for cities and for individuals, it is from the greatest dangers that the greatest glory is to be won.... We must live up to the standard [set by our fathers]: we must resist our enemies in any and every way... (*History of the Peloponnesian War*, 70, 144, Book 1)

4.1.3 Theory of Diminishing Reputational Incentives with Time-in-Office

The theoretical claim underlying this research design is that reputational incentives diminish with time-in-office. This is likely to be so for two reasons: (1) diminishing time horizons with time-in-office, and (2) hardening of reputation with time-in-office.

Time Horizons

The idea that concern for reputation is greatest at the early stages of a set of interactions is found in the first formal models about reputation. In the finitely repeated chain-store game or finitely repeated prisoner's dilemma with uncertainty about types, reputational behavior is most likely early in the game when the benefits to mimicking other types has the most rounds to pay itself off. As the end point of the game approaches, agents become less likely to sustain reputational behavior because they have more incentive to "burn their reputation" for short-term benefits (Axelrod, 1984; Fudenberg and Kreps, 1987; Kreps and Wilson, 1982; Kreps et al., 1982; Milgrom and Roberts, 1982). Similarly, for infinitely repeated games the "Folk Theorem"[7] states that a multiplicity of equilibria may become possible for sufficiently patient actors; an actor's patience, or discount factor, can be interpreted as an increasing function of the probability that the agent believes she will be around in the future. Leaders who are more likely to lose office or die will put less value on the future benefits of having a good leader-specific reputation. In short, longer time horizons provide incentives to build and maintain beneficial reputations.

Hardening of Reputation

A second theoretical argument supports the notion that greater time-in-office will be associated with diminished incentives to maintain reputations. As an agent accumulates a longer history of informative actions, others' inferences about the agent will tend to increase in confidence.[8] As observers' have more confident beliefs about an agent, they will tend to infer less from any subsequent action, thereby reducing the reputational stakes for the agent.[9]

---

[7]So called because the arguments associated with it emerged from the folk wisdom of early game theorists

[8]This is true in expectation, though it may not be true in particular cases for certain distributions of uncertainty. However, for some reasonable distributions "such as the normal and binomial, [additional information can only decrease] the posterior variance". See Gelman et al., 2003, Ch. 2 and http://andrewgelman.com/2011/07/adding_more_inf/

[9]This mechanism depends on the nature of reputational inferences. In games such as the finitely repeated chain-store game or prisoner's dilemma, where there are some (rare) types that others desire to mimic ("Stackelberg types"), then a single deviation from the rare type's behavior will reveal oneself to not be of this type. Thus, irrespective of how long the game has progressed, for some branches of play any subsequent action can be as informative as early actions. However, it is worth noting that for other branches of play—namely those in which the player has revealed itself to not be the Stackelberg type—then subsequent actions will have no influence on others reputational inferences, consistent with the "hardening of reputation" logic.

A complication arises from reputational decay. Over time, unobserved changes in preferences, political coalitions, or other factors should increase others' uncertainty, leading to the decay of reputations. The higher is this rate of reputational decay, the weaker should be the *hardening of reputation* mechanism, since actors will have incentives to reinforce their reputations more frequently. On the other hand, if reputations decay slowly—perhaps because the reputation is about a highly persistent trait—then the *hardening of reputation* mechanism will be more potent.[10]

An additional implication of the reputation hardening mechanism is that the injection of uncertainty into the system should be associated with an increase in reputational incentives. For example, if there is a change in the international balance of power, or a change in a domestic ruling coalition, there will be uncertainty about the strategies and resolve of various parties; these parties will have incentives to rebuild a beneficial reputation. We might test this implication by looking for a jump in the above five predictions following uncertainty-increasing events such as a coup or domestic revolution, a sudden change in alliance structures, etc... Unfortunately, most of these tests will be highly confounded.

A second additional implication of the reputation hardening mechanism is that the effects will be stronger for young leaders and leaders new to political office, than for older or experienced leaders, since the latter will have more established reputations. On the other hand, the time horizon mechanism predicts that changes in reputational incentives associated with time-in-office will be most severe for older leaders, since the future opportunities for them to benefit from a given reputation are more limited by their lifespan. These two last conflicting implications are examined in table 4.4.

Low Concern for Leader-Specific Reputation Implies Attenuation Bias

There are a number of potential concerns that one might have with this account. First, it may be that most reputations are largely not leader-specific, but are primarily specific to other entities such as states. Second, leaders may care about the welfare of others (or be forced to behave in the interests of others), such as their family, clan, political party, or country, and as such they may seek to build reputations that survive their departure from office. Third, leaders may be intrinsically motivated to cultivate their reputation, such as through a desire to build a good legacy, or through deep seated beliefs about the importance of honor.

While these possibilities could pose a problem to this research design, on the whole they should only create *attenuation bias* which means that any estimated association will be biased towards zero and that statistical tests will be less likely to reject a null of no effect. However, contingent on finding a relationship, the presence of attenuation bias implies that the true

---

[10]An additional potential implication of the reputational hardening mechanism is that low decay rates may be associated with a time-in-office ceiling, above which reputational incentives remain constant over time. As the decay rate increases, this ceiling will reduce. However, if the decay rate increases too much, it will at some point be no longer optimal to build a reputation at all.

effect of time-in-office induced reputational incentives is greater and more statistically significant than estimated. That is, if there are time-in-office patterns of behavior that can be persuasively attributed to diminishing reputational incentives, despite the attenuating bias of the limited role of leader-specific reputations, the altruism or organizational influences affecting leaders' behavior, and leaders' intrinsic motivations to build reputation, then the actual effect of concern for reputation is likely be much larger than what is estimated.

It is also worth clarifying that the *hardening of reputation* mechanism is consistent with a concern for reputation that is intrinsically or "psychologically" driven. So long as what one cares about is one's actual reputation (beliefs of others), rather than simply behaving in a certain manner (honorably, virtuously, etc...) then the hardening of reputation will still diminish the concern for reputation of a leader who is intrinsically motivated—as opposed to instrumentally motivated—to maintain a good reputation.

4.1.4 Empirical Findings Related to Time-in-Office

A number of works have looked at how conflict behavior varies with time-in-office and related variables such as the age of a leader and time until election. I will review them here, largely deferring discussion of relevant theory to section 4.3.

Bak and Palmer (2010) look at the targeting of MIDs in a directed-leader-dyad-period setup with respect to time-in-office[11] and age. Bak and Palmer find the effect of time-in-office depends on age; young leaders (less than about 50) have a positive hazard (positive association between time-in-office and being the target of a MID), whereas old leaders (greater than about 70) have a negative association. In addition, "old leaders are more likely to be a target of [MIDs]". Bak and Palmer interpret this as evidence against the "Biden hypothesis" that young, inexperienced leaders (like Obama) are more likely to be the target of MIDs.

Enterline and Williams (2010) similarly examine the "Biden Prophecy". They look at the time until first foreign policy crises for American presidents as a function of age and political experience. Consistent with the "Biden Prophecy", Enterline and Williams find that younger presidents experience their first crises earlier. Enterline and Williams examine a different dependent variable (time to first crisis) than Bak and Palmer, and they employ different control variables, but it is unclear precisely why they get these divergent results. Enterline and Williams also find contradictory results for the association with experience: "congressional experience corresponds to earlier, and state-level experience later, foreign policy crises".

Potter (2007) examines time-in-office as a proxy for the experience of the leader, with

---

[11]Time-in-office is operationalized in a special manner as the time since the leader took office to (i) the day the leader left office in a given period, (ii) the last day of the period of the leader stays in office throughout, or (iii) the day when a MID begins. Unfortunately, this third component will induce tautological endogeneity, since in any observation in which a MID occurs the measured time-in-office will by definition be earlier than the counterfactual. This will generate a negative bias between time-in-office and the dependent variable. A better method would be to just measure time-in-office until the start of the observation.

inexperienced leaders in Potter's formulation more likely to experience conflicts. Potter examines how the probability of the US and UK experiencing a MID or international crisis changes with the leader's time-in-office, controlling for time to election, the age of the leader, and in the US case, whether the president was formerly a vice president. Potter finds that time-in-office is associated with fewer MIDs and fewer crises, age has either zero or a positive association with crises and MIDs, time to election no strong associations, and the vice presidency an inconsistent association (negative significant for MIDs, positive insignificant for crises).

Gelpi and Grieco (2001) argue that "potential challengers may be more likely to target inexperienced leaders". Since democracies have more leadership turnover, they are more likely to have inexperienced leaders who attract challenges from other states. This argument, if correct, could account for the the puzzle of why (often powerful) democracies are more likely to be targeted in militarized disputes than autocracies. Gelpi and Grieco analyze dyad-years and find that democracies are more likely to be targeted in crises, are less likely to initiate crises, and that time-in-office is associated with fewer crises (as both initiator and target). Time-in-office reduces the estimated partial correlation coefficient on democracy by about 50%, though this difference in not statistically significant.

Wolford (2007) derives a number of predictions from a formal model of leader-specific reputation, one of which is directly relevant to time-in-office. Wolford (2007, p. 779) argues that "when conflict is sufficiently costly for the antagonist, an increase in the incumbent protagonist's time in office will decrease the probability of conflict". This prediction is motivated by two arguments. The first, like my theory, draws on the effect of time horizons to alter reputational incentives: "new incumbents have an incentive to demonstrate resolve [early in their time-in-office because] they will receive better bargains over time". The second shares the logic related to the hardening of reputation: "antagonists have an incentive to issue probing demands designed to test the incumbent's resolve... As the informational asymmetry diminishes over time, both the incumbent's incentives to bluff and the adversary's incentives to learn will diminish, however, and the probability of war will often decline as a result." To my knowledge, no one has tested the predictions in (Wolford 2007).

Chiozza and Goemans (2003) seek to study how the risk of losing office affects the probability of crisis initiation. To do so, they model using a two-stage probit the "reciprocal relationship between the probability of losing office and the probability of crisis initiation." They find two countervailing results: [m]ore time in office increases the probability of crisis initiation (model 2), and more periods in office decrease the probability of crisis initiation (models 1 and 2)." (459)

Chiozza and Choi (2003) examine whether new leaders will be more likely to peacefully resolve territorial disputes (because they are more able to break from past policies) or whether states with leaders late in their time-in-office will be more likely to peacefully resolve territorial disputes (because they will have greater experience and will be more secure). Chiozza and Choi also examine whether leaders who have "a reputation for toughness and military competence" are more or less likely to peacefully resolve their disputes. Chiozza and

Choi find that time-in-office is associated with an increased tendency by autocracies to make concessions in territorial disputes, but a decreased tendency in democracies. Prior military career, a history of initiating or being a target of territorial disputes, a history of using force, and a history of not losing crises are all associated with more frequent concessions; a history of defeats in crises is associated with less frequent concessions.

Gaubatz (1991) looks at whether there are patterns in conflict behavior associated with the approach of elections. The approach of elections increases the incentives on re-electable incumbents to behave in a manner preferred by large segments of the electorate. If there is a "pacific public", then the approach of elections for a re-electable incumbent should be associated with fewer wars.(217) On the other hand, if the public tends to reelect incumbents who are involved in militarized disputes because of rally effects or belligerent preferences, then the approach of elections for a re-electable incumbent should be associated with more frequent wars. Gaubatz finds that democracies experience more wars early in the election cycle, irrespective of the regime type of the initiator. While the theory articulated in (Gaubatz, 1991) is of much benefit, a major concern is that the analysis relies on histograms of war onset by years until or since the last election, which is confounded with the size of the sample for each bin (a better measure would be proportion of observations experiencing a war onset). There are also no cross-sectional controls and some of the results are non-monotonic.

Horowitz, McDermott, and Stam (2005) examine leaders' age and the initiation and escalation of MIDs. They find that leader age is positively significantly associated with conflict initiation, the use of force, and the onset of war for most types of countries, though these associations are absent or reversed for personalist regimes. Horowitz, McDermott, and Stam interpret these results as inconsistent with a theory emphasizing the role of testosterone as a determinant of conflict behavior, and consistent with the argument that "older leaders, with more experience and more personal credibility with relevant power brokers, may have more freedom of action" and because their shorter time horizons induce them to take more risky actions in order to secure their legacy.

In summary, depending on the statistical analysis,

- time-in-office is associated with fewer MIDs (Potter, 2007) and crises (Gelpi and Grieco, 2001; Potter, 2007)

- when the leader is young, time-in-office is associated with increased targeting of MIDs; when the leader is old, time-in-office is associated with decreased targeting of MIDs (Bak and Palmer, 2010)

- time-in-office increases, but more periods in office decrease, the probability of crisis initiation (Chiozza and Goemans, 2003)

- time-in-office is associated with an increased tendency by autocrats and decreased tendency by democrats to make territorial concessions (Chiozza and Choi, 2003)

In addition, we have some other related findings:

- the approach of elections are associated with fewer wars (Gaubatz, 1991)

- older leaders are more likely to be a target of MIDs (Bak and Palmer, 2010); in non-personalist regimes older leaders are more likely to experience MIDs, uses of force, and wars, whereas in personalist regimes the opposite associations seem to hold (Horowitz, McDermott, and Stam, 2005)

- younger leaders experience their first crisis sooner in their time-in-office (Enterline and Williams, 2010)

As a research program, there are a number problems with the above set of analyses. Most importantly, none of the above analyses are able to persuasively rule out unknown biases (Rosenbaum, 2010) arising from country-specific or leader-specific differences. Some papers do a better job than others, but ultimately the potential for cross-sectional confounds substantially weakens any inferences drawn from the above analyses. Second, the above results do not appear to be consistent with any underlying theory; even within some studies the results seem to be inconsistent. Time-in-office associations seem to reverse themselves depending on the age of the leader, the regime-type of the country, and the unit of time (days in office vs periods in office). Similarly, age associations seem to be contingent on the dependent variable (MIDs or crises), and the regime-type.

Third, the various theories are too vague and flexible to support a clear causal inference, even if there were consistent empirical patterns. Furthermore, many of the mechanisms proposed are sufficiently flexible that they could be consistent with many kinds of observed behavior: experience could make leaders more able to avoid conflicts (Potter 2007) or could provide leaders with the political capital and policy freedom to engage in conflict (Horowitz, McDermott, and Stam 2005); possession of a reputation for resolve may induce leaders to avoid making concessions in territorial disputes so as to maintain their reputation, or the possession of a reputation for resolve could provide leaders the reputational capital amongst voters to enable them to make concessions (Chiozza and Choi 2003); age, being associated with reduced testosterone in male leaders, may make leaders more peaceful or, being associated with limited time in which to establish one's legacy, may make leaders more militarily risk-acceptant and bold (Horowitz, McDermott, and Stam 2005). The flexibility in the observable implications being drawn from very similar factors and processes means that it is hard to rule out any theory with a given result on a particular dependent variable. Further, the above is an indication that the theoretical space has sufficient flexibility that many equally plausible and parsimonious accounts could be post-hoc proposed to account for any particular set of results. Weak theory is present in social science because the subject of study is complex; however, clear causal inferences will not be possible until stronger theory, with more elaborate, confident, and otherwise unlikely, predictions, can be generated to guide a rich empirical analysis.

This study overcomes the above problems. First, the risk of unmeasured cross-sectional

bias is ruled out through the use of fixed effects at the level of the leader.[12] Second, as will be shown, the multiple results emerging from my study are each consistent with a single clearly articulated theory and are not consistent with the most plausible alternative theoretical accounts. Third, the theory motivating these predictions was articulated prior to this study (Dafoe and Caughey, 2010), without reference to this empirical domain, and gives rise to a more elaborate and risky set of predictions.

4.2 Analysis

As described in section 4.1.1, my theory gives rise to five predictions about how time-in-office induced reputational incentives will influence interstate conflict behavior. Leaders early in their time-in-office should:

1. be more likely to use force in a militarized interstate dispute (MID)

2. experience longer MIDs

3. experience more fatal MIDs

4. be more likely to win their MIDs

5. be less likely to break their security commitments

An elaborate set of predictions provides a number of advantages: through the use of the nonparametric combination of tests (Pesarin and Salmaso, 2010) it will increase the power of one's tests against alternative hypotheses; it allows one to reject other theories that agree on some but not all predictions; and it increases the probability of finding evidence against one's theory, and hence increases one's confidence in any theory that is not rejected.[13]

The first four predictions will be tested on the latest version (v3.1) of the Militarized Interstate Disputes dataset (Ghosn, Palmer, and Bennett, 2004). By Assumption 1, all predictions are conditional on a MID occurring, so the unit of observation is the leader-MID. The data on leaders comes from the latest version (v2.9) of the *Archigos* dataset (Goemans, Gleditsch, and Chiozza, 2009). I assign MIDs to the leader who experienced its onset; the rationale behind this coding decision is that leaders have the most influence over the onset of MIDs, but once a MID is ongoing leaders are more constrained in their influence. I drop MIDs that extend beyond the end of the dataset.

The dependent variables for the first four predictions are:

1. whether force was used by the leader's country in this MID

---

[12]In subsequent non-parametric combination analyses I will also do exact matching on the leader.

[13]I contemplated other predictions, such as on ICB crises; however ICB crises are too rare for the fixed-effects research design used by this paper.

2. the log duration of the MID (expressed in the log of days)

3. the total number of fatalities, expressed in the MID dataset's approximately logarithmic five level coding[14]

4. A victory for the leader's country (or a yield by the opposing country)

I examine the fifth prediction—that leaders are more likely to break their security commitments later in their time-in-office—by using the coding provided in (Leeds, Mattes, and Vogel, 2009, p. 469) for whether an alliance ended due to the leader's state violating its alliance commitments (see Leeds et al., 2002 for additional details about this treaty dataset). Leeds et al. code as a violation whenever "(1) a major provision is violated and the governments do not indicate their intention to continue to recognize the alliance in spite of the violation, or (2) one or more of the allied governments specifically declares that it will no longer recognize or be bound by the alliance agreement despite the fact that the alliance has not reached its scheduled termination date" (Leeds and Savun, 2007, p. 1124).

Tables 4.1 and 4.2 report the estimated coefficients, standard errors, and number of observations and leaders for five separate models run for each of the five dependent variables. Model *All* refers to an analysis on all observations that were present in the combined datasets. Model *All, Reg Entry* then subsets on just those leaders whose entry was coded as *regular*, as opposed to an *irregular* manner such as a coup or a *direct* imposition by a foreign power; this specification is meant to eliminate the possible confounding from initially insecure leaders whose early successful conflict behavior extends their time-in-office. Model *All, Reg Entry Exit* then subsets on only those leaders who had *regular* entries and *regular* departures from office (that is, "according to the prevailing rules, provisions, conventions and norms of the country") or had a *natural death*; this subset also reduces the risk of other forms of endogeneity, such as if coups are a function of performance in war. However, conditioning on how the leader left office could also induce bias, say for example if coups are more likely against leaders whose concern for reputation diminishes the most with time-in-office. Model *Medium Tenure* subsets on leaders who had a *regular* entry and whose total time in office was less than 12 years. Model *Long Tenure* subsets on leaders who had a *regular* entry and whose total time in office was greater than 12 years. I subset on total time-in-office to examine whether the result could be driven by particular end-of-tenure behavior of leaders with very long time-in-office (since extreme values of the independent variable will have larger influence on the estimated coefficients).

Tables 4.1 and 4.2 reveal strong overall support for the five predictions of the theory that time-in-office induced reputation incentives affect conflict behavior. For the *All* model (with no subsetting), all five predictions are statistically significant in the predicted direction. The joint probability of getting this result, against the null of no difference in each of the five

---

[14]0 is no deaths, 1 is 1-25 deaths, 2 is 26-100 deaths, 3 is 101-250 deaths, 4 is 251-500 deaths, 5 is 501-999 deaths, and 6 is greater than 999 deaths. See (Jones, Bremer, and Singer, 1996)

dependent variables, is extremely small and will be calculated in the future using NPC. The results are consistently significant with respect to the *use of force*, are mostly significant with respect to *fatality*, *victory*, and *violation*, and are least robust on *duration*. Most of the results seem to be most clear (more statistically significant and on a smaller n) for the smaller subset of leaders whose total time in office is greater than 12 years, though for *duration* it is the medium tenure leaders with the most significant result.

Dep Var: Use of Force

Model	Estimate	S.E.	N_{MIDs}	$N_{Leaders}$
All	-.0306***	.0067	3842	465
All, Reg Entry	-.0262***	.0076	2925	364
All, Reg Entry Exit	-.0279**	.0091	1963	243
Medium Tenure	-.0327	.0299	1595	269
Long Tenure	-.0258***	.0078	1330	95

Dep Var: (Log) Duration

Model	Estimate	S.E.	N_{MIDs}	$N_{Leaders}$
All	-.0142†	.0088	4812	1109
All, Reg Entry	-.0105	.0103	3713	893
All, Reg Entry Exit	-.0156	.0124	2554	636
Medium Tenure	-.0916**	.034	2321	765
Long Tenure	-.0048	.0104	1392	128

Dep Var: Fatality

Model	Estimate	S.E.	N_{MIDs}	$N_{Leaders}$
All	-.0128*	.0063	4812	1109
All, Reg Entry	-.0139*	.0078	3713	893
All, Reg Entry Exit	-.0181*	.009	2554	636
Medium Tenure	-.0259	.0211	2321	765
Long Tenure	-.013†	.0084	1392	128

$^\dagger p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$ (one-sided)

Table 4.1: **Behavior over Time-in-Office.**
Estimated coefficients on *Years In Office* for various fixed-effects regressions on the respective dependent variables. *Use of Force* is a dichotomous variable; the estimate is in log-odds. Estimates for *Log Duration* and *Fatality* are linear regression coefficients. Models "All, Reg Entry" and "All, Reg Entry Exit" remove from the sample irregular entries into power, and irregular entries and exits. "Medium Tenure" subsets on total years in office ≤ 12; "Long Tenure" on > 12.

The estimated effects are substantial given that this only refers to within leader variation (see table 4.3). Using the baseline probability for the respective dependent variable as the

Dep Var: Victory

Model	Estimate	S.E.	N_{MIDs}	$N_{Leaders}$
All	-.0121	.0096	2280	214
All, Reg Entry	-.0163†	.0101	1829	178
All, Reg Entry Exit	-.0062	.0115	1266	131
Medium Tenure	.0728†	.0483	867	123
Long Tenure	-.0205*	.0104	962	55

Dep Var: Violation

Model	Estimate	S.E.	$N_{AllianceYears}$	$N_{Leaders}$
All	.0504*	.0253	1841	48
All, Reg Entry	.0565*	.0287	1323	32
All, Reg Entry Exit	.022	.029	983	21
Medium Tenure	.178†	.1187	538	22
Long Tenure	.0486*	.0291	785	10

$^\dagger p < 0.10$, $^* p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$ (one-sided)

Table 4.2: **Behavior over Time-in-Office (2).**
Victory and *Violation* are dichotomous variables; the estimate is in log-odds. Models "All, Reg Entry" and "All, Reg Entry Exit" remove from the sample irregular entries into power, and irregular entries and exits. "Medium Tenure" subsets on total years in office ≤ 12; "Long Tenure" on > 12.

prior probability or level of the dependent variable, an increase of ten years in office will be associated with a decrease in the probability of using force of 4.5% (or 10% reduction relative to the baseline), a reduction in the expected duration of a MID by 18 days (from the baseline of the mode of approximately 150 days, which represents an approximate 12% relative change), a decrease in fatalities of about 70 battle deaths (from the second mode of 500-1000 fatalities, a 10% relative reduction), a reduction in the probability of winning of 1% or a 10% reduction relative to the baseline of 10%, and an increased chance of violating a security commitment of about 2.5% which is a 250% relative increase from the baseline of 1%.

In summary, these analyses provide strong support for the five predictions generated from the theory that reputational incentives diminish with time-in-office. I will now subset the analysis by age, era, and regime-type to search for any informative causal heterogeneity.

Table 4.3: **Time-in-Office Analysis with Leader Fixed-Effects.**
Independent Variable: Years in office.

Dep Var	Abs Effect	% Effect	N_{Obs}	$N_{Leaders}$
Use Force	-4.5%***	-10%***	3853	468
Duration[1]	-18* (days)	-12%*	4820	1114
Fatalities[1]	-70* (b.d.)	-10%*	4820	1114
Victory	-1%†	-10%†	2271	212
Violation	2.5%*	250%*	1841	48

†$p < 0.10$, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (one-sided)
Effect estimates for a change of 10 years in office.
[1] Effect estimates depend on baseline.
Absent in full democracies; largely robust to subsetting on regular entry, regular entry and exit, and total time-in-office $\leq$ or > 12.

4.2.1 Subsetting by Age

I first subset by age, splitting the leaders into those who first came to power when they were less than 50 and those over 50 (50 is close to the median for the sample).

Table 4.4: **Analysis on Subsets of Leaders by Age at Start of Term**

Model	Estimate	S.E.	N_{MIDs}	$N_{Leaders}$
Force, Under 50	-.0252**	.0084	1337	146
Force, Over 50	-.0417**	.0176	1583	214
Log Duration, Under 50	-.0072	.0111	1634	352
Log Duration, Over 50	-.0244	.0235	2081	542
Fatality, Under 50	-.0114	.009	1634	352
Fatality, Over 50	-.0303*	.0162	2081	542
Victory, Under 50	-.0183*	.011	904	77
Victory, Over 50	-.0176	.0265	925	100
Violation, Under 50	.0484†	.0295	564	17
Violation, Over 50	.1467†	.1046	759	15

$^†p < 0.10$, $^*p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$ (one-sided)

Table 4.4 reveals that the estimated coefficients are two to three times larger for older leaders (except for *victory*), though statistical significance is present for some dependent variables for each subset. That the estimated coefficients are larger for older leaders provides evidence in favor of the *time horizon* mechanism, over the *hardening reputation* mechanism. This also shows, however, that these patterns are present both in young and old leaders, and so can not be accounted solely by some age related mechanism such as "youthful inexperience" or the political influence and legacy obsessions of older leaders.

4.2.2 Subsetting by Regime Type

I now subset the analysis on the polity score of the country (Marshall and Jaggers, 2002). This subsetting is important to see whether the estimated associations are present under different kinds of regimes. If they are present primarily under democracies then we will want to consider much more carefully potential confounding due to election incentives. In addition, the theory of *influence-specific reputation* predicts that time-in-office induced reputation incentives should be strongest for autocracies because it is there that reputations are most leader-specific.

Table 4.5 reveals that the estimated associations are consistently stronger and more significant under non-democracies (defined as having a combined polity score less than 7).

Table 4.5: **Analysis on Subsets of Leaders by Regime Type**

Model	Estimate	S.E.	N_{Obs}	$N_{Leaders}$
Force, Autocracy	-.0382***	.0104	1554	207
Force, Democracy	-.0199*	.0118	1323	163
Log Duration, Autocracy	-.0235†	.0154	1920	460
Log Duration, Democracy	.0041	.0064	1794	472
Fatality, Autocracy	-.0149†	.011	1920	460
Fatality, Democracy	-.0129	.0123	1794	472
Victory, Autocracy	-.0324*	.0172	869	96
Victory, Democracy	-.0054	.0127	927	85
Violation, Autocracy	.088*	.0489	404	18
Violation, Democracy	.038	.0373	891	14

$^\dagger p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$ (one-sided)

While the only coefficient that is significant under democracy is that for the *use of force*, they are all, except duration, in the expected direction, suggesting support for my theory even under the democracy subset. An NPC test will assess whether this combined evidence is sufficient to warrant rejecting the null at a *p*-value less than 0.05 for democracies. These results suggest that elections are not a confound generating the overall result. Though it is possible that better controls for time until election will sharpen or otherwise affect the results for democracies.

These results are consistent with the theory of *influence-specific reputation*. Note that these results could not arise simply from differences in the mean or variance of conflict behavior by leaders of non-democracies, since the first is accounted for with fixed effects and the second will express itself in larger standard errors. Rather, these results show that an additional year in office for a non-democrat is associated with more sharp reductions in the use of force, in the duration of disputes, and the probability of victory, and with more violations. This is consistent with the theory that what is motivating these changes in behavior is concern for leader-specific reputation, and that leader-specific reputation is most salient for autocracies.

4.3 Alternative Explanations

There are a number of potential alternative explanations that might account for this conflict behavior. However, none of them provide a consistent account for all the behavior observed. Table 4.6 summarizes the most plausible of these. I will briefly describe them here.

Electoral Incentives: Pacific Electorate or Diversionary Theory

There are two explanations that depend on incentives related to the approach of elections. The first—"Pacific Electorate"—argues that the approach of elections will increase the influence of the relatively more peaceful electorate. The second—"Diversionary Theory"—argues that citizens tend to rally behind a leader who is engaged in certain kinds of foreign military conflicts; as such, leaders have incentives to manufacture these kinds of conflicts as elections approach. While we might consider these explanations in more detail, both of them clearly fail to provide a sufficient account for the observed behavior, which has been shown to be strongest amongst autocracies. To the extent that we are interested in this mechanism, an ideal research design is to compare conflict behavior of re-electable incumbents just before an election with formerly re-electable incumbents right after a successful reelection. We could also compare re-electable incumbents who lose the election but are still in office for some months ("lame ducks"; Haynes, 2012) with re-electable incumbents who win the election (though this sample size might be small). Theories of electoral incentives make strong predictions when comparing these categories of leaders, who are otherwise close to each other in time-in-office and time until the next election. In short, the discontinuity in incentives before and after an election provides a clean design for identifying the effect of election incentives.

Learning and Experience

Leaders will gain experience with time-in-office. As Potter (2007) theorized, this may mean they are more able to avoid costly conflicts. However, it is unclear what Potter would predict about an experienced leader's behavior conditional on them experiencing a militarized dispute. I conjecture, in the spirit of Potter's theory, that an experienced leader should have less costly (shorter, less fatal) conflicts. Similarly, it follows from greater experience that a leader should be more likely to win their militarized disputes. It is unclear what effect experience would have on their tendency to use force (maybe more since the leaders are more militarily effective, but maybe less since they don't need force to achieve their aims) or on the tendency to violate security commitments. As table 4.6 makes clear, this explanation is strongly inconsistent with the findings on *victory*, which is where we might most expect to cleanly see the effects of experience. It also otherwise fails to provide a parsimonious account for the increase in violations.

Or we might theorize as Horowitz, McDermott, and Stam (2005) do: greater experience and consolidation of power may provide leaders with more freedom to execute aggressive

Table 4.6: **Testable Implications of Different Explanations**

	Use of Force	Duration	Fatality	Victory	Violation	Other Predictions
Data	–	–	–	–	+	
Reputation (priors)	–	–	–	–	+	
Reputation (horizon)	–	–	–	–	+	
Pacific Electorate	–	–	–	? or +	–?	Only in Democracies
Diversionary Theory	+	+	+	? or –	–?	Discontinuity at Election
Learning/Experience	?	–	–	+	?	
Selection	?	+	+	0^1 or –	? +?	
Disruptive Transition	? or –	? or –	? or –	0	–	
Insecure Transition	–	–	–	?	?	Effect Mostly in First Years
Modern Peace	–	–	–	0	0	Effect Sizes too Large

Green coloring indicates a prediction consistent with the data (light green indicates a weaker prediction); white indicates an unclear prediction; red coloring indicates a prediction inconsistent with the data (maroon indicates a weaker prediction). As is evidence, of the four most plausible alternative theoretical explanations, none generate strong predictions consistent with the five observed variables, and they each generate some predictions that are in error.

[1] (Chiozza and Goemans, 2004) find that victory or draw in a dispute does not affect tenure, though theoretically we might still expect a negative association.

foreign policies. However, such an interpretation would predict increasing *uses of force*, *duration*, and *fatality*, each of which is inconsistent with the data, and it remains inconsistent to observe decreased probability of winning.

Selection into Office

It may be that a leader's conflict behavior influences the probability of that leader staying in power. A leader might be more likely to lose office after experiencing a costly conflict that doesn't end in victory. (Chiozza and Goemans, 2004), however, find having a victory or draw doesn't effect tenure, only losing a dispute does (which does not confound with our analyses).[15] Furthermore, if long bloody conflicts in some circumstances make a leader more likely to lose office (Gartner, 2008), then the observed association will be positive between duration/fatality of MIDs and time-in-office. Thus, it doesn't seem likely that the effect of leader behavior in conflict on tenure length is able to account for the pattern of results.

Disruptive Transitions

It may be that leader transitions are disruptive. New leaders might introduce a new set of preferences and skills to the helm of a state which might destabilize international relations. Hostile powers might seek an opportunity to probe and challenge the new leader. The new leader might challenge the status quo, abandoning some former commitments and pursuing new ambitions. As such, we might expect there to be more conflicts during the period immediately after a new leader comes into power.

It is unclear exactly what this theory would predict for the costliness of conflict. On the one hand, since the probing of rivals might be more tentative and end relatively early, there might be more MIDs but each with lower costs. On the other hand, if probing involves costly escalation so as to reveal the true resolve and preferences of a leader, then the MIDs occurring after a leader transition may be more costly. For the purposes of developing this alternative explanation, I will conservatively accept the interpretation more consistent with the actual results, so that disruptive leader transitions should lead to more uses of force, longer lasting MIDs, with more fatalities, for a new leader. This is thus consistent with three of the predictions.

However, it is not obvious what we should predict about victory. There are more costly conflicts, but should the new leader win these greater abundance of costly conflicts at a greater rate than would that same leader later in his tenure when he can be more selective? Furthermore, if we consider the "probing" mechanism we might expect other states to push harder against the new leader to test his resolve. Unless we also embrace a reputational

---

[15]It is worth noting that Chiozza and Goemans (2004) find that selection effects, if they operate, are only for autocrats. This is consistent with the finding in this paper of heterogeneity by regime-type, though this paper finds some effect for democracies, whereas Chiozza and Goemans (2004, 613) write "neither victory, nor defeat, nor a draw in war affects the hazard of losing office for democratic leaders."

motive on behalf of the new leader, it's not clear why this new leader would push harder back against these probing efforts. Consequently, I assign this alternative explanation a prediction of no effects on victory.

What should we predict about violation? If new leaders disrupt the status quo because of their divergent preferences and skills, then we would expect that if this expresses itself at all in the maintenance of security commitments, that it does so through the violation of older security commitments in the early years of the new leader's tenure. Thus, this explanation would predict a decreasing tendency to violate security commitments with time-in-office, contrary to the observed results.

Insecure Transition

An alternative possibility is that some leaders (specifically autocratic leaders) may have a weak hold on office initially, and they may use military engagements to bolster their authority. This implies that new autocratic leaders will seek opportunities to engage in disputes in which they can generate rally effects or otherwise secure their hold on power (Chiozza and Goemans, 2011). Depending on how we interpret the literature, this might lead us to predict that autocrats early in the time-in-office will be more likely to use force in disputes, and to have longer and more fatal disputes. This should also incentivize leaders to seek victories; however, if the leaders are too eager for a costly conflict they might pick ones in which victory is less likely. As such, I leave the prediction about victories uncertain. It is also not clear how this would relate to support for security commitments. Thus, the first three predictions are identical to those made about reputation (perhaps because domestic audiences tend to support leaders who act in ways that defend national honor).

To evaluate this alternative explanation, it would thus be ideal if we have some other empirical domain on which these theories diverge. To achieve this, I reanalyze the five predictions with a more flexible functional form for time-in-office. I use β_{bK} $TiOBeforeK$ $+\beta_{dK}$ $AfterK$ $+\beta_{aK}$ $TiOAfterK$, where $TiOBeforeK$ is the normal measure of Years-in-Office before the Kth year, but set to 0 for years after K, and $TiOAfterK$ is set to 0 for years before K and is the normal measure of Years-in-Office otherwise. $AfterK$ is a dummy variable for years in office above K. For the purposes of this analysis, I set $K = 4$, though other thresholds could be examined.

If the results reported above are driven by incentives facing leaders as soon as they come into office then we should not see these associations after the first few years (such as by the $K = 4^{\text{th}}$ year). Thus, if early-in-tenure rally effects are driving the results then β_{a4} should be indistinguishable from 0. Table 4.7 reports results with this more flexible functional form.[16] Three of the estimated coefficients are significant in the predicted direction ($\hat{\beta}_{a4}^{Force}$, $\hat{\beta}_{a4}^{Fatality}$, and $\hat{\beta}_{a4}^{Victory}$), one is (insignificant) in the predicted direction ($\hat{\beta}_{a4}^{Violation}$), and one is effectively

---

[16]In a working paper version of this chapter I report a LOESS regression over years-in-office which allows for a visual assessment of the functional form. In the next version of this analysis I intend to estimate a flexible polynomial, and plot the estimated associations and confidence intervals.

zero ($\hat{\beta}_{a4}^{Duration}$). Table 4.7 suggests that the evidence after the first four years is as supportive of the five predictions (taken as an aggregate) as the evidence during the first four years. Table 4.8 then performs this analysis again, but considering only non-democracies. Again, the evidence is as strong, if not stronger, in the years after the first four, as it is during them.

Table 4.7: **Examining Marginal Effects over Time-in-Office**

DepVar	β	Estimate	S.E.	p-values	N_{Obs}	$N_{Leaders}$
Force	b4	-0.0897†	0.0609	0.0704	2919	361
Force	d4	0.0338	0.1878	0.5714	2919	361
Force	a4	-0.0312***	0.0087	0.0002	2919	361
Duration	b4	-0.1680**	0.0652	0.0051	3707	892
Duration	d4	-0.6362***	0.2028	0.0009	3707	892
Duration	a4	0.0000	0.0109	0.4990	3707	892
Fatality	b4	-0.0776*	0.0412	0.0299	3707	892
Fatality	d4	-0.1168	0.1409	0.2037	3707	892
Fatality	a4	-0.0139†	0.0093	0.0671	3707	892
Victory	b4	0.1532	0.0992	0.9387	1826	176
Victory	d4	0.3858	0.3044	0.8975	1826	176
Victory	a4	-0.0198*	0.0113	0.0395	1826	176
Violation	b4	0.0484	0.2539	0.4243	1323	32
Violation	d4	1.2897†	0.8508	0.0648	1323	32
Violation	a4	0.0301	0.0312	0.1673	1323	32

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (one-sided)
This table lists the estimated coefficients for three variables for time-in-office for analyses on five different dependent variables. β_{b4} refers to the coefficient on a measure of years in office for the first 4 years; β_{a4} for the years after the first 4; β_{d4} is a dummy variable for after the first 4 years.

Table 4.8: **Marginal Effects over Time-in-Office (for** $Polity < 7$**)**

DepVar	β	Estimate	S.E.	p-values	$N_{\mathbf{Obs}}$	$N_{\mathbf{Leaders}}$
Force	b4	-.1665*	0.0976	0.0440	1542	203
Force	d4	.2282	0.2823	0.7906	1542	203
Force	a4	-.0509***	0.0119	0.0000	1542	203
Duration	b4	-.1254	0.1031	0.1123	1913	456
Duration	d4	-.6023*	0.3252	0.0323	1913	456
Duration	a4	-.0121	0.0177	0.2462	1913	456
Fatality	b4	-.0551	0.0661	0.2025	1913	456
Fatality	d4	-.1203	0.2102	0.2836	1913	456
Fatality	a4	-.0144	0.0128	0.1306	1913	456
Victory	b4	.3502	0.1855	0.9704	867	95
Victory	d4	1.5212	0.5602	0.9967	867	95
Victory	a4	-.0556**	0.0204	0.0033	867	95
Violation	b4	-.1469	0.3024	0.6864	1158	21
Violation	d4	.8326	0.8657	0.1681	1158	21
Violation	a4	.0303	0.0314	0.1670	1158	21

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (one-sided)

This table lists the estimated coefficients for three variables for time-in-office for analyses on five different dependent variables, only for country-years in which the combined Polity score is less than 7. β_{b4} refers to the coefficient on a measure of years in office for the first 4 years; β_{a4} for the years after the first 4; β_{d4} is a dummy variable for after the first 4 years.

Modern Peace

Could it be that this analysis is picking up a general secular trend in the reduction of costly military conflict associated with the "Modern Peace" or "Liberal Peace"? Figure 4.1 shows the distribution of observations over time, which reveals that a large portion of the sample occurs after 1950. However such an explanation is insufficient for two reasons. First, it doesn't account for the observed results with respect to victories or violations. Second, the estimated effect sizes are too large. If the probability of using force in a MID is reduced by 10% after 10 years (from the starting baseline of 50%), then after 100 years there should be almost no observed acts of force. Similarly for the other dependent variables.

I subset the analysis on era to examine whether the association is particular to post 1950 (Table 4.9). Some results are stronger post-1950 (use of force and violation), while others are stronger pre-1950 (duration, fatality, and victory). This suggests that the aggregate results are not being driven by any particular era.
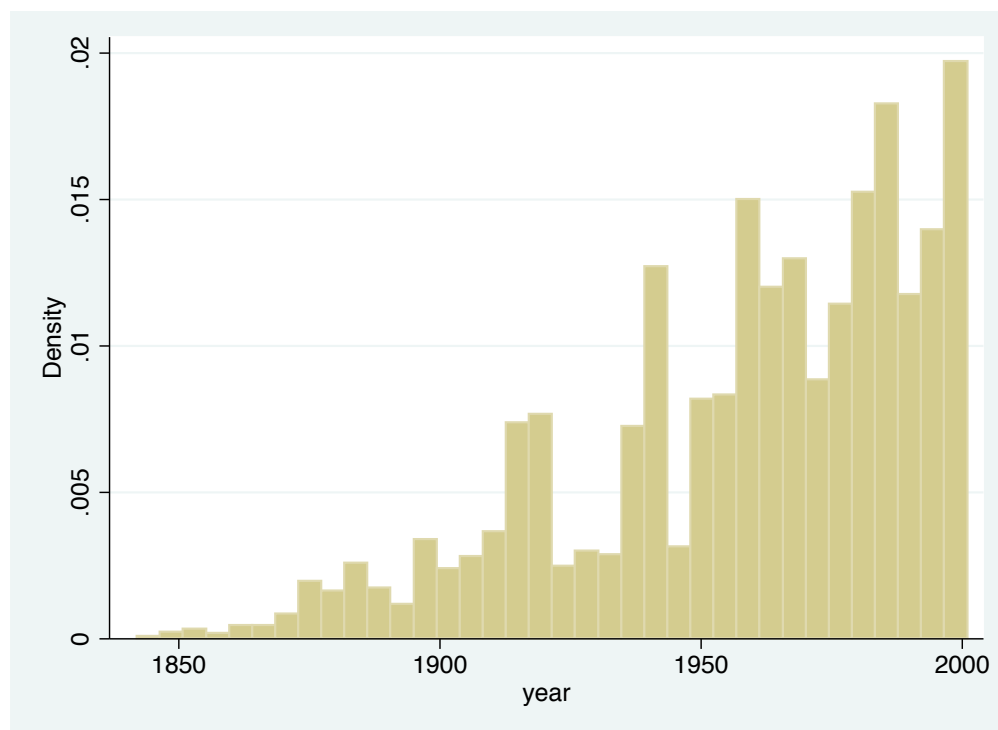

Figure 4.1: **Histogram of Observations by Time**

Table 4.9: **Analysis on Subsets of Leaders by Era**

Model	Estimate	S.E.	N_{Obs}	$N_{Leaders}$
Force, Pre 1950	-.0172*	.0088	1149	138
Force, Post 1950	-.0565***	.0144	1771	222
Log Duration, Pre 1950	-.0128	.011	1494	386
Log Duration, Post 1950	-.0047	.0234	2221	508
Fatality, Pre 1950	-.0202*	.0095	1494	386
Fatality, Post 1950	-.0023	.0096	2221	508
Victory, Pre 1950	-.0238*	.011	939	95
Victory, Post 1950	.0215	.0294	890	82
Violation, Pre 1950	.0467	.0369	491	15
Violation, Post 1950	.071†	.0464	832	17

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (one-sided)

4.4 Conclusion

Leaders throughout history seem to have been very concerned about how other leaders perceive them. This chapter provided a novel examination of the effect of reputational considerations on conflict behavior, leveraging the fact that leaders should care more about reputation early in their time-in-office. This chapter tested the four predictions deduced in chapter 2 about the effects of concern for reputation on conflict behavior, as well as one additional prediction concerning the fulfillment of security commitments. Support for each of these five predictions were found in a leader fixed-effects analysis of state conflict behavior, and the total pattern of results provides strong support for my theory that reputational incentives diminish with time-in-office and that concern for reputation plays a crucial role in interstate conflict escalation. Alternative explanations are ruled out through consideration of critical areas where they predict different outcomes than my theory, and through other critical tests.

At the least, this chapter provides strong empirical evidence that leaders behave differently later in their time-in-office. Irrespective of the cause of this pattern, it is an important and novel finding that leaders, but especially autocratic leaders, engaged in militarized disputes are less likely to use force, to engage in long costly disputes, and to win their disputes, as well as more likely to violate their alliance commitments. The makers of foreign policy will want to be extra wary when confronting new leaders, and to put some consideration into saving their potentially insulting foreign policy demands for the later years of the target leader's tenure. The following chapter puts these results in context and concludes.

Chapter 5

Conclusion

5.1 The Dark Matter of International Relations: Concern for Reputation

Historian Donald Kagan (1995, p. 8) finds that "consideration of practical utility and material gain, and even ambition for power itself, play a [small role] in bringing on wars ... [whereas often] some aspect of honor is decisive." Thomas Schelling (1966, p. 124) argues that "reputation for action ... is one of the few things worth fighting over." Similar quotes about the extreme importance of reputation to leaders as a motive for war can be found by many scholars and leaders, throughout history. My dissertation provides a systematic examination of this neglected insight. By analyzing two distinct empirical domains, each with a plausible strategy for mitigating bias from unobserved confounding, I show that variation in concern for reputation accounts for very large behavioral effects in international relations. Southern US presidents, as compared with non-Southern presidents, are about twice as likely to use force, experience disputes about twice as long, and are three times more likely to win their militarized disputes. Leaders ten years later in their tenure when reputation is less important to them are about 10% less likely to use force, experience disputes 10% shorter, with 10% fewer fatalities, are 10% less likely to win their disputes, and 250% more likely to violate their security commitments; these within-leader effects are attenuated estimates of the full effect of a leader's concern for reputation since they don't include any cross-country or cross-leader sources of variation. Some potential cross-leader causes of concern for reputation include: a leader's cultural background and political beginnings (Goldgeier, 1994), the strategic environment facing a country including being a colonial power, having hostile neighbors, great power status, the balance of capabilities, (Sechser, 2007) and having long-standing rivalries, the international diplomatic culture, the value of latent disputes (valuable disputed territories), the presence of potential separatist groups (Walter, 2006), incentives arising from domestic audiences (Fearon, 1994a; McGillivray and Smith, 2008b; Tomz, 2007a; Weeks, 2008), and in built psychological templates for revenge and honor

(Pinker, 2011; Whitfield, 2011).

In summary, the analyses reported here provide evidence consistent with the strongest claims about the importance of concern for reputation. Concern for reputation is the dark matter of international relations: it is hard to observe and study, but it is necessary for most theories of international relations. The research reported here provides some of the most systematic evidence of this elusive phenomenon to date.

Thucydides and Thomas Hobbes, among others, understood that concern for reputation was one of the primary causes of war. For this reason, Hobbes regarded the reduction of honor considerations following the overcoming of the security dilemma to be one of the primary mechanisms by which a powerful state reduced the incidence of inter-group conflict; contemporary scholars agree (e.g. Pinker, 2011). Deterrent reputations are essential when predation is an otherwise viable strategy. Once an individual or group is sensitive about their coercive reputations the smallest dispute can readily explode into a massive existential conflict. Both the U.S. and the Soviet Union gambled a hugely disproportionate share of their military capacity to winning in their respective quagmires of Vietnam and Afghanistan. Both superpowers escalated dangerously close to nuclear war over relatively minor issues such as the sovereignty over and force deployments in Cuba and Berlin. This reputation-obsessed behavior resonates across the centuries, as far back as the first reliably recorded great-power war: the Peloponnesian War. The Peloponnesian War officially began after the Athenians rejected a very modest Spartan ultimatum *on principle*; the Spartan ultimatum was itself issued to protect the Spartan's reputation as reliable allies; and the whole dispute initially erupted from a domestic political dispute in a minor peripheral town that happened to embody the wounded pride of Corinth under the insolence of its daughter colony Corcyra.

5.2 But Do Reputations Matter?

Many have argued that the obsession over reputation is a gross source of harm. Daryl Press (2005, p. 1), for example, provides one of the most recent and articulate challenges to the notion that reputation is something worth pursuing. Press summarizes: "Tragically, those countries that have fought wars to build a reputation for resolve have wasted vast sums of money and, much worse, thousands of lives." Other scholars who have questioned whether leaders draw or act on reputation inferences include Ted Hopf (1994), Jonathan Mercer (1996), Shiping Tang (2005), and Robert Jervis (1991).

My read of the broader evidence, however, suggests that reputational inferences have, in fact, played a very important role historically, and to a lesser extent remain important today. This isn't to say that the reputational benefits from the Vietnam War, or any particular war, were worth the costs, but that we have reason to believe that those benefits are substantial. I believe this to be the case for a number of reasons: (1) the evidence presented by Press is, in fact, consistent with a theory of reputation; (2) finding evidence to show that reputational inferences do not have substantial welfare implications is actually very hard to do; most tests

of this would be under-powered; (3) we should assign substantial theoretical weight to the instincts and beliefs of leaders; the burden of evidence should be high before we reject a view of the world that so many leaders in so many countries and eras believe to be true; (4) it is rational to draw reputational inferences in stable social systems where there is uncertainty and opportunities for reciprocity and so would be surprising if reputational inferences didn't take place; (5) we have abundant evidence that intelligence organizations believe the history of a leader is informative of how they are likely to behave in the future.

5.2.1   Press's Selection of Cases

Daryl Press sought to investigate whether the "past actions" of a leader affect the perceived credibility of that leader, or whether credibility is primarily a function of current factors such as power and interests. Press looks at a number of cases in which a leader had failed to fulfill his threats in a previous crisis, and searches for evidence that others use this history as evidence for or an argument for why this leader is unlikely to fulfill a current threat. Specifically, Press looks at Nazi inferences about British and French credibility in 1939, American and British inferences about Khrushcev's credibility in the Berlin crises, and U.S. inferences about Soviet credibility during the Cuban Missile Crisis. Examining the historical archives, Press finds little evidence of decision makers relying on the past behavior of their rival, but extensive evidence of them using facts about power and interest to inform their assessment of the credibility of their rival; furthermore, assessments of credibility seem to co-move with changes in power and are not overridden by the history of irresolute actions.

There are a number of reasons, however, why Press' research strategy does not provide evidence against the hypothesis that past actions inform the assessments of credibility. First, Press' design only examines one value of the independent variable: Press only examines cases where the target had failed to fulfill past threats. Lacking variation on the independent variable, Press' design, then, can only draw a causal inference if theories about reputation make strong divergent predictions from Press' theory for the value of the independent variable that Press examines (where leaders have failed to fulfill their past commitments). The theory of reputation that I find plausible, however, makes identical predictions to Press' theory for the kinds of cases he examined.

The theory of reputation that I want to consider here concerns having a reputation *for fulfilling commitments*, which involves inferences that shift credibility away from the baseline inference that would be drawn based on other factors (such as power and interest). A reputation for breaking commitments thus implies that inferences about credibility *should be solely based on other factors* and therefore should be the same as Press' baseline (null) inference. To express this formally, let $P_i(X, Y)$ be the probability that a leader fulfills a threat, given their reputation X and their interests Y in the particular issue, for theory of reputation i. They can have either a reputation for fulfilling threats $X = R$ or a reputation for not fulfilling their threats $X = \neg R$. Similarly, they can have high interests over an issue $Y = I$ or low interests $Y = \neg I$. We will consider Press' theory $i = 0$, what Press

calls "current calculus", where only current factors having to do with power and interest matter, Press' alternative theory of "past actions" ($i = PA$), and my theory of reputation as something that can increase credibility from the baseline as ($i = R$); let $i = D$ refer to the actual data, that is the observed probability of fulfilling a threat given the conditions. Table 5.1 summarizes these predictions.

Table 5.1: **Probability of Fulfilling a Threat**

	High Interest	Low Interest
Reputation for Fulfilling Threats	$P_0(R, I)$ =High $P_R(R, I)$ =V. High $P_{PA}(R, I)$ =High	$P_0(R, \neg I)$ =Low $P_R(R, \neg I)$ =Medium $P_{PA}(R, \neg I)$ =High
Reputation for not Fulfilling Threats	$P_0(\neg R, I)$ =High $P_R(\neg R, I)$ =High $P_{PA}(\neg R, I)$ =Low	$P_0(\neg R, \neg I)$ =Low $P_R(\neg R, \neg I)$ =Low $P_{PA}(\neg R, \neg I)$ =Low

Press' evidence all comes from the bottom row where, the table makes clear, both the null and the above theory of reputation predict the same thing. Press shows that $P_D(\neg R, I)$ =High and $P_D(\neg R, \neg I)$ =Low, and that reputational inferences played no role in assessments of credibility. However, this relationship is consistent with both Press' theory of current calculus and my theory of reputation as enhancing commitment ($i = 0$ and $i = R$). To test my theory of reputation we would need to acquire evidence from one of the cells from the top row. For example, we could test $H_0 : P_D(R, I) = P_D(\neg R, I)$ vs $H_R : P_D(R, I) > P_D(\neg R, I)$; or $H_0 : P_D(R, \neg I) = P_D(\neg R, \neg I)$ vs $H_R : P_D(R, \neg I) > P_D(\neg R, \neg I)$. In short, to test my theory of reputation we need variation on the independent variable; we need to compare R with $\neg R$. Alternatively, we could select cases where leaders had a reputation for fulfilling threats—conditions (R, I) or $(R, \neg I)$—to evaluate whether past actions are used in the inferences drawn by observers, since the above theory of reputation predicts the inferences drawn should be higher than those implied by the baseline set of factors.

5.2.2 The Difficulty of Showing the Absence of Reputational Inferences

A second concern with Press' inference is that it may be very hard to detect reputational inferences using historical (or any) methods with sufficient precision.

Welfare Relevant Changes May Be Too Small to Detect

Let $\Delta\phi(\mathbf{h_1}, \mathbf{h_2}; \mathbf{Z_1}, \mathbf{Z_2})$ represent the change in the belief (ϕ) about the credibility of some particular threat under the counterfactual comparison of two (vectors of) different histories $\mathbf{h_1}$ and $\mathbf{h_2}$ (each of length k), and given the (vectors of) two contexts for the dis-

pute and its counterfactual $\mathbf{Z_1}$ and $\mathbf{Z_2}$ (each of length l). Scholars who wish to argue that reputational inferences don't matter are effectively saying that any two sufficiently different histories of behavior have little (or no) effect on inferences about credibility, H_A : $\forall(\mathbf{h_1}, \mathbf{h_2}, \mathbf{Z_1}, \mathbf{Z_2}) \in (\mathbb{R}^k, \mathbb{R}^k, \mathbb{R}^l, \mathbb{R}^l)$ such that $D_h(\mathbf{h_1}, \mathbf{h_2}) > \epsilon_h$, and $D_Z(\mathbf{Z_1}, \mathbf{Z_2}) < \epsilon_Z$ then $|\Delta\phi(\mathbf{h_1}, \mathbf{h_2}; \mathbf{Z_1}, \mathbf{Z_2})| < |w|$. $D_j(\cdot, \cdot)$ is a function that gives a scalar summary of the difference in the vectors of ($j = h$) histories or ($j = Z$) contexts. In words, this alternative hypothesis says that, holding other factors relatively constant, beliefs about credibility will not change by a substantial amount for a sufficiently large change in the history of behavior. The null hypothesis (H_0), then, is that history of behavior has a sufficiently large effect that it "does matter": $\exists(\mathbf{h_1}, \mathbf{h_2}, \mathbf{Z_1}, \mathbf{Z_2}) \in (\mathbb{R}^k, \mathbb{R}^k, \mathbb{R}^l, \mathbb{R}^l)$ such that $D_h(\mathbf{h_1}, \mathbf{h_2}) > \epsilon_h$, and $D_Z(\mathbf{Z_1}, \mathbf{Z_2}) < \epsilon_Z$ then $|\Delta\phi(\mathbf{h_1}, \mathbf{h_2}; \mathbf{Z_1}, \mathbf{Z_2})| \geq |w|$.[1] w is the least amount that beliefs about credibility would have to change for any given threat that would have a substantial effect on the welfare of the leader or the state. Implicit to any tests of these two hypotheses, then, is that measurements of $\Delta\phi$ smaller than $|w|$ in size must be feasible. Otherwise, the design would lack the measurement power to reject the null hypothesis (H_0) that reputation does matter.

I argue that the effect size that we need to detect, w, is likely much smaller than our ability to detect; our ability to measure ϕ (the belief about the probability a threat is credible), and hence also changes in ϕ, is very coarse, and the effect size that our design needs to be able to detect, w, is very small. Our measurements of the beliefs of observers about the credibility of threats (ϕ) are not that precise, surely not better than with a 5 percentage point standard error. If they were, then historians would employ a more precise vocabulary than ordinary language (namely numbers or a standardized terminology) to communicate these different levels of beliefs. Similarly, a $w < 5\%$ is entirely plausible for certain states. For states expecting to uphold many deterrent threats and employ some compellent threats— namely great powers, states in hostile neighborhoods, or states involved in a rivalry over territory of substantial value—the welfare effects of a slight increase in credibility for any given threat is multiplied by the number of deterrent and compellent threats, and the value of the issues at stake in those other disputes. Even a small change (say $w = 3\%$) in the perceived credibility of each threat will have large welfare implications since this effect is magnified across many ongoing deterrent threats, and, for great powers and conflictual eras, plausibly 100s of expected future coercive situations. Appendix A.2 makes this argument in more detail.

Another reason why it might be particularly hard to detect changes in reputational beliefs using historical evidence is that they may be generally slow moving, and so be swamped in leaders' explicit reflections and in policy debates by other faster changing factors, such as regional military capabilities, the risk of other conflicts breaking out, the reliability of allies, and so forth. Reputation, like other fixed background factors such as whether the opponent

---

[1]We could soften these hypotheses by looking at the average change in $\Delta\phi$ for a "reasonable" set of $(\mathbf{h_1}, \mathbf{h_2}, \mathbf{Z_1}, \mathbf{Z_2})$. The basic point remains.

is a monarchy, may simply be taken for granted in most conversations: it influences thinking without being explicitly stated. This is an especially acute problem in a design such as Press's since he doesn't have variation in the reputation-relevant history of the subjects of the reputational inferences (all of Press's cases focus on leaders who had backed down in a recent sequence of crises).

A final reason why reputational inferences may be hard to detect is that their subjective nature may mean that they are less represented in policy conversations. If policy debates privilege arguments that involve objective factors, such as military capability or geopolitical interests, over more subjective reputational inferences then there will be a systematic bias downwards in the prevalence of reputational arguments in policy debates, relative to their importance in the minds of policy elites.

5.2.3   Other Reasons to Take Reputation Seriously

There are at least three other substantial reasons why we should be cautious about rejecting the hypothesis that leaders draw and act upon reputational inferences. These are that: (1) the belief that policy elites draw reputational inferences is very widespread; (2) it is rational to draw reputational inferences in uncertain environments like the international system; and (3) there does seem to be abundant evidence that reputational inferences about leaders are drawn.

(1) Elites Believe that Reputations Form

If there is one point of agreement amongst all scholars of reputation it is that leaders think that others draw reputational inferences about them, and that these inferences have important implications for others' behavior. Jervis writes that "states seem preoccupied with the credibility of their threats" (Jervis and Snyder, 1991, p. 26). Press (2005, pp. 1-2) writes "The conventional wisdom holds that credibility depends on a country's past behavior— its history of keeping or breaking commitments... Evidence shows that presidents, prime ministers, and dictators believe the conventional wisdom about the sources of credibility."

Scholarly inference should combine, in a rational manner, the weight of our prior beliefs and the weight of new evidence. In the case of reputation there is a very strong belief amongst policy elites and other scholars that reputations form and are acted upon. It may be that all these leaders throughout history have been fundamentally mistaken about the role of reputation; if so, it would be a major finding to be able to demonstrate this. However, it is also plausible that there are flaws in the interpretation of the particular contrary evidence. It is an extraordinary claim that most leaders throughout history have been mistaken about the nature of reputational inferences; as such, we should withhold judgement about such a claim until suitably extraordinary evidence is provided. Scholars who believe that reputational inferences don't form should invest in demonstrating their conjecture with more extensive evidence: employing different case selection strategies, different measurement techniques,

and different methods all together such as laboratory or survey experiments on foreign policy elites.[2]

(2) Rationality of Reputational Inferences

Whenever previous behavior provides *some* additional information about future behavior, conditional on other observable factors, drawing a reputational inference is a rational and prudent thing to do. Further, reputational inferences are often very useful if one has limited cognitive capacity or information, since in many cases the history of past behavior predicts future behavior much better than other combinations of factors.[3] Reputational inferences remain invaluable in a world with perfect information; so long as there is strategic uncertainty—which arises whenever there are multiple equilibria—past behavior can and frequently is used to support certain equilibria (such as Tit-for-Tat in the Iterated Prisoner's Dilemma). Drawing reputational inferences is beneficial whenever (1) there is some uncertainty about an actor's future behavior (arising either from uncertainty over unobserved factors or strategies), and (2) past behavior can be used to predict future behavior. It's hard to imagine that this doesn't apply to international relations. And if it does apply to international relations, it's hard to believe that leaders and foreign policy apparatuses would fail to benefit from the information present in others past behavior.

(3) Evidence of Reputational Inferences

One final argument for why we should continue to believe that reputations probably do form is that there is, in fact, abundant evidence that they do. First, consider that most times that a person appends an adjective to a description of a leader they are drawing an inference about how that leader is likely to behave in the future based on their past behavior. To sample from the first pages of the historical book closest to my computer, I see in Michael Dobbs history of the Cuban Missile Crisis the following quotes: Kennedy refers to Khrushcev as a "fucking liar" (Dobbs, 2008, p. 7; Dallek, 2003, p. 429), in response to reading the CIA report about the Soviet resumption of nuclear testing: "[JFK] complained to his brother that the Soviet leader had behaved like "an immoral gangster... not as a statesman, not as a person with a sense of responsibility." (Dobbs, 2008, p. 7) Khrushchev says about Kennedy after their first meeting that Kennedy is "Not strong enough...Too intelligent and too weak." (Dobbs, 2008, p. 36) Statements of this kind made amongst decision makers are common. It is hard to believe that private statements such as these don't reflect reputational inferences, that is beliefs of the speakers that are based on the other's past actions that inform their

---

[2]In unpublished work I have run a survey experiment on over 1000 respondents who can be reasonably thought to represent the beliefs of some foreign policy elites (specifically, the respondents were readers of Harvard Professor Stephen Walt's foreign policy blog). I found substantial evidence of reputational inferences being drawn.

[3]We see this in statistical models: a single lagged dependent variable often has more predictive power than every other variable, and frequently than the whole set of other variables.

expectations of the other's future behavior (for a paper length examination of the inferences about resolve drawn by Khrushchev and Kennedy during this time, see Lupton, 2012).

We see reputational inferences in the interest rates that sovereign bonds can be sold for. It is well known that a state's history of debt repayment has a clear and direct effect on the kinds of interest rates the state can sell sovereign bonds for (Tomz, 2007b). How plausible is it that a state's history of debt repayment is hugely relevant for inferences about future debt repayment, but a state's history of behavior in military crises is not relevant for inferences about future behavior in military crises.

We see reputational inferences in survey experiments. For example, in unpublished work I administered survey experiments to over 1000 educated foreign-policy interested respondents (the respondents were readers of Stephen Walt's foreign policy blog). My survey involved a scenario of a military crisis in which different aspects (treatments) of the scenario differed for each respondent. I then asked respondents how likely they thought either side in the dispute was to back down before a war broke out. The treatments that had by far the largest effect on perceptions of resolve were each side's history of behavior, much more so than differences in military capability.[4]

In summary, the claim that leaders don't draw or act upon reputational inferences is a very strong one. If true, this would overturn a key assumption on which many theories of international relations are implicitly built, and it would show that the foreign policy of most states has been grossly mismanaged. However, the evidence for this claim is relatively thin; the evidence consists of an absence of clear historical statements invoking reputational inferences to justify foreign policy actions. There are a number of alternative explanations we should explore more seriously, including the possibility of selection bias in certain empirical analyses, the possibility that the relevant magnitude of changes in reputational inferences are too small to reliably detect, and that reputational inferences as subjective and slow-changing inferences tend to be background and unstated features of a policy deliberation. In my judgement, the weight of evidence remains strongly in favor of the notion that reputational inferences do form; those scholars who are convinced otherwise should continue to undertake research that should, if they are correct, provide evidence of their position. The policy and theoretical implications are great of such a result being true. But until then, the conventional wisdom remains persuasive.

5.3 Empirical Extensions

There are many possible extensions of this research. The most obvious is to examine other sources of variation in concern for reputation for the predictions deduced in chapter 2. For example, Todd Sechser (2007) argues that weak target states will have an assurance problem

---

[4]A counterargument to this kind of evidence is that these types of surveys are unrealistic. Since they deprive the respondent of the rich set of information that is available in actual scenarios, one would expect reputational inferences to play a larger role.

and so will want to build a reputation for resisting threats; similarly, Sechser argues that states facing a challenger (or who have faced others) who has a history of making threats will have greater incentives to build a reputation; and states who share a land border with a challenger will have greater incentives to build a reputation. Barbara Walter (2006) argued that governments that face many potential separatist groups, as proxied by the number of ethnic groups, will have greater incentives to build a reputation for resisting and punishing separatist demands. More generally we could create an extensive list of factors that increase incentives to build a reputation. Any factor that increases the value of the future, the number of expected future coercive opportunities (or threats), the stakes of future coercive disputes, or the generalizability or durability of reputation, will increase incentives to build a reputation. Some candidate factors include: living in a hostile neighborhood, having many actual or potential rivals, being a colonial power, being a great power, having one's conflict behavior observed, not having a well formed reputation, requiring support of allies, an increase in the value of territory, norms and culture amongst relevant leaders that make reputational inferences clear, important, and general, and so forth.

Examining such phenomena would be straightforward, but it would be impossible to persuasively overcome confounding. Suppose that after analyzing a number of such sources of variation, the predictions in chapter 2 are mostly or partially supported. There is almost certainly a set of plausible confounds that could bias against the predictions, and thus render the failure to find evidence in favor of the predictions inconsequential. Similarly, evidence in favor of the predictions could plausibly be explained away by possible confounding. The only kind of evidence based on this kind of design that I would find persuasive would be if all the associations were very strongly consistent with the predictions of chapter 2, or very strongly inconsistent, since then explaining away such results would require coming up with a plausible (and parsimonious) set of confounds that could generate or obscure the results in each domain.

There are many other possible extensions of this research including, to list those that I'm currently working on: survey experiments to study various aspects of reputation, theoretical and empirical examination of the extent to which reputations are leader- or state-specific, theoretical elaboration of the implications of reputation as an inference about type vs reputation as an inference about strategies, and elaboration of the relationship between reputation and status.

5.4   Concern for Reputation in International Relations

Leaders throughout history have been concerned—often obsessed—with how other leaders perceive them. Many scholars regard this concern for reputation to be a central cause of war. However, there has been little persuasive systematic evidence in support of this proposition, due in large part to obscuring selection effects and the challenges in drawing causal inferences from observational data. This dissertation seeks to overcome these challenges

through a number of innovations. I develop a family of formal models to deduce a set of predictions that are robust to selection dynamics and different modeling assumptions. I then test these predictions on two empirical domains. The first compares US conflict behavior under Southern presidents—who should care more about reputation—and non-Southern presidents. The second compares conflict behavior under leaders early in their time-in-office with those same leaders later in their time-in-office. Each of these empirical domains was selected because it involved a manipulation of concern for reputation that had plausibly relatively little unobserved bias.

Under both of these two distinct empirical domains I find results very consistent with my theory. Each prediction for which there was even modest statistical power was significantly supported in the analysis of US presidents, and a joint test of these three predictions provided support for my theory robust to a breadth of matching strategies and alternative coding rules. Similarly for the analysis of leaders' time-in-office, all five predictions were significantly associated in the manner expected by my theory, controlling for all cross-leader and cross-state factors. In both cases, lingering alternative explanations are evaluated and ruled out through critical tests. While even the best designed observational study can still succumb to unknown sources of bias, the fact that the rich set of formally derived predictions of my theory are upheld in two distinct carefully chosen empirical domains provides some of the most persuasive evidence to date about the causal effects of concern for reputation.

Effect sizes are large. Southern presidents are about twice as likely to use force in a dispute, and experience disputes twice as long and three times more likely to end in victory, as non-Southern presidents. Leaders ten years later in their time-in-office experience approximately 10% fewer fatalities and uses of force, shorter disputes, and about 250% more frequent violations of their security commitments. Further, these estimated associations only reflect the causal effects of two limited sources of variation in concern for reputation. Historians generally regard non-Southern US presidents to also be very concerned about reputation; the Southern - non-Southern comparison thus underestimates the full effects of concern for reputation. Similarly, the time-in-office analysis only identifies the small layer of concern for reputation that rationally varies and expresses itself with time-in-office. The effects of concern for reputation are underestimated to the extent that reputation is valued intrinsically, is not entirely discounted with increases in time-in-office, or that long-lived institutions such as parliaments, families, or the military incentivize or constrain the behavior of leaders. The full effects of concern for reputation, such as arises for leaders of major powers in hostile neighborhoods in eras when territory is valuable and cultures of honor are deep and prevalent, is plausibly vastly greater than the large effects estimated here.

Concern for reputation expands and amplifies the stakes of smaller disputes. Whereas the nominal material stakes in a dispute might be relatively small (e.g. Berlin or Vietnam during the Cold War), the perceived reputational stakes may be huge. Concern for reputation compresses the stakes of all disputes over which the reputation is relevant into every particular dispute, as if running fuses between each element in a collection of gunpowder

kegs, so that if one happens to detonate they all explode. This is counterproductive from the point of view of limiting the damage from any one militarized dispute, as suggested by the metaphor. It does, however, have strategic benefits. It increases the perceived costs to others from aggression on any issue that engages one's reputations. Like the invention of discrete territorial boundaries (Branch, 2011), it provides a means for articulating and constructing a sharp discontinuity of interest and commitment so as to clearly divide territorial- and issue-space. Without concern for reputation, the boundaries and terms of agreements would always be in flux, for they are drawn over a continuum of interest and power that shifts with shocks to power, economic interests, and political coalitions.[5] Similarly, concern for reputation provides a mechanism for overcoming Lord Palmerston's real-politick insight that nations have no permanent friends or allies, only permanent interests; reputation enables credible alliance commitments by tying one's permanent interest in the preservation of one's reputation to the particular terms of an alliance treaty.

Concern for coercive reputation in international relations, like concern for coercive reputation amongst individuals (Campbell, 1964), is likely increasing in the extent to which coercive reputation is beneficial to the holder. As the utility to military coercion diminishes (Liberman, 1996) due to changes in the nature of the economy (Mansfield and Pollins, 2003; Rosecrance, 1986), nationalism, the spread of liberal institutions, and the introduction of more reliable international coalitions that punish aggression and more clearly articulated and powerful international law (Huth, Croco, and Appel, 2011; Russett and Oneal, 2001), concern for coercive reputations will be of decreasing benefit and we might expect that it will diminish over time, as it already seems to have amongst the leaders of many small democratic developed countries. Such changes might operate amongst the great powers too; the US might not feel that a loss to China in Asia implies the same risk of falling dominoes as was perceived to exist when facing the Soviet Union. If so, then we can expect the world to become much more peaceful, as worries about status and reputation will be less extensively tied to the outcome of more minor disputes.

However, it remains imperative to improve our understanding of the nature and effects of concern for reputation. While interstate conflict is much less common in recent decades and concern for reputation seems to be at historic lows, the consequences of a major power war are potentially more severe than ever. Further, concern for reputation, prestige, and national honor will persist for the foreseeable future as an important factor shaping the decisions of policy elites. The stakes remain great for an improved understanding of how concern for reputation shapes the relations of states.

---

[5]The fact that reputational stakes seem to be greater for defenders of an issue in dispute provides to me the most plausible explanation for why deterrence is so much more effective than compellence, and why initiators of wars tend to lose them (Lebow, 2010, pp. 50-53).

Bibliography

Abreu, D., P. K. Dutfa, and L. Smith (1994). "The Folk Theorem for Repeated Games: A NEU Condition". In: *Econometrica* 62, pp. 939–948.

Abreu, Dilip (1981). "Survey of Repeated Games". In: *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*. Ed. by Oskar Morgenstern and Robert J. Aumann. Mannheim: Bibliographisches Institut, pp. 11–42.

Agresti, Alan (2010). *Analysis of Ordinal Categorical Data*. Hoboken, N.J.: John Wiley and Sons.

Aldrich, John, John Sullivan, and Eugene Borgida (1989). "Foreign Affairs and Issue Voting: Do Presidential Candidates 'Waltz Before A Blind Audience?'" In: *American Political Science Review* 83.1, pp. 123–41.

Almond, Gabriel Abraham (1950). *The American People and Foreign Policy*. 1st. New York: Harcourt.

American National Election Studies (2008). *The ANES 2008 Time Series Study [dataset]*. Stanford University and the University of Michigan [producers].

Angrist, Joshua David and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, N.J.: Princeton University Press.

— (2010). "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics". In: *Journal of Economic Perspectives* 24.2, pp. 3–30.

Axelrod, Robert (1984). *The Evolution of Cooperation*. New York: Basic Books.

Ayers, Edward L. (1989). "Honor". In: *Encyclopedia of Southern Culture*. Ed. by Charles Reagan Wilson and William R. Ferris. Chapel Hill: University of North Carolina Press, pp. 1483–1484.

Bak, Daehee and Glenn Palmer (2010). "Testing the Biden Hypotheses: Leader Tenure, Age, and International Conflict". In: *Foreign Policy Analysis* 6.3, pp. 257–273.

Bartley, Numan (1995). *The New South 1945–1980: The Story of the South's Modernization*. Baton Rouge: Louisiana State UP.

Bennett, D. Scott and Allan Stam III (2010). *Expected Utility and Data Management Program, Version 3.204 [computer program]*. http://www.eugenesoftware.org/.

Bensel, Richard (1984). *Sectionalism and American Political Development: 1880-1980*. Madison: University of Wisconsin Press.

Berman, Larry and Stephen R. Routh (2003). "Why the United States Fought in Vietnam". In: *Annual Review of Political Science* 6, pp. 181–204.

Branch, Jordan (2011). "Mapping the Sovereign State: Technology, Authority, and Systemic Change". In: *International Organization* 65.1, pp. 1–36.

Campbell, Donald T (1975). ""Degrees of Freedom" and the Case Study". In: *Comparative Political Studies* 8.2, pp. 178–193. DOI: 10.1177/001041407500800204.

Campbell, John K. (1964). *Honour, Family and Patronage: A Study of Institutions and Moral Values in a Greek Mountain Community.* Oxford: Clarendon.

Caro, Robert A. (2002). *Master of the Senate.* The Years of Lyndon Johnson. New York: Vintage Books.

Chester, Edward W. (1975). *Sectionalism, Politics, and American Diplomacy.* Metuchen, N.J.: Scarecrow Press.

Chiozza, Giacomo and Ajin Choi (2003). "Guess Who Did What: Political Leaders and the Management of Territorial Disputes, 1950-1990". In: *Journal of Conflict Resolution* 47.3, pp. 251–278.

Chiozza, Giacomo and H. E. Goemans (2003). "Peace through Insecurity: Tenure and International Conflict." In: *Journal of Conflict Resolution* 47.4, pp. 443–467.

Chiozza, Giacomo and Hein E. Goemans (2004). "International Conflict and the Tenure of Leaders: Is War Still 'Ex Post' Inefficient?" In: *American Journal of Political Science* 48.3, pp. 604–619.

Chiozza, Giacomo and Hein E Goemans (2011). *Leaders and International Conflict.* Cambridge University Press.

Clare, Joe and Vesna Danilovic (2010). "Multiple Audiences and Reputation Building in International Conflicts". In: *The Journal of Conflict Resolution* 54.6, pp. 860–882.

Cochran, W. G. and S. Paul Chambers (1965). "The Planning of Observational Studies of Human Populations". In: *Journal of the Royal Statistical Society. Series A (General)* 128.2, pp. 234–266.

Cohen, Dov and Angela Leung (2010). "Virtue and Virility: Governing with Honor and the Association or Dissociation between Martial Honor and Moral Character of U.S. Presidents, Legislators, and Justices". Unpublished manuscript.

Cohen, Dov J. and Richard E. Nisbett (1994). "Self-Protection and the Culture of Honor: Explaining Southern Homicide". In: *Personality and Social Psychology Bulletin* 20.5, pp. 551–67.

Cohen, Dov J., Richard Nisbett, Brian Bowdle, and Norbert Schwarz (1996). "Insult, Aggression, and the Southern Culture of Honor: An 'Experimental Ethnography'". In: *Journal of Personality and Social Psychology* 70.5, pp. 945–60.

Cooper, William J. (1983). *Liberty and Slavery: Southern Politics to 1860.* New York: Knopf.

Crescenzi, Mark, Andrew Enterline, and Stephen Long (Sept. 2008). "Bringing Cooperation Back In: A Dynamic Model of Interstate Interaction". English. In: *Conflict Management and Peace Science* 25.3, pp. 264–280. DOI: 10.1080/07388940802219059.

Crescenzi, Mark J.C. (2007). "Reputation and Interstate Conflict". In: *American Journal of Political Science* 51.2, pp. 382–396.

Crescenzi, Mark J.C., J.D. Kathman, and Stephen B. Long (2007). "Reputation, History, and War". In: *Journal of Peace Research* 44.6, p. 651.

Crescenzi, Mark J.C., J.D. Kathman, K.B. Kleinberg, and R.M. Wood (2012). "Reliability, Reputation, and Alliance Formation". In: *International Studies Quarterly.*

Dafoe, Allan (2009). "Disaggregating Reputation: Type and Commitment". Presented at the Annual Meeting of the American Political Science Association.

Dafoe, Allan and Devin Caughey (2010). "Honor and War: Using Southern Presidents to Identify Reputational Effects in International Conflict". In: Annual Meeting of the International Studies Association.

Dafoe, Allan and Paul K. Huth (2011). "To Whom Does Reputation Adhere? Observational and Survey Experimental Tests of Country-, Leader-, and Influence-Specific Reputation."

Dallek, Robert (2003). *An Unfinished Life.* Little, Brown and Company.

Danilovic, Vesna (2001). "Conceptual and Selection Bias Issues in Deterrence". In: *Journal of Conflict Resolution* 45.1, pp. 97–125.

Davis, Lawrence H. (1985). "No Chain Store Paradox". In: *Theory and Decision* 18.2, pp. 139–144.

Dobbs, Michael (2008). *One Minute to Midnight: Kennedy, Khrushchev, and Castro on the Brink of Nuclear War.* New York: Alfred A. Knopf.

Dolan, Thomas M (July 2010). "Demanding the Impossible: War, Bargaining and Honor". In: *American Political Science Association*, pp. 1–42.

Donelan, Michael (2007). *Honor in Foreign Policy: A History and Discussion.* New York: Palgrave Macmillan.

Downes, A. and T. Sechser (Apr. 2012). "Forthcoming."The Illusion of Democratic Credibility."" In: *International Organization.*

Dunning, Thad (Oct. 2010). "Design-Based Inference: Beyond the Pitfalls of Regression Analysis?" In: *Rethinking Social Inquiry.* Ed. by Henry E Brady and David Collier. Rowman & Littlefield Publishers.

Ellsberg, D. (1968). *The Theory and Practice of Blackmail.* Tech. rep. RAND.

Enterline, Andrew J. and Christopher J. Williams (2010). "Gird Our Loins? The Biden Prophecy, Presidential Experience and the Timing of Foreign Policy Crises".

Fearon, James D. (1992). "Threats to Use Force: Costly Signals and Bargaining in International Costs". PhD thesis. University of California, Berkeley.

— (1994a). "Domestic Political Audiences and the Escalation of International Disputes". In: *The American Political Science Review* 88.3, pp. 577–592.

— (1994b). "Signaling Versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model". In: *Journal of Conflict Resolution* 38.2, pp. 236–269.

— (2002). "Selection Effects and Deterrence". In: *International Interactions* 28.1, pp. 5–29.

Fischer, David Hackett (1989). *Albion's Seed: Four British Folkways in America.* America, A Cultural History. New York: Oxford University Press.

Fite, Gilbert C. (1991). *Richard B. Russell, Jr.: Senator From Georgia.* London: University of North Carolina Press.

Fordham, Benjamin (1998). "The Politics of Threat Perception and the Use of Force: A Political Economy Model of U.S. Uses of Force, 1949–1994". In: *International Studies Quarterly* 42.3, pp. 567–90.

Franklin, John Hope (1970). *The Militant South, 1800–1861.* Cambridge, MA: Belknap Press.

Freeman, Joanne B. (2001). *Affairs of Honor: National Politics in the New Republic.* New Haven: Yale University Press.

Fry, Joseph A. (2002). *Dixie Looks Abroad: The South and U.S. Foreign Relations, 1789-1973.* Baton Rouge: Louisiana State University Press.

— (2006). *Debating Vietnam: Fulbright, Stennis, and their Senate hearings.* Lanham: Rowman and Littlefield.

Fudenberg, Drew and David M Kreps (1987). "Reputation in the Simultaneous Play of Multiple Opponents". In: *The Review of Economic Studies* 54.4, pp. 541–568.

Fudenberg, Drew and Eric Maskin (1986). "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information". In: *Econometrica* 54, pp. 533–554.

Fudenberg, Drew and Jean Tirole (1995). *Game Theory.* Cambridge, MA: MIT Press.

Gartner, Scott Sigmund (2008). "The Multiple Effects of Casualties on Public Support for War: An Experimental Approach". In: *American Political Science Review* 102.1, pp. 95–106.

Gastil, Raymond D. (1971). "Homicide and a Regional Culture of Violence". In: *American Sociological Review* 36.3, pp. 412–427.

Gaubatz, Kurt T. (1991). "Election Cycles and War". In: *Journal of Conflict Resolution* 35.2, pp. 212–244.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (2003). *Bayesian Data Analysis.* Chapman and Hall.

Gelpi, C. and J.M. Grieco (2001). "Attracting trouble: Democracy, leadership tenure, and the targeting of militarized challenges, 1918-1992". In: *The Journal of Conflict Resolution* 45.6, pp. 794–817.

Gelpi, Christopher (1997). "Crime and Punishment: The Role of Norms in Crisis Bargaining". In: *American Political Science Review* 91.2, pp. 339–360.

Gelpi, Christopher and Peter D. Feaver (2002). "Speak Softly and Carry a Big Stick? Veterans in the Political Elite and the American Use of Force". In: *American Political Science Review* 96.4, pp. 779–93.

George, Alexander L. and Richard Smoke (1974). *Deterrence in American Foreign Policy: Theory and Practice.* New York: Columbia University Press.

Ghosn, Faten, Glenn Palmer, and Scott Bennett (2004). "The MID3 Data Set, 1993–2001: Procedures, Coding Rules, and Description". In: *Conflict Management and Peace Science* 21, pp. 133–54.

Goemans, H. E., Kristian S. Gleditsch, and Giacomo Chiozza (2009). "Introducing Archigos: A Data Set of Political Leaders". In: *Journal of Peace Research* 46.2, pp. 269–283.

Goldgeier, James M. (1994). *Leadership Style and Soviet Foreign Policy: Stalin, Khrushchev, Brezhnev, Gorbachev*. Johns Hopkins University Press.

— (1997). "Psychology and Security". In: *Security Studies* 6.4, pp. 137–66.

Good, Phillip (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*. 3rd. New York: Springer.

Greenstein, Fred I. (2000). *The Presidential Difference: Leadership Style from FDR to Clinton*. New York: Martin Kessler Books.

Greif, Avner (1989). "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders". In: *The Journal of Economic History* 49.4, pp. 857–882.

Guisinger, Alexandra and Alastair Smith (2002). "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises". In: *The Journal of Conflict Resolution* 46.2, pp. 175–200.

Guzman, Andrew T. (2008). *How International Law Works: a Rational Choice Theory*. Oxford, UK: Oxford University Press.

Harrington, David (2005). "Linear Rank Tests in Survival Analysis". In: *Encyclopedia of Biostatistics*. Ed. by Peter Armitage and Theodore Colton. 2nd. John Wiley and Sons.

Haynes, Kyle (Oct. 2012). "Lame Ducks and Coercive Diplomacy:" in: *The Journal of Conflict Resolution*.

Hermann, Margaret G. and Thomas W. Milburn (1977). *A Psychological Examination of Political Leaders*. Free Press.

Hero, Alfred O. (1965). *The Southerner and World Affairs*. Baton Rouge: Louisiana State University Press.

Herring, George C. (2008). *From Colony to Superpower: U.S. Foreign Relations Since 1776*. The Oxford History of the United States. New York: Oxford University Press.

Hobbes, Thomas and Richard Tuck (1996). *Leviathan*. Cambridge Texts in the History of Political Thought. Cambridge University Press.

Hopf, Ted (1994). *Peripheral Visions: Deterrence Theory and American Foreign Policy in the Third World, 1965-1990*. University of Michigan Press.

Horowitz, Michael, Rose McDermott, and Allan C. Stam (2005). "Leader Age, Regime Type, and Violent International Relations". In: *Journal of Conflict Resolution* 49, pp. 661–685.

Hospodor, Gregory S. (1999). "'Bound by All the Ties of Honor': Southern Honor, the Mississippians, and the Mexican War". In: *Journal of Mississippi History* 61.Spring, pp. 1–28.

Howell, William G. and Jon C. Pevehouse (2005). "Presidents, Congress, and the Use of Force". In: *International Organization* 59.1, pp. 209–232.

Hunt, Michael H. (1987). *Ideology and U.S. Foreign Policy*. New Haven: Yale UP.

Huth, Paul K. (1988). *Extended Deterrence and the Prevention of War*. Yale University Press.

— (1997). "Reputations and Deterrence: A Theoretical and Empirical Assessment". English. In: *Security Studies* 7.1, pp. 72–99.

Huth, Paul K., Sarah E. Croco, and Benjamin J. Appel (2011). "Does International Law Promote the Peaceful Settlement of International Disputes? Evidence from the Study of Territorial Conflicts Since 1945". In: *American Political Science Review* 105.2, pp. 415–436.

Huth, Paul K. and Bruce Russett (1984). "What Makes Deterrence Work: Cases from 1900 to 1980". In: *World Politics* 36.4, pp. 496–526.

— (1993). "General Deterrence Between Enduring Rivals: Testing Three Competing Models". In: *American Political Science Review* 87.1, pp. 61–73.

James, Patrick and John R. Oneal (1991). "The Influence of Domestic and International Politics on the President's Use of Force". In: *Journal of Conflict Resolution* 35.2, pp. 307–32.

Jervis, Robert (2002). "Signaling and Perception: Drawing Inferences and Projecting Images". In: *Political Psychology*. Ed. by Kristen R. Monroe. Lawrence Erlbaum Associates.

Jervis, Robert and Jack Snyder (1991). *Dominoes and Bandwagons: Strategic Beliefs and Great Power Competition in the Eurasian Rimland*. Oxford University Press.

Johnston, Alastair I. (1998). *Cultural Realism: Strategic Cculture and Grand Sstrategy in Chinese History*. Princeton University Press.

Johnston, Alastair Iain (1995). "Thinking about Strategic Culture". In: *International Security* 19.4, pp. 32–64.

Jones, Daniel M., Stuart A. Bremer, and J. David Singer (1996). "Militarized Interstate Disputes, 1816-1992: Rationale, Coding Rules, and Empirical Patterns". In: *Conflict Management and Peace Science* 15.2, pp. 163–213.

Joshi, Shashank (2008). "Honor in International Relations". Working Paper No. 2008-0146. Weatherhead Center for International Affairs, Harvard University.

Kagan, Donald (1995). *On the Origins of War and the Preservation of Peace*. New York: Doubleday.

Keohane, Robert (1984). *After Hegemony*. New York: Princeton University Press.

Kier, Elizabeth (1995). "Culture and Military Doctrine: France Between the Wars". In: *International Security* 19.4, pp. 65–93.

Kinder, Donald (2011). "Cambridge Handbook of Experimental Political Science". In: ed. by James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia. Cambridge University Press. Chap. Campbell's Ghost.

King, Gary, Robert Owen Keohane, and Sidney Verba (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.

Kissinger, Henry (1979). *The White House Years*. New York: Little, Brown, and Company.

Krause, Sharon R. (2002). *Liberalism with Honor*. Cambridge, Mass.: Harvard University Press.

Kreps, David M. and Robert Wilson (1982). "Reputation and Imperfect Information". In: *Journal of Economic Theory* 27, pp. 253–279.

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson (1982). "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma". In: *Journal of Economic Theory* 27, pp. 245–252.

Kydd, Andrew H. (2005). *Trust and Mistrust in International Relations.* Princeton University Press.

Larson, Deborah Welch (1985). *Origins of Containment: A Psychological Explanation.* Princeton: Princeton UP.

Lebow, Richard Ned (Oct. 2008). *A Cultural Theory of International Relations.* New York: Cambridge University Press.

— (Mar. 2010). *Why Nations Fight.* New York: Cambridge University Press.

Lebow, Richard Ned and Janice Gross Stein (1990). "Deterrence: The Elusive Dependent Variable". In: *World Politics* 42.3, pp. 336–369.

Lee, Matthew R., William B. Bankston, Timothy C. Hayes, and Shaun A. Thomas (2007). "Revisiting the Southern Culture of Violence". In: *The Sociological Quarterly* 48, pp. 253–275.

Leeds, Brett Ashley (2003). "Alliance Reliability in Times of War: Explaining State Decisions to Violate Treaties". In: *International Organization* 57, pp. 801–827.

Leeds, Brett Ashley, Andrew G. Long, and Sara McLaughlin Mitchell (2000). "Reevaluating Alliance Reliability: Specific Threats, Specific Promises". In: *Journal of Conflict Resolution* 44.5, pp. 686–699.

Leeds, Brett Ashley, Michaela Mattes, and Jeremy S. Vogel (2009). "Interests, Institutions, and the Reliability of International Commitments". In: *American Journal of Political Science* 53.2, pp. 461–476.

Leeds, Brett Ashley and Burcu Savun (2007). "Terminating Alliances: Why Do States Abrogate Agreements?" In: *Journal of Politics* 69.4, pp. 1118–1132.

Leeds, Brett Ashley, Jeffrey M. Ritter, Sara McLaughlin Mitchell, and Andrew G. Long (2002). "Alliance Treaty Obligations and Provisions, 1815-1944". In: *International Interactions* 28, pp. 237–260.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden-Day.

Leventoglu, Bahar and Ahmer Tarar (2005). "Prenegotiation public commitment in domestic and international bargaining". In: *The American Political Science Review* 99.3, pp. 419–433.

Liberman, Peter (1996). *Does Conquest Pay?* Princeton University.

Lind, Michael (2003). *Made in Texas: George W. Bush and the Southern Takeover of American Politics.* New York: Basic Books.

Logevall, Fredrik (1999). *Choosing War: The Lost Chance for Peace and the Escalation of War in Vietnam.* Berkeley: University of California Press.

Lupton, Danielle (Mar. 2012). "The Khrushchev Crises:Lessons in Leader-Based Reputations for Resolve". In: *Annual Meeting of the International Studies Association.*

Mansfield, Edward D. and Brian M. Pollins (2003). *Economic Interdependence and International Conflict.* Ann Arbor: The University of Michigan Press.

Markey, Daniel (1999). "Prestige and the Origins of War: Returning to Realism's Roots". In: *Security Studies* 8.4, pp. 126–72.

Marshall, Monty G. and Keith Jaggers (2002). *Polity IV Project: Political Regime Characteristics and Transitions.*

McElreath, Richard (2003). "Reputation and the Evolution of Conflict". In: *Journal of Theoretical Biology* 220.3, pp. 345–357.

McGillivray, Fiona and Alastair Smith (2005). "The Impact of Leadership Turnover and Domestic Institutions on International Cooperation". In: *Journal of Conflict Resolution* 49.5, p. 639.

— (2008a). "Credibility in Compliance and Punishment: Leader Specific Punishments and Credibility". In: *The Journal of Politics* 68.02, pp. 248–258.

— (2008b). *Punishing the Prince: A Theory of Interstate Relations, Political Institutions, and Leader Change.* Princeton, N.J.: Princeton University Press.

McGillivray, Fiona and Allan C. Stam (2004). "Political Institutions, Coercive Diplomacy, and the Duration of Economic Sanctions". In: *Journal of Conflict Resolution* 48.2, pp. 154–172.

McMahon, RJ (1991). "Credibility and World Power: Exploring the Psychological Dimension in Postwar American Diplomacy". In: *Diplomatic History.*

McWhiney, Grady (1988). *Cracker Culture: Celtic Ways in the Old South.* Tuscaloosa: University of Alabama Press.

Meernik, James (1994). "Presidential Decision Making and the Political Use of Military Force". In: *International Studies Quarterly* 38.1, pp. 121–38.

Mercer, Jonathan (1996). *Reputation and International Politics.* Cornell Studies in Security Affairs. Ithaca, New York: Cornell University Press.

Mickey, Robert W. (forthcoming). *Paths Out of Dixie: The Democratization of Authoritarian Enclaves in America's Deep South, 1944–1972.* Princeton: Princeton UP.

Milgrom, Paul and John Roberts (1982). "Predation, Reputation and Entry Deterrence". In: *Journal of Economic Theory* 27, pp. 280–312.

Miller, Gregory (2003). "Hypotheses on Reputation: Alliance Choices and the Shadow of the Past". In: *Security Studies* 12.3, pp. 40–78.

Morgan, Stephen L. and Christopher Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* Cambridge University Press.

Morgenthau, Hans Joachim (1948). *Politics Among Nations: The Struggle for Power and Peace.* New York: A. A. Knopf.

Moritz, Mark (2008). "A Critical Examination of Honor Cultures and Herding Societies in Africa". In: *African Studies Review* 51.2, pp. 99–117.

Morrow, James D. (1989). "Capabilities, uncertainty, and resolve: A limited information model of crisis bargaining". In: *American Journal of Political Science* 33.4, pp. 941–972.

Morrow, James D. (2000). "Alliances: Why Write Them Down?" In: *Annual Review of Political Science* 3, pp. 63–83.

— (2010). "Eight questions for A Cultural Theory of International Relations". In: *International Theory* 2.3, pp. 475–480.

Nalebuff, Barry (1991). "Rational Deterrence in an Imperfect World". In: *World Politics* 43, pp. 313–335.

Neyman, A (1999). "Cooperation in Repeated Games When the Number of Stages is not Commonly Known". In: *Econometrica* 67.1, pp. 45–64.

Nisbett, Richard E. and Dov Cohen (1996). *Culture of Honor: The Psychology of Violence in the South*. New directions in social psychology. Boulder, Colo.: Westview Press.

Notter, Harley (1965). *The Origins of the Foreign Policy of Woodrow Wilson*. New York: Russell & Russell.

O'Neill, Barry (1999). *Honor, Symbols, and War*. University of Michigan Press.

— (2006). "Nuclear Weapons and National Prestige". In: *Cowles Foundation Discussion Paper*.

Ostrom, Charles W. and Brian L. Job (1986). "The President and the Political Use of Force". In: *American Political Science Review* 80.2, pp. 541–66.

Park, Jong Hee (2010). "Structural Change in U.S. Presidents' Use of Force". In: *American Journal of Political Science* 54.3, pp. 766–82.

Partell, Peter J and Glenn Palmer (1999). "Audience Costs and Interstate Crises: An Empirical Assessment of Fearon's Model of Dispute Outcomes". In: *International Studies Quarterly* 43.2, pp. 389–405.

Pearl, Judea (2010). "On a Class of Bias-Amplifying Variables that Endanger Effect Estimates". In: *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*. Ed. by P. Grunwald and P. Spirtes. Corvallis, OR.: AUAI, pp. 417–424.

Peristiany, John G. (1966). *Honour and Shame: The Values of Mediterranean Society*. Chicago: University of Chicago Press.

Pesarin, Fortunato and Luigi Salmaso (2010). *Permutation Tests for Complex Data: Theory, Applications, and Software*. Wiley Series in Probability and Statistics. John Wiley and Sons.

Pinker, Steven (2011). *The Better Angels of our Nature*. New York: Viking.

Pitt-Rivers, Julian (1968). "Honor". In: *International Encyclopedia of the Social Sciences*. Ed. by David Sills. Vol. 6. New York: Macmillan, pp. 503–11.

Potter, Philip B. K. (2007). "Does Experience Matter? American Presidential Experience, Age, and International Conflict". In: *The Journal of Conflict Resolution* 51.3, pp. 351–378.

Press, Daryl G. (2005). *Calculating Credibility*. Ithaca, NY: Cornell University Press.

Quirk, Robert E. (1967). *An Affair of Honor: Woodrow Wilson and the Occupation of Veracruz*. New York: W.W. Norton and Co.

Radner, Roy (1980). "Collusive Behavior in Noncooperative Epsilon-Equilibria of Oligopolies with Long but Finite Lives". In: *Journal of Economic Theory* 22.2, pp. 136–154.

Risjord, Norman K. (1961). "1812: Conservatives, War Hawks, and the Nation's Honor". In: *William and Mary Quarterly* 18.3, pp. 196–210.

Robinson, Paul (July 2006). *Military Honour and the Conduct of War: From Ancient Greece to Iraq*. Routledge.

Rosecrance, Richard (1986). *The Rise of the Trading State: Commerce and Conquest in the Modern World*. Basic Books.

Rosenbaum, Paul R. (2002). "Covariance Adjustment in Randomized Experiments and Observational Studies". In: *Statistical Science* 17.3, pp. 286–304.

— (2010). *Design of Observational Studies*. Springer Series in Statistics. New York: Springer.

Russett, Bruce (1963). "The Calculus of Deterrence". In: *The Journal of Conflict Resolution*, pp. 1–14.

Russett, Bruce and John Oneal (2001). *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York: W. W. Norton and Company.

Sartori, Anne (2002). "The Might of the Pen: A Reputational Rheory of Communication in International Disputes". In: *International Organization* 56.1, pp. 121–149.

— (2005). "Deterrence by Diplomacy". In: *Princeton University Press*.

Schelling, Thomas C. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

— (1966). *Arms and Influence*. New Haven: Yale University Press.

Schultz, Kenneth A. (1998). "Domestic Opposition and Signaling in International Crises". In: *American Political Science Review* 92.4.

— (1999). "Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War". In: *International Organization* 53.2, pp. 233–266.

Sechser, Todd S. (2007). "Winning Without a Fight: Power, Reputation, and Compellent Threats in International Crises". PhD thesis. Stanford University.

— (Aug. 2010). "Goliath's Curse: Coercive Threats and Asymmetric Power". In: *International Organization* 64.4, pp. 627–660.

Sechser, Todd S. and Alexander B. Downes (2010). "Democracy, Threats, and Credibility: A Reassessment". Presented at the Annual Meeting of the American Political Science Association.

Sekhon, Jasjeet S. (2009). "Opiates for the Matches: Matching Methods for Causal Inference". In: *Annual Review of Political Science* 12.1, pp. 487–508.

— (2011). "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The `Matching` package for R". In: *Journal of Statistical Software* 42.7. http://sekhon.berkeley.edu/matching/, pp. 1–52.

Sekhon, Jasjeet S. and Rocío Titiunik (2012). "When Natural Experiments Are Neither Natural Nor Experiments". In: *American Political Science Review* 106.1, pp. 35–57.

Selten, Reinhard (1978). "The Chain Store Paradox". In: *Theory and Decision* 9.2, pp. 127–159.

Shackelford, Todd K. (2005). "An Evolutionary Psychological Perspective on Cultures of Honor". In: *Evolutionary Psychology* 3, pp. 381–391.

Simmons, Beth (2010). "Treaty Compliance and Violation". In: *Annual Review of Political Science* 13, pp. 273–296.

Small, Dylan S., Thomas R. Ten Have, and Paul R. Rosenbaum (2008). "Randomization Inference in a Group-Randomized Trial of Treatments for Depression". In: *Journal of the American Statistical Association* 103.481, pp. 271–79.

Snyder, Glenn Herald and Paul Diesing (1977). *Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises*. Princeton, N.J.: Princeton University Press.

Snyder, Jack and Erica D. Borghard (2011). "The Cost of Empty Threats: A Penny, Not a Pound". In: *American Political Science Review* 105.03, pp. 437–456.

Sorin, S. (1992). "Repeated Games With Complete Information". In: *Handbook of Game Theory*. Ed. by R. J. Aumann and S. Hart. Amsterdam: North-Holland. Chap. 4, pp. 71–107.

Stevenson, Dennis Matthew (2004). *War, Words, and the Southern Way: The Florida Acquisition and the Rhetoric of Southern Honor*. MA thesis, University of Florida, Gainsville.

Stewart, Frank Henderson (1994). *Honor*. Chicago: University of Chicago Press.

Tang, Shiping (2005). "Reputation, Cult of Reputation, and International Conflict". In: *Security Studies* 14.1, pp. 34–62.

Thucydides. *History of the Peloponnesian War*. New York: Penguin Books. Translated by Rex Warner.

— *The Peloponnesian War*. Translated by Richard Crawley.

Tingley, Dustin H. and Barbara F. Walter (Dec. 2011a). "Can Cheap Talk Deter?: An Experimental Analysis". English. In: *The Journal of Conflict Resolution* 55.6, pp. 996–1020.

— (2011b). "The Effect of Repeated Play on Reputation Building: An Experimental Approach". In: *International Organization* 65.2, pp. 343–365.

Tomz, Michael (2007a). "Domestic Audience Costs in International Relations: An Experimental Approach". In: *International Organization* 61.4, pp. 821–40.

— (2007b). *Reputation and International Cooperation: Sovereign Debt across Three Centuries*. Princeton, NJ: Princeton University Press.

— (2008). "Reputation and the Effect of International Law on Preferences and Beliefs". Stanford University. Unpublished manuscript. http://www.stanford.edu/~tomz/working/Tomz-IntlLaw-2008-02-11a.pdf.

Tomz, Michael and Jessica Lea Weeks (2012). "An Experimental Investigation of the Democratic Peace".

Van Evera, Stephen (1984). "The Cult of the Offensive and the Origins of the First World War". In: *International Security* 9.1, pp. 58–107.

Vandello, Joseph A., Dov Cohen, and Sean Ransom (2008). "U.S. Southern and Northern Differences in Perceptions of Norms About Aggression: Mechanisms for the Perpetuation of a Culture of Honor". In: *Journal of Cross-Cultural Psychology* 39.2, pp. 162–77.

Wagner, R. Harrison (1992). "Rationality and Misperception in Deterrence Theory". In: *Journal of Theoretical Politics* 4.2, pp. 115–141.

Walter, Barbara F. (2006). "Building Reputation: Why Governments Fight Some Separatists but not Others". In: *American Journal of Political Science* 50.2, pp. 313–330.

— (2009). *Reputation and Civil War: Why Separatist Conflicts Are So Violent.* New York: Cambridge University Press.

Weeks, Jessica Lea (2008). "Autocratic Audience Costs: Regime Type and Signaling Resolve". In: *International Organization* 62.1, pp. 35–64.

Weeks, Jessica Lea and Dara Kay Cohen (2009). "Red Herrings? Fishing Trawler Disputes, Regime Types, and Interstate Conflict". Paper presented at the International Studies Association annual convention.

Wen, Quan (1994). "The 'Folk Theorem' for Repeated Games with Complete Information". In: *Econometrica* 62.4, pp. 949–954.

Wheelan, Joseph (2003). *Jefferson's War: America's First War on Terror, 1801–1805.* New York: Carroll and Graf.

Whitfield, John (2011). *People Will Talk: The Surprising Science of Reputation.* John Wiley & Sons. ISBN: 9781118114643.

Wittkopf, Eugene R. (1990). *Faces of Internationalism: Public Opinion and American Foreign Policy.* Durham, NC: Duke UP.

Wolf, Reinhard (Feb. 2011). "Respect and Disrespect in International Politics: the Significance of Sstatus Recognition". English. In: *International Theory* 3.1, pp. 105–142. DOI: [10.1017/S1752971910000308](10.1017/S1752971910000308).

Wolford, Scott (2007). "The Turnover Trap: New Leaders, Reputation, and International Conflict". In: *American Journal of Political Science* 51.4, pp. 772–788.

Woolley, John T. and Gerhard Peters (2004). *George W. Bush: Address Before a Joint Session of the Congress on the State of the Union, January 20, 2004. The American Presidency Project* [online], Santa Barbara, CA, http://www.presidency.ucsb.edu/ws/index.php?pid=29646#axzz1NgV3pM8l, accessed May 28, 2011.

Wright, Gavin (1986). *Old South, New South: Revolutions in the Southern Economy Since the Civil War.* New York: Basic Books.

Wyatt-Brown, Bertram (1982). *Southern Honor: Ethics and Behavior in the Old South.* New York: Oxford University Press.

— (2005). "The Ethic of Honor in National Crises: The Civil War, Vietnam, Iraq, and the Southern Factor". In: *Journal of The Historical Society* 5.4, pp. 431–60.

Appendix A

A.1 Appendix for Chapter 4

The appendix contains: Two tables of leaders who fit the empirical predictions (p. 110 and 111); a table of leaders who don't fit the empirical predictions (p. 112)

The list of leaders who fit the empirical predictions may be inspected to see if any other explanations suggest themselves for the behavior common to these leaders, or whether the reputational theory has any particular resonance with any them. Similarly, the list of leaders who don't fit the empirical predictions may be examined to see if an alternative explanation suggests itself, or if perhaps some scope conditions of the reputational theory may be present in these cases.

The list of leaders who fit the empirical predictions may be inspected to see if any other explanations suggest themselves for the behavior common to these leaders, or whether the reputational theory has any particular resonance with any them. Similarly, the list of leaders who don't fit the empirical predictions may be examined to see if an alternative explanation suggests itself, or if perhaps some scope conditions of the reputational theory may be present in these cases.

Table A.1: **List of (regular entry) leaders who best fit the theory, sorted by combined rank.**

Rank	M-Rank	Leader Name	Country	Dates	Manner of Exit	Force R	Dur. R
.098	.056	Sadat	Egypt	1970-1981	Irregular	24	7
.111	.046	Hassan II	Morocco	1961-1991	Natural Death	22	64
.116	.064	Stalin	Russia	1923-1952	Natural Death	21	875
.179	.121	Estrada	Philippines	1998-2000	Regular	59	36
.184	.12	Carter	United States	1977-1980	Regular	71	966
.216	.109	Gorbachev	Russia	1985-1987	Irregular	120	837
.218	.097	Marcos	Philippines	1965-1985	Irregular	135	2
.223	.158	Hussein Ibn Talal El-Hashim	Jordan	1952-1991	Natural Death	27	1069
.234	.129	Brezhnev	Russia	1964-1982	Natural Death	117	1112
.272	.038	Kim Il-Sung	North Korea	1948-1995	Natural Death	5	0
.278	.039	Chiang Kai-shek	Taiwan	1950-1974	Natural Death	2	4
.298	.165	Rahman	Malaysia	1959-1968	Regular	148	903
.304	.143	Menzies	Australia	1949-1972	Regular	180	956
.311	.157	Gandhi, R.	India	1984-1989	Regular	100	30
.325	.075	Deng Xiaoping	China	1980-1997	Natural Death	4	1105
.341	.172	De Gaulle	France	1958-1964	Regular	188	248
.343	.199	Bhutto	Pakistan	1971-1974	Irregular	161	387
.351	.153	Le Duan	Vietnam	1969-1986	Natural Death	221	1077
.376	.143	Kim Jong-Il	North Korea	1994-2001	Still in office	114	35
.389	.2	Chaudhri M. Ali	Pakistan	1955-1956	Regular	211	270
.401	.149	Hua Guofeng	China	1976-1979	Regular	105	225
.416	.209	Obuchi	Japan	1998-1999	Natural Death	231	65
.421	.275	Sharif	Pakistan	1990-1991	Irregular	163	859
.424	.268	Gandhi, I.	India	1980-1984	Irregular	124	1088
.43	.101	Saddam Hussein	Iraq	1979-2001	Foreign	0	1116
.437	.094	Erbakan	Turkey	1996-1997	Irregular	14	257
.438	.218	Chun Doo Hwan	South Korea	1980-1987	Regular	149	302

Table A.2: **List of (regular entry) leaders who best fit the theory, sorted by modified combined rank.**

Rank	M-Rank	Leader Name	Country	Dates	Manner of Exit	Force R	Dur. R	Fat. R
.99	.019	Eisenhower	United States	1953-1961	Regular	11	32	24
.523	.023	Alexander II	Russia	1855-1880	Irregular	8	62	141
.622	.023	Ferdinand I	Romania	1914-1926	Natural Death	10	161	43
.681	.034	Alexander	Yugoslavia	1918-1934	Irregular	34	22	26
.535	.035	Ben Gurion	Israel	1955-1963	Regular	28	24	48
.272	.038	Kim Il-Sung	North Korea	1948-1995	Natural Death	5	0	2
.278	.039	Chiang Kai-shek	Taiwan	1950-1974	Natural Death	2	4	63
.54	.04	Shevardnadze	Georgia	1992-2001	Irregular	37	42	60
.723	.042	Masaryk	Czechoslovakia	1918-1934	Irregular	35	107	34
.111	.046	Hassan II	Morocco	1961-1991	Natural Death	22	64	142
.555	.055	Pedro II	Brazil	1840-1875	Irregular	46	77	93
.098	.056	Sadat	Egypt	1970-1981	Irregular	24	7	14
.557	.057	H. Aliyev	Azerbaijan	1993-2001		20	28	79
1.031	.063	Johnson	United States	1963-1968	Regular	32	17	7
.116	.064	Stalin	Russia	1923-1952	Natural Death	21	875	13
.569	.069	Francis Joseph I	Austria-Hungary	1848-1918	Natural Death	7	1052	1
1.063	.07	Hitler	Germany	1933-1945		15	944	5
1.065	.074	Mussolini	Italy	1922-1943	Irregular	1106	1113	99
.325	.075	Deng Xiaoping	China	1980-1997	Natural Death	4	1105	3
.658	.079	Yeltsin	Russia	1991-2001	Regular	1	163	38
.579	.079	Wilhelm II	Germany	1888-1918	Irregular	13	1109	18
.583	.083	Houphouet-Boigny	Cote d'Ivoire	1960-1993	Natural Death	83	20	109
.531	.087	Mengistu Marriam	Ethiopia	1977-1988	Irregular	85	16	94
1.082	.09	Al-Assad H.	Syria	1971-1998	Natural Death	25	19	56
.59	.09	Tudjman	Croatia	1990-1997	Natural Death	49	201	20
.437	.094	Erbakan	Turkey	1996-1997	Irregular	14	257	558.5
1.06	.095	Ford	United States	1974-1976	Regular	48	146	171

Table A.3: **List of (regular entry) leaders who had the worst fit with the theory, sorted by combined rank.**

Rank	Leader Name	Country	Dates	Manner of Exit	Force R	Dur. R	Fat. R	Win R
2.145	Reagan	United States	1981-1989	Regular	1110	1095	1036	1023
2.113	Atlee	United Kingdom	1945-1952	Regular	1105	1107	1057	1100
2.002	Khrushchev	Russia	1953-1965	Regular	1114	60	6	52
1.994	Chamberlin	United Kingdom	1937-1940	Regular	1108	932	126	98
1.977	Bush	United States	1989-1993	Regular	993	1053	57	1015
1.955	Plaek Pibulsongkram	Thailand	1938-1943	Regular	1012	920	68	67
1.926	Heath	United Kingdom	1970-1973	Regular	1019	927	82	558.5
1.917	Eden	United Kingdom	1955-1956		901	1044	1061	997
1.895	Wilson	United Kingdom	1964-1969	Regular	971	1100	558.5	84
1.878	MacMillan	United Kingdom	1957-1963	Regular	918	1093	558.5	558.5
1.872	Nixon	United States	1969-1973	Regular	967	29	80	558.5
1.866	Daladier	France	1938-1940	Regular	939	907	1033	1024
1.85	Wilson	United Kingdom	1974-1976	Regular	955	172	558.5	558.5
1.824	Caetano	Portugal	1968-1973	Irregular	1092	10	967	558.5
1.8	Kennedy	United States	1961-1963	Irregular	903	151	178	25
1.799	Salazar	Portugal	1932-1967	Natural Death	1103	866	16	42
1.78	Petain	France	1940-1941	Regular	889	823	937	558.5
1.77	Gomulka	Poland	1956-1968	Irregular	1009	928	558.5	1068
1.743	Adenauer	Germany	1949-1963	Regular	956	969	947	558.5
1.717	Chirac	France	1995-2001	Still in office	885	178	983	991
1.681	Guggiari	Paraguay	1928-1935	Regular	1098	1111	1111	1107
1.663	Lubbers	Netherlands	1982-1994	Regular	1086	1090	999	1111
1.658	Jabir As-Sabah	Kuwait	1978-1990	Foreign	1073	1101	1105	1106
1.655	Konoe, F.	Japan	1937-1940	Regular	1077	1075	1030	1094
1.654	Pelloux	Italy	1898-1900	Regular	1084	933	1110	1042

A.2 Ability to Detect Reputational Inferences

[1] Let $\Delta\phi(\mathbf{h_1}, \mathbf{h_2}; \mathbf{Z_1}, \mathbf{Z_2})$ represent the change in the belief (ϕ) about the credibility of some particular threat under the counterfactual comparison of two (vectors of) different histories $\mathbf{h_1}$ and $\mathbf{h_2}$ (each of length k), and given the (vectors of) two contexts for the dispute and its counterfactual $\mathbf{Z_1}$ and $\mathbf{Z_2}$ (each of length l). Scholars who wish to argue that reputational inferences don't matter are effectively saying that any two sufficiently different histories of behavior have little (or no) effect on inferences about credibility, $H_A : \forall(\mathbf{h_1}, \mathbf{h_2}, \mathbf{Z_1}, \mathbf{Z_2}) \in (\mathbb{R}^k, \mathbb{R}^k, \mathbb{R}^l, \mathbb{R}^l)$ such that $D_h(\mathbf{h_1}, \mathbf{h_2}) > \epsilon_h$, and $D_Z(\mathbf{Z_1}, \mathbf{Z_2}) < \epsilon_Z$ then $|\Delta\phi(\mathbf{h_1}, \mathbf{h_2}; \mathbf{Z_1}, \mathbf{Z_2})| < |w|$. $D_j(\cdot, \cdot)$ is a function that gives a scalar summary of the difference in the vectors of ($j = h$) histories or ($j = Z$) contexts. In words, this alternative hypothesis says that, holding other factors relatively constant, beliefs about credibility will not change by a substantial amount for a sufficiently large change in the history of behavior. The null hypothesis (H_0), then, is that history of behavior has a sufficiently large effect that it "does matter": $\exists(\mathbf{h_1}, \mathbf{h_2}, \mathbf{Z_1}, \mathbf{Z_2}) \in (\mathbb{R}^k, \mathbb{R}^k, \mathbb{R}^l, \mathbb{R}^l)$ such that $D_h(\mathbf{h_1}, \mathbf{h_2}) > \epsilon_h$, and $D_Z(\mathbf{Z_1}, \mathbf{Z_2}) < \epsilon_Z$ then $|\Delta\phi(\mathbf{h_1}, \mathbf{h_2}; \mathbf{Z_1}, \mathbf{Z_2})| \geq |w|$.[2] w is the least amount that beliefs about credibility would have to change for any given threat that would have a substantial effect on the welfare of the leader or the state. Implicit to any tests of these two hypotheses, then, is that measurements of $\Delta\phi$ smaller than $|w|$ in size must be feasible. Otherwise, the design would lack the measurement power to reject the null hypothesis (H_0) that reputation does matter.

As argued above, our ability to measure ϕ (the belief about the probability a threat is credible), and hence also changes in ϕ, is probably very coarse. Our measurements of the beliefs of observers about the credibility of threats (ϕ) are not that precise, surely not better than with a 5 percentage point standard error. If they were, then historians would employ a more precise vocabulary than ordinary language (namely numbers or a standardized terminology) to communicate these different levels of beliefs.

How large is w? In section 5.2.2 I argued that it is plausibly less than 5%. This section will now attempt to estimate w more explicitly. Let W be the minimum welfare relevant change in beliefs over all threats; that is, W is the sum of the change in beliefs of the credibility of threats that would just be relevant to the welfare considerations of leaders or a state.[3] Let k be the number of compellent and deterrent threats taking place during a leader's tenure. Again for convenience, I assume that reputations don't decay over leadership tenure, but then reset completely with leader turnover. This implies that the minimum welfare relevant change in beliefs of credibility of a given threat is the total relevant change divided by the

---

[1]The following text slightly repeats section 5.2.2

[2]We could soften these hypotheses by looking at the average change in $\Delta\phi$ for a "reasonable" set of $(\mathbf{h_1}, \mathbf{h_2}, \mathbf{Z_1}, \mathbf{Z_2})$. The basic point remains.

[3]For this analysis I conservatively assume that utility is linear in changes in beliefs about credibility for different threats. The more plausible assumption that changes have diminishing returns would reinforce my argument.

number of expected threats: $w = W/k$.[4]

How large is k? We can think of k as the number of compellent and deterrent threats taking place during a leader's tenure (of a given importance). Compellent threats are observable but relatively rare. Deterrent threats, however, are not observable when they are not challenged, and most are probably rarely challenged. Every territorial border that would be contested if the defender lacked sufficient capability to defend it represents a deterrent threat. We can roughly estimate the number of these unobserved deterrent threats by considering how many militarized interstate disputes (MIDs) occur during the tenure of a typical leader. MIDs may exaggerate the number of threats if some MIDs represent unimportant and low level uses of force (Downes and Sechser, 2012). However, most MIDs probably reflect an ongoing instance where a deterrent threat is at work; for example, when a small group of soldiers cross a border and engage in an unauthorized exchange of fire, this reveals that there is a latent deterrent threat over the border requiring a military presence. MIDs more likely under-estimate the number of deterrent threats since they don't count all those deterrent threats that are never tested. For example, during the Cold War the US and Soviet Union each had an implicit deterrent threat protecting every country in its sphere of influence; only a handful of these escalated into a MID. Looking after 1950 when the MID data is most reliable, we see that the U.S. and the U.K. experienced a mean of 13 MIDs per leader (median=10). Russia (and the Soviet Union) and China experienced a mean of 30 MIDs per leader (median=11). Thus, let's estimate k to be conservatively around 10 for great powers during 1950-2000, though it could conceivably be as high as 50, and at more conflictual periods of history it could be much greater.[5]

Let's estimate W very conservatively at 70%.[6] That is, we are assuming that only actions that can change the credibility of a threat by 70 percentage points or more[7] is "welfare relevant". This is surely an exaggeration—surely it would have profound welfare implications for a state if it could transform a highly implausible threat into a highly plausible one—and thus is a conservative assumption for the purposes of this argument. Then, with $k = 10$, we would need to be able to detect effect sizes of at least $w = 7\%$ in the beliefs of observers about any given threat in order to argue that welfare relevant changes in reputational beliefs don't take place for great powers. As argued above, $w < 10\%$ is likely too small.

Another means of crudely calibrating the necessary effect sizes is to use data on the effects of military capabilities on public support for the use force. Tomz and Weeks (2012) report that 3.5% fewer U.K. respondents are willing to support a preemptive attack on a soon-to-be nuclear state if that state is "as strong", as opposed to "half as strong", as the U.K. We

---

[4]If we had diminishing marginal utility to increases in beliefs of credibility, then $w < W/k$.

[5]For example, Israel has a mean and median of 10 MIDs per leader. Israel is not a major power, but it is involved in a number of ongoing disputes and rivalries with its neighbors.

[6]Note that this is a huge effect. If we started from a baseline credibility of 10%, then the effect would be a 700% relative increase in credibility

[7]Or can increase the credibility of many threats, so that the change in probabilities sums to greater than 70%

could crudely transform that shift in public opinion into a shift in credibility by doubling or quadrupling it (to 7% or 14%).A justification for this transformation is as follows. Suppose a threat is credible if more than a certain proportion, α, of the population supports the threat. Suppose that α is somewhere between $0 < a$ and $b < 1$ (and $a < b$). Suppose also that baseline support for the threat is drawn from a uniform distribution between a and b: $s \ U(a, b)$. Then, an increase in support for the threat by z translates into an increase in the credibility of the threat by $z/(b-a)$. If we suppose that typical threats have support somewhere in the range $a = 0.25, b = 0.75$, then a change in 3.5% support for the threat increases the credibility of the threat by 7%.[8] If the effect on credibility of a doubling in military capabilities is understood as welfare relevant, than an improvement in one's reputation that achieves a similar magnitude of benefit ($w = 7\%$) is also welfare relevant. And as argued above, an $w = 7\%$ is likely below the threshold that can be reliably measured using historical methods.

---

[8]The closer the baseline support is likely to be to the necessary threshold, the greater the effect of any given change in public support.