

**Rationalist Causes of War:
Mechanisms, Experiments, and East Asian Wars**

by

Ch-yuan Kaiy Quek

B.A., Economics and Government, Cornell University (2005)

Submitted to the Department of Political Science
in Partial Fulfillment of the Requirements for the Degree of

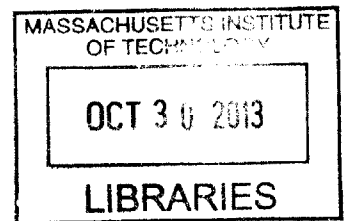
Doctor of Philosophy in Political Science

at the

Massachusetts Institute of Technology

September 2013

ARCHIVES



© 2013 Ch-yuan Kaiy Quek. All Rights Reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly
paper and electronic copies of this thesis document in whole or in part
in any medium now known or hereafter created.

Signature of Author: _____
Department of Political Science
July 10, 2013

Certified By: _____
Kenneth A. Oye
Associate Professor of Political Science
Thesis Supervisor

Accepted By: _____
Roger Petersen
Arthur and Ruth Sloan Professor of Political Science
Chairman, Graduate Program Committee

**Rationalist Causes of War:
Mechanisms, Experiments, and East Asian Wars**

by

Ch-yuan Kaiy Quek

Submitted to the Department of Political Science on July 10, 2013
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Political Science

Abstract

This dissertation specifies and tests rationalist mechanisms of war. Why would rational states fight each other despite their incentives for peaceful bargains that would avoid the costs of war? In the rationalist theory of war, private information and the commitment problem are the key causes of war. I study the effects of these factors – and the mechanisms regulating their effects – through randomized experiments, historical analysis of the decision processes in three wars, and a comparative study of all international wars fought in East Asia in the last century.

This is the first integrated study of rationalist causes of war that combines randomized experiments with historical cases. Despite a wide theoretical literature, there are few empirical tests of rationalist explanations for war. I use experimental and historical evidence to show that the commitment problem has strong positive effects on conflict. The effects of private information are less clear. Next, I specify six mechanisms that regulate the effects of the commitment problem and the private-information problem: three mechanisms (exogenous, endogenous, and inadvertent enforcement) for the first problem and three mechanisms (signaling with sunk cost, implementation cost, and salient contradiction) for the second. The experimental and historical evidence largely converge. Each of the three enforcement mechanisms calms the commitment problem and reduces the risk of conflict. Evidence for the three signaling mechanisms is mixed. Finally, I use the case universe of East Asian wars to assess the relevance of the mechanisms, suggest theoretical refinements, and infer alternative theories of war.

Thesis Supervisor: Kenneth A. Oye

Title: Associate Professor of Political Science

Acknowledgements

This thesis would have been impossible without the three members of my committee: Kenneth Oye, Stephen Van Evera, and James M. Snyder.

They have shaped my work profoundly. I am thankful to them for teaching me so much, and I leave MIT knowing that I still have much more to learn from them. They have the sharpest minds and the biggest hearts – a truly beautiful combination. Their dedication touched me on several occasions. (How would you feel, for instance, if you see your advisor crossing the road while reading your work and not watching out for traffic?) Numerous anecdotes come to mind and melt me as I write. I will remember them always.

This thesis has also benefited from the thoughts of many others who took time to digest the different pieces. For Prospectus: I thank Sandeep Baliga for an inspiring conversation, and Amartya Sen for reading my two-pager and making me think deeply about the nature of rationality. For Chapter 1: I thank Dan Altman, David Jae and Nick Miller for reading a summary draft and offering advice. For Chapter 2: I thank Adam Berinsky, Nathan Black, Ernst Fehr, Phil Haun, Gabriel Lenz, Randall Lewis, Alvin Roth, Mike Sances, David Singer, Dustin Tingley, anonymous reviewers, and seminar participants at MIT, APSA, ISA and MPSA. Chris Butler, Daina Chiba, Anthony Fowler, Nehemia Geva, Kentaro Hirose, Meredith Sarkees, Gerald Schneider, Josh Shiffrin, David Weinberg and Xu Yiqing also gave useful comments. For Chapter 3: I thank Dan Altman, Mark Bell, Nathan Black, Chris Clary, He Yinan, David Jae and Wayne Tan. The experiments in Chapter 4 were presented separately to different audiences: I thank Dan Altman, Mark Bell, Dave Clark, Andrew Coe, Eric Grynaviski, Chad Hazlitt, Hou Yue, Mike Sances, Dustin Tingley, Joshua Wu and seminar participants at Harvard, MIT, ISA and MPSA for their thoughts. MIT's Department of Political Science, Center for International Studies, Political Experiments Research Lab and Behavioral Research Lab, as well as Harvard's Decision Science Lab, facilitated my experiments.

Experimental design is both science and art. I was fortunate to be in Cambridge at a time when both Ernst Fehr and Al Roth were present. The evening when Ernst went over my design and shut down each and every source of noise was an evening of enlightenment. The engineering impulses that animated Al Roth's experimental mind were beautiful to behold – and worth the frostbite of ten winter morning walks to Harvard.

Suzanne Berger, Roger Petersen, Daniel Posner, Richard Samuels, David Andrew Singer, Edward Steinfeld, Lily Tsai and Kathleen Thelen offered wisdom at key junctures in my MIT career. Susan Twarog and Diana Gallagher made the ship sail smoothly. My friends above, and M. S. Kyoung, B. K. Song, Maeda Kentaro, Eitan Hersh, Jonas Nahm, Miranda Priebe, LFRZ and many others, made a hard MIT softer, and full of flavor and cheer. And I am who I am because of the nurturing kindness of my grandparents; the constant love of my parents, James Quek and Mary Chng; and the personification of nonsense and joy that is my younger brother QB. I am grateful to them all.

Kai Quek

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction	7
Chapter 2. Rationalist Explanations for War: Experimental and Historical Evidence	46
Chapter 3. Decisions, Processes and War: Evidence from the Sino-Japanese Wars	101
Chapter 4. Enforcement, Signaling and War: Evidence from Ten Experiments	188
Chapter 5. Wars in East Asia: Assessing Old Hypotheses, Inferring New Hypotheses	277
Chapter 6. Conclusion	348
Appendices A1-A5 (Experimental Protocols and Instructions)	358
Appendices B1-B2 (Supplementary Tables)	438

List of Figures

- 2.1 Crisis Bargaining Game
- 2.2 Incidence of War in Rounds 5-6
- 2.3 Percentage of Reneged Offers (No Enforcement)
- 2.4 Percentage of Wars in Stage 1 (No Enforcement)
- 2.5 Incidence of War (By Enforcement Conditions)
- 2.6 Incidence of War in Rounds 1-10
- 2.7 Incidence of War (By Information Conditions)
- 2.8 Incidence of War (Rounds 11-15)

- 4.1 Crisis Bargaining Game with Inadvertent Enforcement
- 4.2 Incidence of War Across Experimental Conditions (Rounds 1-5)
- 4.3 Incidence of War in Control Group (Rounds 1-8)
- 4.4 Incidence of War in Treatment Group (Rounds 1-8)
- 4.5 Average Final Offer in Treatment Group (Rounds 1-8)
- 4.6 Crisis Bargaining Game with Endogenous Enforcement
- 4.7 Incidence of War (With and Without Endogenous Enforcement)
- 4.8 Sunk-Cost Signaling Game
- 4.9 Proportion of Signalers Sending the Costly Threat (Experiments 3 and 4)
- 4.10 Proportion of Receivers Who Stayed Out (Experiments 3 and 4)
- 4.11 Average Credibility Scores (Experiments 3 and 4)
- 4.12 (A) Proportion of Believers and (B) Average Credibility Scores in Experiment 6
- 4.13 (A) Proportion of Believers and (B) Average Credibility Scores in Experiment 7
- 4.14 (A) Proportion of Believers and (B) Average Credibility Scores in Experiment 8 (First Test)
- 4.15 (A) Proportion of Believers and (B) Average Credibility Scores in Experiment 8 (Second Test)
- 4.16 (A) Proportion of Believers and (B) Average Credibility Scores in Experiment 8 (Third Test)
- 4.17 Distribution of Believers Across Four Korean-War Counterfactuals
- 4.18 Average Credibility Scores Across Four Korean-War Counterfactuals
- 4.19 (A) Proportion of Believers and (B) Average Credibility Scores Under Korean-War Counterfactual 0 (“Acheson/Withdrawal”)
- 4.20 (A) Proportion of Believers and (B) Average Credibility Scores Under Korean-War Counterfactual 1 (“No Acheson/No Withdrawal”)

- 5.1 War in East Asia, 1900-2000
- 5.2 Comparative Intensity of War (Battle Fatalities)
- 5.3 Comparative Intensity of War (War Duration)
- 5.4 Relative Power Between War Parties (Based on CINC Scores)

List of Tables

- 1.1 Overview of Key Empirical Results
- 1.2 International Wars in East Asia, 1900-2000

- 2.1 Experimental Design
- 2.2 Incidence of War Across Conditions
- 2.3 Average Number of Wars Per Player
- 2.4 Logit Estimates of Determinants for the Decision for War (Rounds 1-10)
- 2.5 Incidence of War (Rounds 11-15)

- 3.1 Predictions vs. Evidence from the Third Sino-Japanese War
- 3.2 Predictions vs. Evidence from the Second Sino-Japanese War

- 4.1 Summary of Experimental Findings
- 4.2 Player Incidence of War
- 4.3 Incidence of Endogenous Enforcement
- 4.4 Incidence of War Outcomes (With and Without Endogenous Enforcement)
- 4.5 Summary of Experiments
- 4.6 Logit Estimates – Determinants of Costly Threat
- 4.7 Logit Estimates – Determinants of Staying Out
- 4.8 Ordered Logit Estimates – Determinants of Credibility Score
- 4.9 Average Credibility Scores in Experiment 5
- 4.10 Logit and Ordered Logit Estimates – Effect of Costly Threat

- 6.1 Review of Key Empirical Results

- B1 Logit Estimates of Determinants for the Decision for War: Rounds 11-15
- B2 Logit Estimates of Determinants for War: Rounds 1-5
- B3 Logit Estimates of Determinants for War Decision in Stage 1: Rounds 1-5
- B4 Logit Estimates of Determinants for War in the Control Group
- B5 Logit Estimates of Determinants for War in the Treatment Group
- B6 OLS Estimates of Determinants for Final Offer Size: Treatment Group
- B7 Logit Estimates of Determinants for War: Experiment 2
- B8 Logit Estimates of Determinants for War Decision in Stage 1: Experiment 2

Chapter 1

Introduction

The problem of conflict and cooperation is arguably the central problem in social science. Of all conflict, the stakes are highest in international wars. In the 20th century, more than 30 million combatants – and countless civilians – were killed in these wars. Given the costs involved, one might expect states to rationally avoid going to war with one another. Yet wars recur in history. Their recurrence suggests two logical possibilities that can be cast into two separate questions. First, if wars are irrational, what prevent states from choosing rationally? Second, if wars are *not* irrational, then what are the conditions under which war *is* a rational choice? The two questions encompass the theoretical essence behind the causes and prevention of war. This thesis focuses on the second question because it is narrower but no less important than the first.

Why do rational states fight despite their incentives for a peaceful deal that can avoid the costs of war? In the rationalist theory of war, asymmetric information and the commitment problem are the key causes of war.¹ First, asymmetric information may foster war-causing miscalculations. This occurs when states have private information and the incentive to misrepresent. Second, a commitment problem may obstruct a peaceful bargain. This happens when two sides cannot commit to an agreement in the present because one side has an incentive to break the agreement in the future.

¹ James Fearon, "Rationalist Explanations for War," *International Organization* 49, No. 3 (1995): 379-414.

I study the effects of these two factors, and I specify and test the mechanisms regulating their effects. Because the rationalist explanations are very general, by themselves they offer limited specific guidance on how to prevent wars. How can we prevent war-causing miscalculations, when asymmetric information *always* exists in international politics? What specific mechanisms help to shut down the commitment problem? To answer these questions, we need greater specification in the rationalist theory of war. My theoretical goal is to break down rationalist explanations for war into specific mechanisms with direct implications for the prevention of war. My empirical goal is to study their effects. I do this through a series of experiments, case studies of the decision process in three wars, and a comparative analysis of all international wars fought in East Asia since 1900.

This is the first integrated study of rationalist causes of war that combines randomized experiments with fine-grained historical research. Despite a wide theoretical literature, there are few empirical tests of rationalist explanations for war. I provide a series of empirical tests using experimental and historical evidence. I also map out six mechanisms for shutting down the commitment and information problems. I test these mechanisms with laboratory and Internet experiments, and assess their realism with historical cases. The experimental and historical findings tend to converge. All three enforcement mechanisms (exogenous, endogenous, and inadvertent enforcement) have strong peace-promoting effects. The effects of the three signaling mechanisms (signaling with sunk cost, implementation cost, and salient contradiction) are mixed. Table 1.1 provides a sweeping summary of the empirical results.

Table 1.1: Overview of Key Empirical Results

	Experimental support for the mechanism *	Historical evidence for the mechanism *
<i>2 Rationalist Explanations</i>		
Commitment problem	Yes	Yes
Private information	No	Omitted #
<i>3 Enforcement Mechanisms</i>		
Exogenous Enforcement	Yes	Yes
Endogenous Enforcement	Yes	Yes
Inadvertent Enforcement	Yes	Yes
<i>3 Signaling Mechanisms</i>		
Costly Signaling	Mixed ##	Inconclusive
Costly Implementation	Yes	Yes
Contradictory Signaling	Yes	Limited

Notes: Chapters 2, 3, 4 and 5 discuss the evidence. Chapter 6 (Conclusion) summarizes the findings.

* For experiments: "Yes" denotes the detection of a significant treatment effect in the specific experiments implemented. For historical evidence: "Yes" suggests that evidence from the case universe supports the mechanism on balance. # The private-information explanation for war is difficult to test in a non-experimental setting: Private information always exists in international crises. It is also impossible to know every piece of private information held by each actor (as they are private). ## "Yes" for the signaler end of the mechanism; "No" for the receiver end of the mechanism.

The purpose of this chapter is to explain the first column and the first row of the table.

Section 1 in this chapter explains the theoretical mechanisms. Section 2 explains the methods behind the empirical results.

Wars have complex origins. No single theory can fully explain how and why war occurs. In policy-making, rationalist explanations for war should always be considered in conjunction with other potential causes that can range from organizational pathologies to individual misperceptions. There are few examples of a purely “rational” or “non-rational” war in history; a historical crisis often involves a complex mix of rational and non-rational elements. As scientists, however, it is important for us to identify and test the distinct mechanisms separately to understand their causal implications for complex cases. The purpose of this thesis is to establish a rationalist baseline for our theoretical understanding. In a definitional sense, we cannot fully define what are non-rationalist causes of war without first defining what are the rationalist causes. With this baseline, we may understand better how rationalist and non-rationalist factors interact to shape the phenomenon of war.

But first: What is the rationalist baseline?

1. Theories

The rationalist theory of war is not without criticisms,² but the assumption of rationality³ remains foundational in many theories of international politics.⁴ James Fearon's paper in

² Rationalist theory omits potential psychological influences on rational choice, and is often seen as less realistic than "bounded rationality", which emphasizes the cognitive limits in human decision-making. See, e.g., Herbert Simon, "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* 69, No. 1 (1955): 99-118; Ariel Rubinstein, *Modeling Bounded Rationality* (Cambridge, MA: MIT Press, 1998); Daniel Kahneman, "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review* 93, No. 5 (2003): 1449-75. Rationalist theory based on unitary states also treats national security interests as apparent and unproblematic. The domestic and ideational contestations that occur over the definition of "national interests" are omitted. For a critique of the restrictive assumptions used in Fearon's rationalist models, see Jonathan Kirshner, "Rationalist Explanations for War?" *Security Studies* 10, No. 1 (2000): 143-50.

³ While tighter assumptions of rationality are used for analytic modeling purposes, a broader definition of rationality can be used to interpret the substance of rationalist explanations for war in historical cases. I define rationality as the *tendency* to compare available options in terms of relative costs and benefits based on available information. A rational actor is one who is moved by that tendency; rational choice is the choice of an option that is perceivably better (or at least not worse) than others. As rationality is defined as a *behavioral tendency* rather than an absolute capability or an achieved state of affairs, this definition of a rational actor is broad. Indeed, it would be hard to identify many national decision-makers who do not fall under this general conception of rationality. But while this conception is broad, it does not contradict the spirit of the basic economic conception of rational choice as the "smart maximization of self-interest", dealing with "the question of how each player should act in order to promote his own interests most effectively". Amartya Sen, *The Idea of Justice* (Cambridge, MA: Harvard University Press, 2009), 179; John Harsanyi, *Rational Behavior and Bargaining Equilibrium in Games and Social Situations* (Cambridge, UK: Cambridge University Press, 1977), 16. Neither does the definition contradict the usual characterizations of rational actors in international relations. Definitions of rationality in international relations include James Morrow's characterization of rationality as "choosing the best means to gain a determined set of ends" and Charles Glaser's assumption of states being "able to identify and compare options, evaluating the prospects that they will succeed, as well as their costs and benefits." James Morrow, *Game Theory for Political Scientists* (Princeton, NJ: Princeton University Press, 1994), 17; Charles Glaser, *Rational Theory of International Politics: The Logic of Competition and Cooperation*. (Princeton, NJ: Princeton University Press, 2010), 1.

⁴ Rationality is a fundamental assumption in most of the existing realist and liberal theories of international politics, including seminal works such as Hans Morgenthau, *Politics Among Nations: The Struggle for Power and Peace* (New York: A. A. Knopf, 1948), Kenneth Waltz, *Theory of International Politics* (Reading, MA: Addison-Wesley, 1979), and Robert Keohane, *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton, NJ: Princeton University Press, 1984). Since the 1960s, scholars have used assumptions of rational choice to model strategic interaction in international relations. See, e.g., Thomas Schelling, *The Strategy of Conflict* (Cambridge, MA: Harvard University Press, 1960); Glenn Snyder and Paul Diesing, *Conflict Among Nations: Bargaining, Decision Making, and System Structure in International Crises* (Princeton, NJ: Princeton University Press, 1977); Donald Wittman, "How a War Ends: A Rational Model Approach," *Journal of Conflict Resolution* 23, No. 4 (1979): 743-63; R. Harrison Wagner, "The

1995 is the seminal work that defines the rationalist theory of war. The paper proposed three rationalist explanations for war. The first two are based on private information and the commitment problem.⁵ The third is based on issue indivisibility: a peaceful bargain with a proportionate division of value may not be possible when the issues are inherently indivisible.

The rationalist literature has converged on the first two explanations for war as “the full set of rationalist explanations that are both theoretically coherent and empirically plausible”.⁶ Both explanations are prominent in the wider literature on conflict and

Theory of Games and the Problem of International Cooperation,” *American Political Science Review* 77, No. 2 (1983): 330-46; Axelrod Robert, *The Evolution of Cooperation* (New York: Basic Books, 1984); Steven Brams, *Superpower Games: Applying Game Theory to Superpower Conflict* (New Haven, CT: Yale University Press, 1985); Kenneth Oye, ed., *Cooperation under Anarchy* (Princeton, NJ: Princeton University Press, 1986). Thomas Schelling pioneered the strategic analysis of conflict from a bargaining perspective. Bruce Bueno de Mesquita’s work decomposed the decision for war in terms of net expected utilities. Schelling, *Strategy of Conflict*; Thomas Schelling, *Arms and Influence* (New Haven: Yale University Press, 1966); Bruce Bueno De Mesquita, *The War Trap* (New Haven, CT: Yale University Press, 1981). Charles Glaser recently offered a general theory of international politics based on rational choice, in which international actions are determined through the interaction of state motivations with material and informational variables. See Glaser, *Rational Theory*.

⁵ While the commitment-problem explanation for war has not been explicitly articulated before Fearon’s 1995 paper, the idea that asymmetric information fosters war-causing miscalculations is well established. Historical studies of war have often connected the decision for war with miscalculations and misperceptions, e.g., Richard Smoke, *War: Controlling Escalation* (Cambridge, MA: Harvard University Press, 1977), and Richard Ned Lebow, *Between Peace and War: The Nature of International Crisis* (Baltimore, MD: Johns Hopkins University Press, 1981). Major theoretical works in international relations have also made the same connection, e.g., Robert Jervis, *Perception and Misperception in International Politics* (Princeton, NJ: Princeton University Press, 1976); Waltz, *Theory of International Politics*. Indeed, the security dilemma exists only because uncertainty exists; the security dilemma vanishes if there is complete information on intentions. Geoffrey Blainey suggested that wars are caused by disagreements in interstate estimates of capabilities and resolve: war occurs when states diverge in their estimates and war ends when the process of fighting reveals sufficient information for their estimates to converge. See Geoffrey Blainey, *The Causes of War* (New York: Free Press, 1988). States in the prewar stage may also initiate an escalatory action with the intent of making their opponent reveal its private information. For instance, a state might use a “limited and reversible probe” with an initial challenge to the status quo, which forces the opponent to signal its resolve or weakness in defending the status quo. Alexander George, *Avoiding War: Problems of Crisis Management* (Boulder, CO: Westview Press, 1991), 381.

⁶ Fearon, “Rationalist Explanations,” 380. Most scholars agree that most issues are potentially divisible; and even when they are not, they may be effectively divisible through side payments.

cooperation in political science and economics.⁷ Scholars of war have extended the private-information explanation⁸ and the commitment-problem explanation⁹ in different directions. The commitment-problem logic is also believed to play a critical role in a wide variety of political and economic phenomena, ranging from regime change and civil wars to international economic relations and the protection of human rights.¹⁰

Monica Toft and Ron Hassner provided a contrary view by arguing that subjective factors such as religiosity may make certain issues less divisible than others. Robert Powell showed that the indivisibility problem is essentially one form of the commitment problem. Monica Toft, *The Geography of Ethnic Violence: Identity, Interests, and the Indivisibility of Territory* (Princeton, NJ: Princeton University Press, 2005); Ron Hassner, "The Path to Intractability: Time and the Entrenchment of Territorial Disputes," *International Security* 31, No. 3 (2007): 107-38; Robert Powell, "War as a Commitment Problem," *International Organization* 60, No. 1 (2006): 169-203.

⁷ See survey articles in James Morrow, "The Strategic Setting of Choices: Signaling, Commitment, and Negotiation in International Politics," in *Strategic Choice and International Relations*, ed. David Lake and Robert Powell (Princeton, NJ: Princeton University Press, 1999), 77-114; Robert Powell, "Bargaining Theory and International Conflict," *Annual Review of Political Science* 5 (2002): 1-30; Dan Reiter, "Exploring the Bargaining Model of War," *Perspectives on Politics* 1 (2003): 27-43; Michelle Garfinkel and Stergios Skaperdas, "Economics of Conflict: An Overview," in *Handbook of Defense Economics*, ed. Todd Sandler and Keith Hartley (Amsterdam: North-Holland, 2007), 649-709; Matthew Jackson and Massimo Morelli, "The Reasons for War: An Updated Survey," in *Handbook on the Political Economy of War*, ed. Christopher Coyne and Rachel Mathers (Cheltenham, UK: Edward Elgar, 2011), 34-57.

⁸ See, e.g., R. Harrison Wagner, "Bargaining and War," *American Journal of Political Science* 44, No. 3 (2000): 469-84; Darren Filson and Suzanne Werner, "A Bargaining Model of War and Peace: Anticipating the Onset, Duration, and Outcome of War," *American Journal of Political Science* 46, No. 4 (2002): 819-37; Branislav Slantchev, "The Principle of Convergence in Wartime Negotiations," *American Political Science Review* 97, No. 4 (2003): 621-32; Andrew Kydd, "Which Side Are You On? Bias, Credibility, and Mediation," *American Journal of Political Science* 47, No. 4 (2003): 597-611; Robert Powell, "Bargaining and Learning While Fighting," *American Journal of Political Science* 42, No. 2 (2004): 344-61; Alastair Smith and Allan Stam, "Bargaining and the Nature of War," *Journal of Conflict Resolution* 48, No. 6 (2004): 783-813. Wagner and Smith and Stam constructed models in which states with conflicting priors on their chances of victory converge in their estimates in the course of fighting. Slantchev's principle of convergence suggests that wars end when further fighting is no longer informative once the divergent expectations of the belligerents have sufficiently converged. In Filson and Werner's model, wars start with the attacker's underestimation of the defender's capabilities and overestimation of the defender's concessions; and wars end when the defender's capabilities are sufficiently revealed through battle patterns and bargaining behaviors.

⁹ See, e.g., Robert Powell, *In the Shadow of Power: States and Strategies in International Politics* (Princeton, NJ: Princeton University Press, 1999); Robert Powell, "War as a Commitment Problem"; Bahar Leventoglu and Branislav Slantchev, "The Armed Peace: A Punctuated Equilibrium Theory of War," *American Journal of Political Science* 51, No. 4 (2007): 755-71.

¹⁰ See, e.g., Daron Acemoglu and James Robinson, *Economic Origins of Dictatorship and Democracy* (New York: Cambridge University Press, 2006); Barbara Walter, *Committing to Peace: The Successful Settlement of Civil Wars* (Princeton, NJ: Princeton University Press, 2002); James Fearon, "Why Do Some Civil Wars Last So Much Longer Than Others?" *Journal of Peace Research* 41, No. 3

I express the commitment-problem explanation as a commitment theory of war (Theory 1) and the private-information explanation as a signaling theory of war (Theory 2). In Theory 1, war is more likely when one side expects a significant increase in its *future vulnerability* that the opponent *cannot commit* not to exploit.¹¹ In Theory 2, war is more likely when private information cannot be signaled credibly.

Enforcement mechanisms

Theory 1 is driven by a stylized fact: *Actors respond to future vulnerabilities*. One side sees a future vulnerability that is exploitable by its opponent. It wants to prevent future exploitation. It can do so with a peaceful bargain if the opponent can commit not to exploit the other in the future. This solution is sustainable if there are mechanisms to enforce the commitment.

The rationalist theory of war is unduly constrained if we focus solely on formal bargains. To broaden the theory, I use a general definition of enforcement: To enforce is to disincentivize one from breaking an implicit or explicit contract or status-quo arrangement.¹² When we think more deeply about the nature of enforcement, its purpose

(2004): 275-301; Avner Greif, Paul Milgrom, and Barry Weingast, "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild," *Journal of Political Economy* 102, No. 4 (1994): 745-76; J. Lawrence Broz, "Political System Transparency and Monetary Commitment Regimes," *International Organization* 56, No. 4 (2002): 861-87; Olivier Jeanne, "Debt Maturity and the International Financial Architecture," *American Economic Review* 99, No. 5 (2009): 2135-48; Beth Simmons, *Mobilizing for Human Rights: International Law in Domestic Politics* (New York: Cambridge University Press, 2009).

¹¹ The rationale for this formulation is explained later in this section.

¹² To "enforce" is to "compel observance of or compliance with (a law, rule, or obligation)," according to *The New Oxford American Dictionary* (New York: Oxford University Press, 2010).

is essentially to prevent deviation from a prior arrangement. Hence, enforcement prevents commitment problems. The commitment problem exists when there is an incentive to renege. The commitment problem vanishes when the incentive to renege is removed. Enforcement, in a general sense, is the removal of the incentive to renege. And renegeing, in a broad sense, can apply to both explicit agreements as well as implicit status-quo arrangements.

I specify three enforcement mechanisms:

Mechanism M1 (Exogenous Enforcement):

War is less likely with *exogenous enforcement*.

Exogenous enforcement occurs when the incentive to renege is removed by an enforcer who is *not* one of the parties in conflict.¹³ In international relations, the absence of exogenous enforcement is seen as the root of the commitment problem: violent conflict happens – or peaceful cooperation does not occur – because there is no higher authority to enforce agreements or status-quo arrangements under international anarchy.¹⁴ Within nations, the government acts as the enforcing authority. Hence, contractual compliance and peaceful cooperation are more likely.¹⁵ Between nations, there is no world government that acts as an enforcer. Hence, compliance and cooperation are less likely.

¹³ In a bargaining framework, the enforcer is exogenous if it is not a bargainer in the bargaining game.

¹⁴ See, e.g., Fearon, “Rationalist Explanations”; Oye, *Cooperation Under Anarchy*; Waltz, *Theory of International Politics*.

¹⁵ Thomas Hobbes, *Leviathan* (Oxford: Oxford University Press, 1996 [1651]).

But exogenous enforcement can occur without a world government. It can occur as long as there is an external actor willing and able to act as the enforcer. Exogenous enforcement has played an important role in calming civil wars, as well as in peacekeeping missions in volatile international situations.¹⁶ It reduces the risk of war by increasing the cost of breaking a peaceful settlement or status-quo arrangement. When leaders see an external actor that can act as an enforcer, they are less likely to risk war. Conversely, leaders are more likely to risk war when they believe that exogenous enforcement is unlikely. Hence, in 1929, the Soviet Union attacked Manchuria with the calculation that other great powers would not intervene in a Sino-Soviet war.¹⁷ In 1931, conspirators at the Kwantung Army sparked the Second Sino-Japanese War with the calculation that the League of Nations would be ineffectual.¹⁸ The attacker's calculations were correct in both cases.

Exogenous enforcement can take one of two forms. It takes a "hard" form when the costs are imposed through military measures. It takes a "soft" form when the costs are imposed through non-military measures, such as economic sanctions, diplomatic isolation, or the withdrawal of international cooperation. Both forms should reduce the risk of war, though our attention often falls on the former rather than the latter.

¹⁶ See, e.g., Patrick Regan, "Third Party Interventions and the Duration of Intrastate Conflicts," *Journal of Conflict Resolution* 46, No. 1 (2002): 55-73; Walter, *Committing to Peace*; Virginia Page Fortna, "Does Peacekeeping Keep Peace? International Intervention and the Duration of Peace after Civil War," *International Studies Quarterly* 48, No. 2 (2004): 269-92.

¹⁷ Cheng Tianfang, *A History of Sino-Russian Relations* (Washington DC: Public Affairs Press, 1957), 154; Zeng Yeying, Huang Daoxuan, and Jin Yilin, *Zhonghua Minguo Shi* [History of Republican China], Vol. 7 (Beijing: Zhonghua Shuju, 2011), 221.

¹⁸ See Chapter 3.

In international relations, the idea of exogenous enforcement always assumes an external actor who acts as the enforcer *intentionally*. The external actor plays the role of the enforcer by intention; it has full knowledge that it is playing the role. But exogenous enforcement can also be achieved with an external actor that has *no intention* to act as an enforcer. This is inadvertent enforcement.

Mechanism M2 (Inadvertent Enforcement):

War is less likely with *inadvertent enforcement*.

Inadvertent enforcement occurs when a third-party rival enforces a peaceful bargain or status-quo arrangement even though it has no intention to do so. This occurs when war would create a strategic advantage for a third-party rival who might later threaten the parties at war. This expectation helps to enforce the peace by reducing the incentive to fight. The enforcement is inadvertent because third-party rivals have no intention of enforcing the bilateral peace. Indeed, they may benefit if the peace collapses. But they end up preserving the peace despite their preferences to the contrary.

The driving force in inadvertent enforcement is the fear of a third-party rival gaining a strategic advantage from a war between two parties: the greater the expected advantage, the greater the fear, and the greater the force of inadvertent enforcement. The fear creates *inadvertent doves*: even hawkish leaders may become war-averse when war makes them vulnerable to a third-party rival. This fear promotes peace by encouraging an inadvertent dove to make decisions or take measures to avoid war.

Consider the ironical case of Major-General Ishiwara Kanji, the radical militarist who had masterminded the Second Sino-Japanese War. On the eve of the Third Sino-Japanese War, however, the same Ishiwara pushed hard to contain the crisis with a localized peaceful settlement. He highlighted “the consideration over the war with the Soviet Union” as the “key reason” for avoiding a war with China: “That is, in the event of a protracted war [with China], if the Soviet Union attacks, Japan at this point would not be ready to handle it.”¹⁹ A similar reasoning based on the same Sino-Soviet-Japanese triad was used on the eve of the 1929 Sino-Soviet War, when Marshal Zhang Xueliang urged Nanjing to adopt a negotiated settlement with the Soviets “in order to prevent a third party [Japan] from fishing an advantage [from the situation].”²⁰

Inadvertent enforcement is a general theory of war avoidance. But it has specific implications for rationalist explanations for war. In particular, it offers a specific mechanism (Mechanism M2) for shutting down the commitment problem – one of the two central factors in the rationalist theory of war.

The strategic logic of M2 has ancient origins and connects to practical military strategy.²¹

However, M2 as a war-prevention mechanism remains largely undeveloped and untested

¹⁹ Record of Interview between Prince Takeda and Ishiwara Kanji, 1939, in Tianjin Municipal Translation Team for the *Daihon'ei Rikugunbu*, ed., *Riben Diguo Zhuyi Qinhua Ziliao Changbian: Da Ben Ying Lujunbu Zhaiyi* [Extensive Sources on the Japanese Imperialist Invasion of China: Selected Translations of *Daihon'ei Rikugunbu*] (Sichuan Renmin Chubanshe, 1987), 334.

²⁰ Xue Xiantian, *Minguo Shiqi Zhongsu Guanxi Shi: 1917-1949* [History of Sino-Soviet Relations in the Republican Era: 1917-1949], Vol. 1 (Beijing: Zhonggong Dangshi Chubanshe, 2009), 228.

²¹ I drew the strategic logic from a story in the ancient Chinese classic, *Zhan Guo Ce* [Strategies of the Warring States]: “A clam was sunbathing itself when a snipe suddenly poked its beak into the clam. In an instant the clam shut its shell and trapped the snipe's beak. Neither the clam nor the snipe could [commit to a mutually-beneficial bargain]. At last, a fisherman came and caught them both.”

in political science.²² The connection between M2 and the commitment problem has also never been made. But M2 is historically important. The war-restraining effects of M2 – and the war-promoting effects of its removal – may be found in the decision-making process in some of the most costly wars in history, including the Third Sino-Japanese War, the Pacific War and the Vietnam War.²³

M2 makes three basic assumptions. First, it assumes that war is costly and that it weakens those at war relative to their third-party rival.²⁴ Hence, the existence of a third-party rival reduces the expected utility of war. Hence, it promotes decisions to avoid war. This assumption often holds when the third party is a state actor. Second, inadvertent

(Translated by author.) Two lessons can be drawn. First, the clam and the snipe died of a commitment problem. Second: fighting a costly war weakens the fighting parties and benefits third parties. A similar logic is also implied in practical military strategies that emphasize the need to avoid fighting two enemies at once, such as the strategy of “divide-and-conquer.”

²² The most closely related work is Blainey, *Causes of War*, chapter 4. The chapter does not develop specific theoretical propositions, but it discusses how third parties at the international level promote war through neutrality or assistance. That war may be prevented through non-neutrality is implied. Other related works in international relations include the literature on the calculus of third-party intervention; alliance and balancing behavior; and third-party intervention in civil conflict. See Michael Altfeld and Bruce Bueno de Mesquita, “Choosing Sides in Wars,” *International Studies Quarterly* 23, No. 1 (1979): 87-112; Chac-Han Kim, “Third-Party Participation in Wars,” *Journal of Conflict Resolution* 33, No. 4 (1991): 659-77; Suzanne Werner, “Deterring Intervention: The Stakes of War and Third-Party Involvement,” *American Journal of Political Science* 44, No. 4 (2000): 720-32; Kevin Siqueira, “Conflict and Third-Party Intervention,” *Defence and Peace Economics* 14, No. 6 (2003): 389-400; Renato Corbetta, “Determinants of Third Parties’ Intervention and Alignment Choices in Ongoing Conflicts, 1946-2001,” *Foreign Policy Analysis* 6, No. 1 (2010): 61-85; Waltz, *Theory of International Politics*; Stephen Walt, *The Origins of Alliances* (Ithaca, NY: Cornell University Press, 1987); Alastair Smith, “Alliance Formation and War,” *International Studies Quarterly* 39, no. 4 (1995): 405-25; Powell, *Shadow of Power*; Regan, “Third Party Interventions”; Walter, *Committing to Peace*; Fortna, “Does Peacekeeping Keep Peace?”; Michael Findley and Tze Kwang Teo, “Rethinking Third-Party Interventions into Civil Wars: An Actor-Centric Approach,” *Journal of Politics* 68, No. 4 (2006): 828-37. In all these cases, however, a deliberate choice is made by the third-party intervener to support an ally, terminate a conflict, or enforce the peace. Inadvertent enforcement focuses on how third-party rivalry – at both domestic and international levels – enforces the peace despite the contradictory intent and preference of the third-party. To my knowledge, the argument has not been systematically developed and tested in the existing literature.

²³ See the section on inadvertent enforcement in Chapter 5.

²⁴ The model in Chapter 4 distinguishes between two variants of inadvertent enforcement: A symmetric variant in which the third party is a rival of both warring parties, and an asymmetric variant in which the third party is a rival of one warring party but not the other.

enforcement assumes the perceived efficacy of the third-party rival: the third party is perceived as capable of depressing the war payoffs to the extent that the incentive to fight is suppressed.²⁵ Inadvertent enforcement is compromised if the third-party rival is perceived as too weak to make much difference, or if the expected war utility remains positive and higher in the present than in the future.

Finally, inadvertent enforcement assumes the existence of a *commitment problem with the third-party rival*. Otherwise, a warring party can simply make an alliance or strike a bargain with the third party not to attack it while it is weakened by the war. After the war has started, power shifts in the favor of the third-party rival, reducing its incentive to fulfill the prewar bargain. As a consequence, the commitment problem with a third party preserves the peace between the two parties. This leads to a paradoxical implication: Commitment problems can lead to war – but under inadvertent enforcement, an indirect commitment problem is also necessary for the preservation of peace.

While the focus is on inadvertent enforcement at the international level, the theory can also extend into the domestic level. When the third-party rival is a domestic rival, inadvertent enforcement provides a domestic-level rationalist theory of peace: War is less likely if it creates a potential political vulnerability that domestic opponents *cannot commit* not to exploit. In this case, a commitment problem at the domestic level calms the commitment problem at the international level, and reduces the risk of war.

²⁵ Here, M2 assumes the perception that the third-party rival has depressed the war payoffs sufficiently such that the incentive to renege is removed. This assumption is difficult to test with observational data. All observed cases of peace would suggest that the assumption did hold through revealed preferences, while all observed cases of war would suggest it did not hold. I will use an experiment to set up a fair test.

Mechanism M3 (Endogenous Enforcement):

War is less likely with *endogenous enforcement*.

The commitment problem can also be resolved through endogenous enforcement. Endogenous enforcement occurs when the incentive to renege is removed by one of the parties in conflict.²⁶ With M3, the set of enforcement mechanisms is complete.²⁷

Endogenous enforcement is not emphasized in the rationalist theory of war. There seems to be a common assumption that the parties in conflict are powerless to deal with the commitment problem that confronts them. This is a necessary modeling assumption: Without it, rational states will be expected to remove the commitment problem by themselves and avoid the suboptimal outcome of a costly war. This means that the commitment-problem explanation for war vanishes without this assumption.

But states may not be as impotent as assumed. States may work actively to manage the potential shift in power that causes the commitment problem (and potentially war). They can exercise strategic or tactical restraint: They may reduce their military spending. They may stop developing potential sources of military advantage, such as nuclear-weapon programs. They may reverse policies that cause a sharp change in future relative power, such as military installations on strategic territories or oil embargoes on oil-dependent

²⁶ The enforcer is endogenous if it is one of the bargainers in the bargaining game.

²⁷ Completeness is constructed by definition since the set of endogenous enforcement mechanisms is the logical complement of the set of exogenous enforcement mechanisms.

nations. Military spending, nuclear-weapon programs, or oil embargoes will benefit one side in the future. But they can also lead to a costly war in the present *before* the rising side reaches its favorable future and the declining side its unfavorable future. Hence, it may be better for both sides to achieve a peaceful bargain by reducing the commitment problem.

American reservations on imposing an oil embargo on Japan in 1940-41 echoed the logic of endogenous enforcement through tactical restraint. In July 1940, the Roosevelt administration decided not to impose an oil embargo on Japan for fear that it would provoke the Japanese.²⁸ But a total oil embargo was eventually implemented on 1 August 1941, with disastrous consequences. Since the Japanese military machine was highly dependent on foreign oil imports, the embargo would greatly strengthen American power at Japan's expense in the *future*. But precisely so, it made Japan desperate in the *present*. Hence, on the very eve of the Pacific War, U.S. diplomats rushed together a proposal that would temporarily lift the embargo to prevent Tokyo from resorting to extremities. Washington believed that if it could stem the steep decline in Japan's future power, Tokyo's desperation would be lessened, and a war would be avoided in at least the near future.²⁹

²⁸ Scott Sagan, "The Origins of the Pacific War," in *The Origin and Prevention of Major Wars*, ed. Robert Rotberg, Theodore Rabb, and Robert Gilpin (Cambridge, UK: Cambridge University Press, 1989), 328; Waldo Heinrichs, *Threshold of War: Franklin D. Roosevelt and American Entry into World War II* (New York: Oxford University Press, 1988), 134; Jonathan Utley, *Going to War with Japan, 1937-1941* (Knoxville, TN: University of Tennessee Press, 1985), 83-101.

²⁹ Akira Iriye, *The Origins of the Second World War in Asia and the Pacific* (London: Longman, 1987), 180. The rise and demise of the proposal are discussed in Chapter 5 of this thesis.

Endogenous enforcement can take different forms.³⁰ In this thesis, I focus on endogenous enforcement through strategic or tactical restraint. The breakdown of endogenous enforcement can lead to a conflict spiral that makes both parties worse off than before.³¹ In this light, policies of restraint have been discussed in the wider international relations literature, although their connection to the commitment problem has not been made.³² Intuitively, there is a plausible connection between the incentive for mutual restraint and the general rarity of war despite the anarchic conditions of world politics.³³ Practically, states can control their military strategies and investments, and they often seek to optimize rather than maximize their military potential.³⁴ Theoretically, states have the incentive to resolve the commitment problem by themselves to avoid the suboptimal outcome of a costly war. But do they attempt to do so in reality? Does endogenous enforcement prevent conflict? Does its reversal promote conflict? I address these questions in the empirical chapters.

³⁰ Theoretically, endogenous enforcement can also be achieved if the declining or disadvantaged power successfully enhances its relative power through internal or external balancing strategies. The race for greater power, of course, can lead to trouble even before endogenous enforcement is achieved. Strategic reactions from the opponent, such as arms racing or counter-alliances, can make war more likely *before* endogenous enforcement is attained. This thesis focuses on strategic restraint, which appears to be a more straightforward and promising way of achieving endogenous enforcement. However, I also draw suggestive implications from the historical cases for other forms of endogenous enforcement.

³¹ See Jervis, *Perceptions and Misperceptions*, chapter 3.

³² See, esp., Charles Glaser, "Political Consequences of Military Strategy: Expanding and Refining the Spiral and Deterrence Models," *World Politics* 44, No. 4 (1992): 519-25; Charles Glaser, "Realists as Optimists: Cooperation as Self-Help," *International Security* 19, No. 3 (1994): 50-90; and Charles Glaser, "The Security Dilemma Revisited," *World Politics* 50 (1997): 171-201. On the concept of unilateral restraint to resolve the arms race-disarmament dilemma, see Charles Osgood, *An Alternative to War or Surrender* (University of Illinois Press, 1962), and an assessment of the concept in George Downs and David Rocke, *Tacit Bargaining, Arms Races, and Arms Control* (Ann Arbor, MI: University of Michigan Press, 1990), 41-51.

³³ The causal effect implied here is hard to identify with observational data. Chapter 4 sets up an experimental test for the effect of unilateral restraint in the shadow of a commitment problem.

³⁴ On the guns-and-butter tradeoff faced by states in their armament policies, see Powell, *Shadow of Power*, chapter 2.

Windows and commitments

Theory 1 suggests that war is more likely when one side expects a significant increase in its *future vulnerability* that the opponent *cannot commit* not to exploit. This formulation combines the commitment-problem explanation in the rationalist literature with the window theory in the realist literature. A window is an impending shift in relative power: Window theory suggests that war is more likely when there is a window.³⁵ Fearon

³⁵ The idea that a shift in power causes war is known since Thucydides, who attributed the cause of the Peloponnesian War to the rising power of Athens and the alarm it created in Sparta. The most developed formulation of window theory is found in Stephen Van Evera, *Causes of War: Power and the Roots of International Conflict* (Ithaca, NY: Cornell University Press, 1999), chapter 4. Van Evera differentiated between a “window of vulnerability”, which is a “growing defensive weakness” or a “negative power shift”, and a “window of opportunity”, which is a “fading offensive advantage” or a “positive power shift”. In Dale Copeland, *The Origins of Major War* (Ithaca, NY: Cornell University Press, 2000), it is argued that the expected trends in the relative power differences between great powers create different incentives for a major war depending on the polarity of the international system. Prominent articulations and applications of window theory include Stephen Van Evera, “The Cult of the Offensive and the Origins of the First World War,” *International Security* 9, No. 1 (1984): 58-107; Richard Ned Lebow, “Windows of Opportunity: Do States Jump through Them?” *International Security* 9, No. 1 (1984): 147-86; Jack Levy, “Declining Power and the Preventive Motivation for War,” *World Politics* 40, No. 1 (1987): 82-107; Marc Trachtenberg, *The Craft of International History: A Guide to Method* (Princeton, NJ: Princeton University Press, 2006), chapter 4; Thomas Christensen, “Windows and War: Trend Analysis and Beijing’s Use of Force,” in *New Directions in the Study of China’s Foreign Policy*, ed. Alastair Iain Johnston and Robert Ross (Stanford, CA: Stanford University Press, 2006), 50-85; M. Taylor Fravel, *Strong Borders, Secure Nation: Cooperation and Conflict in China’s Territorial Disputes* (Princeton, NJ: Princeton University Press, 2008). Power transition theories share the same central variable: the shift in power over time. In window theory, it is the declining state that triggers a war. In power transition theory, however, it may be the rising state that initiates a war. The rising state triggers a war to revise the existing status quo where the distribution of benefits favors the dominant state. War becomes more likely as the power of the rising state approaches the power of the dominant state in the system. See A. F. K. Organski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1981); Jacek Kugler and Douglas Lemke, *Parity and War: Evaluations and Extensions of the War Ledger* (Ann Arbor, MI: University of Michigan Press, 1996); Ronald Tammen et al., *Power Transitions: Strategies for the Twenty-First Century* (New York: Chatham House, 2000). Robert Gilpin emphasized a preventive war by the dominant state rather than a revisionist war by the rising state: the dominant state tends to overload its commitments and overstretch its capabilities over time, which creates vulnerabilities in the dominant state and opportunities for the rising state. As the power of the rising state becomes threatening, a preventive war to eliminate the threat becomes attractive to the dominant state. See Robert Gilpin, *War and Change in World Politics* (Cambridge, UK: Cambridge University Press, 1981); and also Richard Ned Lebow and Benjamin Valentino, “Lost in Transition: A Critical Analysis of Power Transition Theory,” *International Relations* 23, No. 3 (2009): 389-410. Lebow and Valentino argued that the power transition theory is overly deterministic. They also found little

suggested that the relationship between war and the commitment problem can take one of three forms:³⁶ a preventive war by a declining state; a preemptive war motivated by the existence of a first-move advantage;³⁷ or a war over strategic territory that influences the ex-post distribution of power.³⁸ Despite their different manifestations, all three arise from a power shift across time.³⁹ In other words, there should be no commitment problem without a window.⁴⁰ And a window without a commitment problem should not lead to war: rational states would choose a peaceful bargain to avoid the costs of war.

Theory 1 explicitly connects both the window theory and the commitment problem, as neither alone is complete. I identify the cause of war in the perceived *potential vulnerability* generated by the impending power shift, which the opponent *cannot commit* not to exploit. A power shift has no decisional impact if there is no potential vulnerability

evidence that wars between rising and dominant powers are fought with the primary objective to defend or revise the status quo. See also the analysis in Powell, *Shadow of Power*, chapters 3-4.

³⁶ Fearon, *Rationalist Explanations*, 401-9.

³⁷ Thomas Schelling identified “[t]he premium on haste – the advantage, in case of war, in being the one to launch it or in being a quick second in retaliation if the other side gets off the first blow” as “undoubtedly the greatest piece of mischief that can be introduced into military forces, and the greatest source of danger that peace will explode into all out war.” Schelling, *Arms and Influence*, 227. See also Leventoglu and Slantchev, “The Armed Peace,” 755-71. Leventoglu and Slantchev constructed a complete-information model in which the war incentive varies with the size of the first-strike advantage. The incentive changes in the course of fighting. Specifically, if first-strike advantages (which determine victory) were sufficiently reduced by the destruction of resources during the war, the incentive for peaceful settlement would increase.

³⁸ The first connects to the theory of preventive war in the realist literature. See Alfred Vagts, *Defense and Diplomacy: The Soldier and the Conduct of Foreign Relations* (New York: King's Crown Press, 1956), chapter 8; Van Evera, “The Cult of the Offensive,” 58-107; and Van Evera, *Causes of War*, chapter 4. The latter two connect to the theories of war due to first-move advantage and resource cumulativeness in Van Evera, *Causes of War*, chapters 3 and 5. Robert Powell offered a fourth mechanism based on costly deterrence: States choose to fight earlier than later when the expected cost of deterrence in the long term exceeds the expected cost of eliminating the threat altogether in the present. Powell, “War as a Commitment Problem,” 192-4.

³⁹ Powell, “War as a Commitment Problem,” 180-8. Realists focus more on shifts in military power while rationalists focus more on shifts in bargaining power. The two are equivalent: changes in military power translate into changes in bargaining power.

⁴⁰ This assumes the absence of asymmetric information, which is a separate cause of war.

perceived by the decision-maker. A future power shift is never a certain fact until it occurs. What matters at the point of decision is whether a potential vulnerability is perceived.

Signaling mechanisms

Theory 2 highlights the private-information explanation for war: *War is more likely when private information cannot be signaled credibly.* A signal is a piece of information communicated by one party (the “signaler”) and observed by another (the “receiver”). In international politics, signaling involves the making of threats or promises to deter, compel, or persuade. A credible signal is one that convincingly shows that the signaler is resolved to fulfill its threat or promise. War-causing miscalculations are more likely when private information cannot be credibly signaled. I focus on three signaling mechanisms:

Mechanism M4 (Costly Signaling):

War is more likely when the threats to use force are not *costly in themselves.*

M4 comes from the theory of costly signaling in international relations.⁴¹ To deter the opponent, one’s resolve must be credibly signaled. To be credible, the signal must

⁴¹ The theory of costly signaling in international politics is elaborated in James Fearon, “Threats to Use Force: Costly Signals and Bargaining in International Crises,” PhD dissertation (University of California, Berkeley, 1992); James Fearon, “Domestic Political Audiences and the Escalation of International Disputes,” *American Political Science Review* 88, No. 3 (1994): 577-92; and James Fearon, “Signaling Foreign Policy Interests: Tying Hands Versus Sinking Costs,” *Journal of Conflict Resolution* 41, No. 1 (1997): 68-90. Generally, signals of resolve become credible if they create some

differentiate between resolved and unresolved states by carrying some costs or risks that would discourage unresolved states from sending that signal.⁴² Fearon suggested that states have two ways to signal their resolve.⁴³ The first is “tied-hands signaling”, which binds the signaler to a higher cost of backing down if the opponent does not back down, but is costless to the signaler if the opponent backs down.⁴⁴ Such signals may involve “audience costs” – the political costs suffered by a national leader for backing down after making an international threat.⁴⁵

cost that an unresolved signaler would not be willing to incur. Thomas Schelling might have been the first to explore the strategic use of costs to signal resolve. Robert Jervis combined ideas from Schelling and Erving Goffman in the first comprehensive discussion of signaling in international relations. Robert Powell and Barry Nalebuff extended the insights with a focus on deterrence. See Schelling, *Strategy of Conflict*; Robert Jervis, *The Logic of Images in International Relations* (Princeton, NJ: Princeton University Press, 1970); Erving Goffman, *The Presentation of Self in Everyday Life* (Garden City, NY: Doubleday, 1959); Robert Powell, *Nuclear Deterrence Theory: The Search for Credibility* (Cambridge, UK: Cambridge University Press, 1990); Barry Nalebuff, “Rational Deterrence in an Imperfect World,” *World Politics* 43, No. 3 (1991): 313-35. In economics, the use of costly signals to transmit information is formally explored in Michael Spence, “Job Market Signaling,” *Quarterly Journal of Economics* 87, No. 3 (1973): 355-74. However, costless signaling is occasionally informative. Crawford and Sobel provided the seminal model of costless signaling (“cheap-talk”) in which informative equilibria may exist if the divergence of interests between signaler and receiver is sufficiently small. Joseph Farrell showed that cheap talk might also inform players in coordinating on a Nash equilibrium that is more optimal than the equilibrium by default. See Vincent Crawford and Joel Sobel, “Strategic Information Transmission,” *Econometrica* 50, No. 6 (1982): 1431-51; Joseph Farrell, “Cheap Talk, Coordination, and Entry,” *RAND Journal of Economics* 18, No. 1 (1987): 34-39. Focusing on the role of diplomacy in international crises, Anne Sartori (2005) suggested that “talk” might not be “cheap” as it can involve reputational costs tied to a state’s ability to use deterrent threats in future disputes. Stephen Van Evera and Branislav Slantchev also highlighted the incentive of high-resolved states to muffle their signaling under certain conditions – for example, to feign weakness and achieve tactical surprise. Anne Sartori, *Deterrence by Diplomacy* (Princeton, NJ: Princeton University Press, 2005); Van Evera, *Causes of War*, 45-51; Branislav Slantchev, “Feigning Weakness,” *International Organization* 64, No. 3 (2010): 357-88.

⁴² A threat is “rendered credible when the act of sending it incurs or creates some cost that the sender would be disinclined to incur or create if he or she were in fact not willing to carry out the threat For a threat to increase the target’s belief that the sender would be willing to fight, it must be more likely that a resolved state would make the threat than an unresolved state. Thus, to be credible, a threat must have some cost or risk attached to it that might discourage an unresolved state from making it.” Fearon, “Signaling Foreign Policy Interests,” 69.

⁴³ Fearon, “Signaling Foreign Policy Interests,” 68.

⁴⁴ Fearon’s model suggests that states are better off with tied-hands signals than with sunk-cost signals as tied-hands signals are not costly in themselves; however, they create a higher risk of war. Fearon, “Signaling Foreign Policy Interests,” 71, 78-81.

⁴⁵ James Fearon, “Domestic Political Audiences,” 577.

The second is “sunk-cost signaling”, which creates direct costs that cannot be recovered, and which does not affect the relative value of escalation or compromise. One example is the costly maintenance of nuclear missiles in Western Europe by the United States in the Cold War: Those missiles were unlikely to affect the outcome of a nuclear war under mutually assured destruction, but the sunk costs helped to signal U.S. resolve in defending its European allies.⁴⁶

Fearon classified tied-hands signaling and sunk-cost signaling as two different forms of costly signaling.⁴⁷ Here, I use a strict definition of costly signaling to include only sunk-cost signaling.⁴⁸ The reason is because a sunk-cost signal is costly in itself, but a tied-hands signal is not – it is costless to the signaler if the opponent backs down.⁴⁹

Costly signaling theory is prominent in the international relations literature, but there are few empirical tests in the literature demonstrating that costly signals are indeed more credible.⁵⁰ I fill the empirical gap and reconsider the theoretical logic of costly signaling based on the evidence in Chapters 3 to 5.

⁴⁶ Branislav Slantchev, *Military Threats: The Costs of Coercion and the Price of Peace* (Cambridge, UK: Cambridge University Press, 2011), 36. The example is attributed to Barry O'Neill.

⁴⁷ Fearon, “Signaling Foreign Policy Interests,” 68-70.

⁴⁸ Spence, “Job Market Signaling.”

⁴⁹ Fearon pointed out that tied-hands signals in his model have a similar effect “on the challenger's behavior as do sunk-cost signals – they minimize the risk of a challenge by signaling that the defender will fight for sure – but they are not in and of themselves costly. [W]ith sunk-cost signals, any type of the defender that chooses to signal pays the costs upfront for doing so, whereas with tying-hands signals the audience costs created are never paid in equilibrium because no type backs down after creating them.” Fearon, “Signaling Foreign Policy Interests,” 81. Of course, since the opponent may not back down, a tied-hands signal is always risky to the signaler.

⁵⁰ One reason is that there are few pure cases of sunk-cost signals in international politics: Fearon highlighted that the effects of sunk-cost and tied-hands signals are often mixed together in real-world cases, but “it is important to see ... that two distinct mechanisms are at work, and we need to analyze

Beyond costly signaling, Daryl Press summarized four theories of credibility in international relations.⁵¹ In one of them – the current-calculus theory – threats are credible if the threatener has sufficient capabilities and interests in fulfilling them. The theory articulates a relationship emphasized in the deterrence literature between credibility and the calculation of opponent interests.⁵² However, what is “current calculus” is difficult to pin down. It includes intrinsic interests and relative capabilities,

them separately as ideal types to understand the strategic logic of mixed cases.” Fearon, “Signaling Foreign Policy Interests,” 70. In contrast, there is a major empirical debate on the existence of domestic audience costs in international crises. Michael Tomz showed the existence of domestic audience costs measured by public approval ratings in the United States. Trager and Vavreck test the sensitivity of domestic audience costs to party labels and the rhetoric used to issue the threat. The study found that vague threats lead to lower audience costs than specific threats, but party treatments do not affect the level of audience costs. Levendusky and Horowitz tested the sensitivity of domestic audience costs to three domestic-level variables: the president’s justification for backing down (“because he received new information”), the reactions of other political elites, and the party of the president. The study found substantial effects for the first two variables but minimal effects for the third. Recently, Snyder and Borghard and Marc Trachtenberg argued that there is little historical evidence to support the logic of domestic audience costs in international crises. Kenneth Schultz pointed out that audience costs are like the “dark matter” of international relations: they are theoretically useful but difficult to observe directly. A chief reason for the difficulty is strategic selection bias. As Tomz suggested: “If leaders take the prospect of audience costs into account when making foreign policy decisions, then in situations when citizens would react harshly against backing down, leaders would tend to avoid that path, leaving little opportunity to observe the public backlash.” Michael Tomz, “Domestic Audience Costs in International Relations: An Experimental Approach,” *International Organization* 61, No. 4 (2007): 822. See Matthew Levendusky and Michael Horowitz, “When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs,” *Journal of Politics* 74, No. 2 (2012): 323-38; Robert Trager and Lynn Vavreck, “The Political Costs of Crisis Bargaining: Presidential rhetoric and the Role of Party,” *American Journal of Political Science* 55, No. 3 (2011): 526-45; Jack Snyder and Erica Borghard, “The Cost of Empty Threats: A Penny, Not a Pound,” *American Political Science Review* 105, No. 3 (2011): 437-56; Marc Trachtenberg, “Audience Costs: an Historical Analysis,” *Security Studies* 21 (2012): 3-42.

⁵¹ Daryl Press, *Calculating Credibility: How Leaders Assess Military Threats* (Ithaca, NY: Cornell University Press, 2005).

⁵² See Alexander George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974); Paul Huth and Bruce Russett, “Deterrence Failure and Crisis Escalation,” *International Studies Quarterly* 32, No. 1 (1988): 29-45; Elli Lieberman, “The Rational Deterrence Theory Debate: Is the Dependent Variable Elusive?” *Security Studies* 3, No. 3 (1994): 384-427; Paul Huth, “Deterrence and International Conflict: Empirical Findings and Theoretical Debates,” *Annual Review of Political Science* 2, No. 1 (1999): 25-48. Branislav Slantchev studied military coercion as a class of costly signals with endogenous effects on the military balance: Military threats, such as troop mobilization, not only have a signaling effect, but also a direct effect on the relative distribution of power, which in turn reduces the opponent’s expected war payoff and incentivizes acquiescence. Branislav Slantchev, “Military Coercion in Interstate Crises,” *American Political Science Review* 99, No. 4 (2005): 533-47; Slantchev, *Military Threats*.

but the relationship between the two is not always clear. In particular, “interest” is conceptually broad and empirically very slippery. I propose a narrow formulation based on the rational implementability of the signaled threat. This theoretical formulation is more limited but also more precise. The narrow theory – *the costly implementation theory* – suggests a signaling mechanism based on the perceived costs of implementation.⁵³

Mechanism M5 (Costly Implementation):

War is more likely when the threats to use force are *costly to implement*.

Ceteris paribus, a signal is less credible when its implementation is costly. Observers undervalue the signal because they know that it would be costly for a signaler to do what it signaled it would do. A rational signaler may still send the signal as long as it is not completely incredible. Indeed, the signaler may be compelled to send the signal if *not* sending the signal is informative in itself. That is, while sending the signal may *not* convince the opponent of the signaler’s resolve, not sending the signal *would* convince the opponent of the signaler’s non-resolve.

But the more costly it is to implement the threat, the less credible is the signal, and the more likely that miscalculations would arise and lead to war. A threat is less credible if implementing it is costly. A threat is more credible if implementing it is cheap.

⁵³ A more general theory of credibility balances the cost of implementation with the cost of *non-implementation*. The general theory is more complete, but it also involves reading the signaler’s interest over a contested issue and assigning proportionate credibility weights based on the inferred interest. Costly implementation theory is a special theory of credibility that sidesteps the problem with a narrower and more parsimonious formulation. It brackets the cost of non-implementation.

By implication, strategic manipulations of implementation costs can enhance credibility. This implication highlights the connection as well as the difference between M4 and M5: a signal with sunk cost can also affect credibility *if* it also affects the cost of implementation. If part of the implementation cost is already paid in sending the costly signal (e.g. the cost of mobilization), then the signaled threat (e.g. the use of force) becomes cheaper to implement, and thus more credible.

Costly implementation theory, in its general form, has suggestive implications for the concept of audience costs (tied-hands signaling): The credibility effect of audience costs rests ultimately on the rational implementability of the threat. Audience costs make a threat more credible because they make the threat more *costly to not implement*.

Mechanism M5 has important historical implications. It resonates in several wars in East Asia. In 1950, the United States believed that overt Chinese intervention in Korea was unlikely, despite at least three warnings from Beijing and the deployment of hundreds of thousands of Chinese troops in Manchuria.⁵⁴ Washington assumed that it would be too costly for China to implement its threat to fight the United States.⁵⁵ On the eve of the 1962 Sino-Indian War, India escalated her “forward policy” despite repeated warnings from China, partly because she thought that the Chinese threat to use force would be too

⁵⁴ Chen Jian, *China's Road to the Korean War: The Making of the Sino-American Confrontation* (New York: Columbia University Press, 1994), 163-4, 170; William Stueck, *Rethinking the Korean War: A New Diplomatic and Strategic History* (Princeton, NJ: Princeton University Press, 2002), 89; Thomas Christensen, *Worse Than a Monolith: Alliance Politics and Problems of Coercive Diplomacy in Asia* (Princeton, NJ: Princeton University Press, 2011), 67-8.

⁵⁵ Stueck, *Rethinking the Korean War*, 100-1; Roscmay Foot, *The Wrong War: American Policy and the Dimensions of the Korean Conflict, 1950-1953* (Ithaca, NY: Cornell University Press, 1985), 80-2, 86.

costly to implement. Other examples include the Russo-Japanese War, the Sino-Soviet War, and the Changkufeng War.⁵⁶

Mechanism M6 (Contradictory Signaling):

War is more likely when there is a *salient contradiction* in the signals of resolve sent in a crisis.

Signals are often evaluated serially rather than one in isolation from the others. Signals are less credible when they involve a salient contradiction. A *contradiction* occurs when signals of strong resolve are mixed with signals of weak resolve in the series of signals sent. The contradiction is *salient* if it is registered and considered by the opponent. Signal consistency differentiates between resolved and unresolved states on the assumption that a resolved state is more willing and able to signal its resolve consistently. Signal inconsistency may reflect disunity or disorganization within the signaling state. It may imply that the signaler has yet to make a clear and resolved decision. It may suggest a strategic intent to maintain some ambiguity and hedge between competing policy alternatives. In all these cases, contradictory signaling reveals the lack of resolve.

The logic in M6 is straightforward and intuitive:⁵⁷ Credibility is compromised when our signals of resolve are mutually contradictory. Flip-flopping is bad for credibility. U.S.

⁵⁶ See the section on costly implementation in Chapter 5.

⁵⁷ M6 is straightforward because it focuses purely on whether a salient contradiction exists. Finer-grained theories of different levels of credibility under different permutations and combinations of contradictory signals will be much more complex. As M6 has not yet been systematically articulated and tested in the existing literature, it is important to establish the core mechanism first before we move to more detailed specifications of the mechanism.

signaling in the period leading to the Korean War provides a classic example. The U.S. withdrew its troops from South Korea in June 1949, but expanded and made permanent the provisional U.S. advisory team in South Korea, with increased commitments in American aid, in July 1949.⁵⁸ However, in January 1950, Secretary of State Dean Acheson publicly excluded Korea from the American defense perimeter. Hence, U.S. signals of resolve had salient contradictions. Concluding that “the prevailing mood [in the U.S.] is not to interfere,” Stalin approved Kim Il-sung’s proposal to attack South Korea.⁵⁹

While the logic of M6 haunts the road to the Korean War, the mechanism remains underdeveloped in the international relations literature on signaling. The closest are the “ingrained-lessons” and “past-actions” theories of credibility summarized by Daryl Press.⁶⁰ In the former, leaders’ assessments of credibility depend on their country’s specific experiences in previous crises.⁶¹ In the latter, an opponent’s threat is more credible if the opponent has a history of fulfilling its threats.⁶² But none of these theories

⁵⁸ Stueck, *Rethinking the Korean War*, 73-79; Doris Condit, *History of the Office of the Secretary of Defense: The Test of War, 1950-1953* (Washington, DC: Historical Office, Office of the Secretary of Defense, 1984), 45.

⁵⁹ Stueck, *Rethinking the Korean War*, 73.

⁶⁰ Press, *Calculating Credibility*.

⁶¹ Ernest May, Yuen Foong Khong, and Marijke Breuning have argued that historical analogies influence credibility estimates, as suggested by the ingrained-lessons theory. Ernest May, *Lessons of the Past: The Use and Misuse of History in American Foreign Policy* (New York: Oxford University Press, 1973); Yuen Foong Khong, *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965* (Princeton, NJ: Princeton University Press, 1992); Marijke Breuning, “The Role of Analogies and Abstract Reasoning in Decision-Making: Evidence from the Debate over Truman’s Proposal for Development Assistance,” *International Studies Quarterly* 47, No. 2 (2003): 229-45.

⁶² See contrary arguments in Paul Huth and Bruce Russett, “What Makes Deterrence Work? Cases from 1900 to 1980,” *World Politics* 36, No. 4 (1984): 496-526; and Jonathan Mercer, *Reputation and International Politics* (Ithaca, NY: Cornell University Press, 1996). Daryl Press used a set of historical cases to support the current-calculus theory. His rejection of the past-actions theory, however, remains controversial. Anne Sartori showed that a reputation based on a history of honest diplomacy matters during crises, and that the ability to use deterrent threats in future disputes becomes compromised if a

deal directly with signaling mechanisms, since past histories and ingrained lessons are not manipulable in a given crisis. M6 differs in its focus on signals *within a given crisis*, without inferring credibility from other historical episodes prior to and separate from the current crisis.

state develops a reputation for bluffing. Mark Crescenzi argued that past actions influence credibility estimates when relative capabilities are equivalent. In international finance, Michael Tomz showed that debtor states that had a history of repaying their loans in tough times enhanced their reputation and received lower interest rates in future loans. See Press, *Calculating Credibility*; Sartori, *Deterrence by Diplomacy*; Mark Crescenzi, "Reputation and Interstate Conflict," *American Journal of Political Science* 51, No. 2 (2007): 382-96; Michael Tomz, *Reputation and International Cooperation: Sovereign Debt across Three Centuries* (Princeton, NJ: Princeton University Press, 2007).

2. Methods

There is very limited empirical work on rationalist explanations for war.⁶³ I build an empirical complement to the literature on two fronts: (1) I use randomized experiments to construct clean tests of the core mechanisms in Theories 1 and 2, and the specific mechanisms in M1-M6. (2) I use historical case studies to evaluate if these mechanisms are realistic in actual crises.

Experimental methods and formal models

Observationally, it is impossible to rule out the effects of confounders that are unmeasured or unobservable. It is also difficult to measure variations in the information or commitment environment. For example, the private-information explanation by itself is very difficult to test with observational data: Private information *always* exists in international politics, and we can never know every piece of private information held by each actor (since they are *private*). With an experiment, however, we can control the information parameters and identify their causal effects on the outcome variable. An experiment also eliminates the difficulties associated with potential strategic selection

⁶³ While the rationalist literature on war has made significant advances theoretically, empirical work in this area remains limited. The key theoretical variables in the rationalist literature are empirically slippery. For instance, it is not clear how private information or the commitment problem may be coded and measured uncontroversially with quantitative data. The rationalist theory of war has been qualitatively assessed in various historical cases. The most comprehensive effort is David Lake, “Two Cheers for Bargaining Theory: Assessing Rationalist Explanations of the Iraq War,” *International Security* 35, No. 3 (2010): 7-52. These efforts have provided valuable insights, although coding and measuring private information and the commitment problem – and controlling for potential confounders – remain a major challenge. Earlier studies that touched on important elements of the rationalist theory with historical cases include Blainey, *Causes of War*, and Van Evera, *Causes of War*. Glaser’s *Rational Theory* broadened the explanatory scope of the rationalist theory and provided useful historical comparisons.

effects.⁶⁴ I use an experimental approach to control and manipulate the information and commitment environments, and construct clean tests of their effects on conflict decisions.

Fearon's private-information and commitment-problem models are basically mechanisms of bilateral conflict.⁶⁵ Each is derived from a strategic-interaction mechanism that can be formally generalized and experimentally tested. Do these mechanisms cause bilateral conflict? Do they predict the human behavior that is actually observed? I model these mechanisms formally and test them with strategic-interaction experiments. Of course, since specific historical contexts cannot be fully replicated in the experiments, we cannot know whether the causal effect observed in the lab would also be observed in a *specific* case, say the Korean War or the Vietnam War, where a multitude of confounders existed. But we can distill the *general* theoretical mechanism and test it in a clean *ceteris-paribus* setting where all confounders are eliminated by experimental design.

I use this approach to test the enforcement mechanisms in M1 to M3 based on laboratory experiments. To test the signaling mechanisms, I conduct Internet experiments in which signals are experimentally manipulated based on M4 to M6. The laboratory experiments involve medium-*n* samples of students from MIT and Harvard University. The Internet experiments involve large-*n* diverse national samples recruited across the United States.

⁶⁴ For example, if leaders have already achieved endogenous enforcement through policies of restraint *before* the crisis, then we are less likely to observe the crisis happening, which makes it difficult for us to test the endogenous enforcement mechanism in an observational setting. It is also likely that a crisis that occurs despite an ex-ante policy of restraint is systematically different from crises that occur in the absence of ex-ante restraint. These potential problems are avoided with a randomized experiment under which strategic selection bias is cleansed out.

⁶⁵ Fearon, "Rationalist Explanations."

Experiments have high internal validity. They allow us to replicate the *ceteris-paribus* condition assumed in scientific theories. In particular, strategic-interaction experiments supplement our stylized models of hypothetical game-theoretic behavior with real data on actual human behavior. These experiments isolate the strategic mechanisms of formal models in a controlled setting, and test if human decision-making converge to or diverge from the equilibrium predictions in those models.

External validity, however, is an important concern. Two limitations are important to note. The first is that the subject pool is not a representative sample of national leaders caught within an international crisis. It is very difficult to obtain such a sample. However, while we cannot eliminate the possibility that certain findings may be affected by the subject samples in the experiment, the data do provide real behavioral information on strategic decision-making given our variables of interest.

The second limitation is that the stakes in the experiments are smaller than the stakes involved in actual interstate wars. We cannot scale up the stakes to involve life and death of real human beings, as decisions for wars do. Nonetheless, it should be noted that formal theories of war are not dependent on “stake effects.” Formally, the models are driven by specific payoff structures, which *are* replicated in the experiment. Practically, we should also consider if small stakes tend to under-estimate or over-estimate the treatment effect. If it is the former, and we find that a large treatment effect is detected *despite* the small stakes, then it is reasonable to expect that the treatment effect in the

experiment is likely to be an *under-estimation* of the true treatment effect in the real world where the stakes are higher.⁶⁶

Experiments replicate the *ceteris-paribus* condition often assumed in scientific theories but rarely achievable with observational data. Hence, carefully designed experiments provide special advantages for testing causal mechanisms in political science. But the investigation should not end here. We also need to understand how these mechanisms operate within complex policymaking processes where a multitude of historical contingencies and confounders exist. It is important to probe the external validity of the experimental results based on actual historical observations. This is a crucial task performed through historical cases.

Historical cases

My mechanisms are decision-based mechanisms. Hence, the unit of observation in the historical analysis is *not* the final outcome (war), but instances of pro-war or anti-war decisions or proposals made by decision-makers during the crisis.

My case selection uses a two-stage procedure. First is the *period selection*. The ideal timeframe is one with the highest contemporary relevance. The 20th century provides a good fit. Second is the *region selection*. The universe of all wars in the 20th century can

⁶⁶ For instance, being “reneged upon” in a commitment-problem situation leads to the loss of dollars in the context of our experiment, whereas the equivalent loss may be measured in life and death in the context of international politics. If we believe that people are less trusting and more sensitive to the commitment problem (our treatment variable) in the latter case, then the treatment effect isolated in the lab may in fact be an under-estimation of the true treatment effect in the real world.

be partitioned based on the major geographical regions of the world. My strategy is to select a major geographical region and study the *total universe of cases* within that region. This minimizes selection bias at the individual *case* level, while accepting a selection bias at the *region* level. The glass here is both half-empty and half-full. It is half-empty, because it remains unknown if the empirical results are valid in other regions. It is half-full, because the findings established in the case universe are valid within a major region of the world. These findings may or may not have global external validity. But understanding how the causes of war operate within a major geographical region of the world is valuable and non-trivial in itself.

A regional case universe may be chosen on three criteria. First, the case universe has a mix of wars with different intensities and actors. Second, the case universe is not too large as to make careful case studies infeasible. Third, the selected region has prominence in terms of its global economic or political importance, or in terms of its share of world population. Thus, war and peace in this region can make the most difference to the most number of people, and to the world.

East Asia satisfies these criteria. It also has critical influence on global security in the present and projected future. The region has become a foremost priority in U.S. security policy under the Obama administration, with the U.S. shifting the majority of its naval fleet to the Asia-Pacific by 2020.⁶⁷ The U.S. National Intelligence Council projected that

⁶⁷ “Leon Panetta: US to deploy 60% of navy fleet to Pacific,” *BBC News*, 1 June 2012.

Asia would have greater overall power than the U.S. and Europe combined by 2030.⁶⁸ And international peace in East Asia cannot be taken for granted. The risks of conflict are growing in a multipolar Asia, given the absence of a robust regional security framework and the erosion of the post-Cold War equilibrium anchored by U.S. primacy.⁶⁹

I have also chosen the East Asian universe for three practical reasons. First, there are no comparative historical studies of 20th-century East Asian wars in political science. This is the first of its kind. Second, several wars in the case universe remain unexamined in the political science literature. A study of such wars will provide useful theoretical information and empirical documentation for future scholars. The third reason is the potential for cross-validation: The canonical theories of war in international politics – including realist and rationalist theories – are largely inspired by modern Western history. Since case studies of East Asian wars are relatively limited in political science, it remains uncertain whether rationalist theories of war do hold in the East Asian universe. The answer to this question will be intrinsically interesting and important.

My case universe consists of all interstate wars fought in East Asia in the 20th century.

I identified a total of 16 cases based on the Correlates of War 3.0 classification:

⁶⁸ National Intelligence Council, *Global Trends 2030: Alternative Worlds* (National Intelligence Council, 2012), ix, 15-19.

⁶⁹ *Ibid.*, viii-ix, 61-69, 76-78. See also Aaron Friedberg, “Ripe for Rivalry: Prospects for Peace in a Multipolar Asia,” *International Security* 18, No. 3 (1993): 5-33; Thomas Berger, “Set for Stability? Prospects for Conflict and Cooperation in East Asia,” *Review of International Studies* 26, No. 3 (2000): 405-28; Aaron Friedberg, *A Contest for Supremacy: China, America, and the Struggle for Mastery in Asia* (New York: W.W. Norton, 2011). For counterarguments, see David Kang, “Getting Asia Wrong: The Need for New Analytical Frameworks,” *International Security* 27, No. 4 (2003): 57-85; Amitav Acharya, “Will Asia's Past Be Its Future?” *International Security* 28, No. 3 (2004): 149-64.

Table 1.2: International Wars in East Asia, 1900-2000 ⁷⁰

War	Year-I	Intensity*
Boxer War	1900	Low
Sino-Russian War	1900	Low
Russo-Japanese War	1904	High
German-Japanese War	1914	Low
Sino-Soviet War	1929	Low
Second Sino-Japanese War	1931	Medium
Third Sino-Japanese War	1937	High
Changkufeng War	1938	Low
Nomonhan War	1939	Medium
Franco-Thai War	1940	Low
Pacific War	1941	High
Korean War	1950	High
Sino-Indian War	1962	Low
Vietnam War	1965	High
Cambodian-Vietnamese War	1975	Low
Sino-Vietnamese War	1979	Medium

Notes: Year-I: Year of war initiation. * Low: $F < 10000$; High: $F \geq 100000$; Medium: $10000 \leq F < 100000$ (F = Combat fatalities).

I conduct two primary case studies based on two wars with major historical consequences – the Second Sino-Japanese War and the Third Sino-Japanese War. These cases use archival records that allow for more accurate measurements of variations in the dependent variable – the instances of pro-war or anti-war decisions or proposals made by decision-makers during the crisis. My selection criterion for the primary cases is of a pragmatic nature: Archival records of decision-making at the highest level on both sides of the conflict must be available and reliable. Two-sided archival evidence allows us to see the interactive decisions and perceptions that drove each crisis into war. The Pacific War also satisfies this criterion, as do (to some extent) the Russo-Japanese War and the

⁷⁰ Correlates of War Inter-State War Data Version 3.0. See Meredith Reid Sarkees, “The Correlates of War Data on War: An Update to 1997,” *Conflict Management and Peace Science* 18, No. 1 (2000): 123-44. The two world wars defined in the dataset are restricted to only the East Asian theater – the German-Japanese War (in World War I) and the Pacific War (in World War II). I have excluded the 1987 Sino-Vietnamese border conflict, as there is no clear evidence that Beijing or Vietnam made a formal decision for war in that case. The Cambodian-Vietnamese War can be split into three sub-wars: Cambodia invaded Vietnamese-held islands in 1975; and Vietnam launched large-scale attacks deep into Cambodia in 1977 and 1978.

Korean War. I choose the Sino-Japanese wars for three reasons. First, they are the most under-studied cases among the five. Hence, a detailed analysis of the decision process in these cases makes a useful documentary contribution. Also, as the existing literature offers limited prior information on these cases, I have no clear prior expectations on whether the details of these cases would fit or contradict the rationalist mechanisms. Finally, this is a pair of cases that involve the same dyad in two different time periods, which allows for sharper intertemporal comparisons. Chapter 3 discusses the methodological strategy in detail. In addition, Chapter 2 also contains a detailed study of Japanese calculations on the eve of the Pacific War based on archival decision records.

My analysis focuses on rationalist mechanisms of war. I assess if the rationalist mechanisms help to shed light on the complex processes that motivate decisions for war. I also highlight specific evidence that suggest alternative causal forces at work. It should be emphasized that the historical analysis is *not* designed to show that rationalist theories of war are correct and other theories are wrong; or that rationalist theories explain more of certain cases than other theories. Indeed, various theories of war can reasonably operate within the commitment or information environments implied in the rationalist mechanisms.⁷¹ Few theories of war are directly orthogonal in that the validity of one must necessarily require the invalidation of the other. War is extremely complex and almost always involves multiple causes. The scientific problem is to show whether a hypothesized mechanism is indeed causal and to measure its causal force on the observed

⁷¹ For instance, a historical decision for war may driven by a commitment problem in conjunction with other facilitating factors, such as economic cycles, the polarity of the regional or international system, or the particularities of specific regime types.

outcome, *controlling for all other potential causes*.⁷² The experimental component of this thesis is designed for this problem. A randomized experiment isolates one hypothesized factor and rules out all other potential factors by design. Hence, the experimental method constructs clean tests of the conflict mechanisms at the heart of the rationalist theories.

In this thesis, the experimental method bridges two powerful research traditions – formal models and historical case studies – that typically have little contact. The models formalize the strategic logic and generate equilibrium predictions. The experiments test the causal effects and detect systematic deviations from the predictions. The historical cases test the realism of the mechanisms with detailed information on decision-making in actual crises. Theoretical insights are compared and cumulated as we iterate between the models, the cases, and the experiments.

⁷² One strategy is show that the observed evidence supports a hypothesized mechanism but contradicts *all other plausible mechanisms*: hence, the hypothesized mechanism is both causal and dominant. But limitations in data and measurement make it difficult to implement this strategy with decisive results.

3. Preview

Chapter 2, “*Rationalist Explanations for War: Experimental and Historical Evidence*,” studies the private-information and commitment-problem explanations for war. The experiments show that the commitment problem has large and robust effects on conflict decisions. The effects of private information are surprisingly subdued in the shadow of shifting power. A historical case study based on the Pacific War probes the external validity of the positive finding and tests for theoretical omissions.

Chapter 3, “*Decisions, Processes and War: Evidence from the Sino-Japanese Wars*,” analyzes the decision processes leading to the Sino-Japanese wars in 1931 and 1937. The analysis clarifies the decision contexts where the hypothesized mechanisms operated (or didn’t). The two case studies intersect archival evidence from both sides of the conflict. Systematic two-sided archival analysis is rare in the study of war. Using archival evidence from both sides, the chapter reveals the interactive dynamics that drove the crises to war.

Chapter 4, “*Enforcement, Signaling and War: Evidence from Ten Experiments*,” describes medium-*n* laboratory experiments and large-*n* Internet-based experiments that test the enforcement and signaling mechanisms. Results show that inadvertent enforcement reduces the incidence of conflict with decisive impact. Peaceful settlements are also more likely when endogenous enforcement is achieved. Meanwhile, experiments on costly signaling reveal an unexpected asymmetry between the behaviors of signalers

and receivers: the logic of sunk-cost signaling works well at the signaler's end but not at the receiver's end. High-resolved signalers are more likely to send sunk-cost signals, but receivers do not necessarily respond according to the signalers' expectation. This asymmetry leads to wasted resources and mutually undesirable outcomes for both parties.

Chapter 5, "*Wars in East Asia: Assessing Old Hypotheses, Inferring New Hypotheses*," provides a comparative study of all international wars in East Asia fought in the 20th century. The chapter exploits within-case and cross-case evidence to evaluate the observable implications of my hypothesized mechanisms. It turns out that the experimental and historical evidence generally converge. The chapter also infers alternative theoretical ideas from the historical cases, mainly: (1) War is more likely when concessions made in one bargaining dyad negatively affects relative power in another bargaining dyad. (2) War is more likely when conquest is suddenly easy and cheap due to an exogenous change in the international environment. (3) Many East Asian wars have roots in two types of domestic principal-agent problems. (4) Audience costs can promote war decisions. Domestic and international audience costs matter, but so do *personal audience costs* – the political costs to a career politician or bureaucrat for opposing the position taken by powerful colleagues or superiors.

Chapter 6, "*Conclusion*," crystallizes the knowledge generated from the earlier chapters. It evaluates the strength of the key findings and their implications based on the observational and experimental evidence accumulated in the dissertation.

Chapter 2

Rationalist Explanations for War: Experimental and Historical Evidence

1. Introduction

Private information and the commitment problem are the fundamental causes of war in the rationalist literature. Despite their theoretical prominence, there are few empirical tests of rationalist explanations for war: Do private information and the commitment problem have causal effects on the risk of conflict? How large are the effects? How much more peaceful is a world with credible enforcement or public information compared to a world without?

I offer empirical evidence on these questions in this chapter. I use a randomized laboratory experiment to study the causal effects of private information and the commitment problem on conflict. I translate the formal theory to a laboratory setting with real people, with some useful simplifications to make the theory experimentally testable. I also conduct a case study of the decision process in a historical crisis to assess the realism of my positive experimental findings.

The commitment problem sharply increased the incidence of conflict in the experiment. Despite the costs of war, decision-makers with a commitment problem gravitated towards war. The causal effect was robust over repeated rounds in both sequential and

simultaneous-moves interactions. The effect of private information, however, was surprisingly subdued in the experiment: the overall incidence of conflict in the shadow of shifting power was almost the same with or without private information.

I supplement my analysis with a case study of Japan's policy deliberations leading to the Pacific War. The case includes a shock that created a sharp and unambiguous perception of an impending power shift. The shock came when the United States imposed a total oil embargo on oil-dependent Japan. The embargo would sharply reduce Japan's future relative power. Based on the decision-making records before and after the embargo, I show that Japanese leaders quickly shifted to an active prowar position against the U.S. when they were suddenly confronted by a commitment problem induced by the shadow of shifting power.

The formal, experimental and historical evidence point in the same direction: The logic of the commitment problem has a strong influence on the decision for war. The mix of methods also shows how the experimental method connects two research traditions – formal models and historical studies – that typically have little contact.

Related literature

This study is the first to jointly evaluate Fearon's rationalist explanations for war in the laboratory.¹ Using a 2x2 experimental design, I explore whether the commitment

¹ James Fearon, "Rationalist Explanations for War," *International Organization* 49, No. 3 (1995): 379-414.

problem affects the risk of conflict in different information environments, and whether private information has a significant effect on conflict with and without a commitment problem. The 2x2 design is useful since private information and commitment problems often arise jointly rather than separately.

In international relations, three earlier game-theoretic experiments are closest to this work. Two of them relate to the role of asymmetric information. Butler, Bellman and Kichiyev examined whether a status-quo advantaged actor can achieve a better bargain under different information conditions, and found that the status-quo advantaged actor gets more when there is private information.² Tingley and Wang experimentally manipulated the cost of backing down to examine its effects on how players update their beliefs in a crisis bargaining game with incomplete information.³ The study found that players update their beliefs in ways that are generally consistent with the perfect Bayesian equilibrium. Finally, Dustin Tingley studied the commitment problem by manipulating experimental treatments based on the probability of repeated interactions (“shadow of the future”).⁴ The commitment problem is induced at two different levels with two different probabilities of repeated play (0.3 and 0.7). The study is insightful and it shows how bargaining behavior changes along the shadow of the future. However, due to its specific research focus, the experiment does not set up a control condition that

² Christopher Butler, Mary Bellman, and Oraz Kichiyev, “Assessing Power in Spatial Bargaining: When Is There Advantage to Being Status-Quo Advantaged?” *International Studies Quarterly* 51, No. 3 (2007): 607-23.

³ Dustin Tingley and Stephanie Wang, “Belief Updating in Sequential Games Of two-Sided Incomplete Information: An Experimental Study of a Crisis Bargaining Model,” *Quarterly Journal of Political Science* 5, No. 3 (2010): 243-55.

⁴ Dustin Tingley, “The Dark Side of the Future: An Experimental Test of Commitment Problems in Bargaining,” *International Studies Quarterly* 55, No. 2 (2011): 521-44.

completely eliminates the prospect of a commitment problem. Since a positive probability of repeated interaction exists in both experimental groups, subjects may perceive a potential commitment problem in either experimental condition, albeit at two different probabilities (0.3 and 0.7). In this chapter, I construct a pure commitment problem in the treatment condition under which players confront the commitment problem with absolute certainty. For the control group, I construct a perfect-enforcement device that eliminates all possibilities of renegeing. The device shuts off the commitment problem completely and facilitates a clean identification of its effect on conflict.

In economics, the most relevant studies involve bargaining experiments such as the ultimatum-game experiment.⁵ There is a sizeable literature on the effects of information asymmetry on bargaining. In general, there are more bargaining breakdowns when information is asymmetric, but the support for particular equilibrium predictions is sometimes weak.⁶ The effect and significance of asymmetric information appear to depend on the particularities of the information environment and the specific form of

⁵ Ultimatum game experiments, in which a proposer interacts with a responder over the division of a prize, have consistently shown a convergence to a 50-50 fair split that contradicts rational-choice predictions based on subgame perfection. This result has been replicated across different cultures, experience levels, and stake sizes. However, while the experiments reported here involve bilateral bargaining over the division of a prize, their design and game models diverge from the ultimatum-game literature, with substantially different results. See Alvin Roth, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir, "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review* 81, No. 5 (1991): 1068-95; Miguel Costa-Gomes and Klaus Zauner, "Ultimatum Bargaining Behavior in Israel, Japan, Slovenia, and the United States: A Social Utility Analysis," *Games and Economic Behavior* 34, No. 2 (2001): 238-69; Robert Slonim and Alvin Roth, "Learning in High Stakes Ultimatum Games: an Experiment in the Slovak Republic," *Econometrica* 66, No. 3 (1998): 569-96; Lisa Cameron, "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia," *Economic Inquiry* 37, No. 1 (1999): 47-59.

⁶ See Robert Forsythe, John Kennan, and Barry Sopher, "An Experimental Analysis of Strikes in Bargaining Games with One-Sided Private Information," *American Economic Review* 81, No. 1 (1991): 253-78.

interaction. The experiments here focus specifically on asymmetric information on the cost of war in the shadow of shifting power.⁷

On the other hand, there are few experiments that focus on the commitment problem in bargaining games. The closest strands of literature in experimental economics are three. The first consists of experiments that focus on the communication of commitment, particularly the making of a promise and its generally positive effect on the level of cooperation.⁸ These experiments are mostly silent on the commitment problem.⁹ The second strand focuses on experimental manipulations of the probability of repeated future interactions (“shadow of the future”).¹⁰ These manipulations can induce a commitment problem, though of a different form from that created by the shadow of shifting power. A closely related work is by McBride and Skaperdas, in which the winner of a period-1

⁷ This allows us to jointly evaluate the effect of asymmetric information and the commitment problem with the same experimental design and game-theoretic model, as discussed earlier. Future work may test the effect of asymmetric information in other forms. For surveys of bargaining experiments conducted across information environments with different specific features, see John Kennan and Robert Wilson, “Bargaining with Private Information,” *Journal of Economic Literature* 31, No. 1 (1993): 45-104; Alvin Roth, “Bargaining Experiments,” in *The Handbook of Experimental Economics*, ed. John Hagel and Alvin Roth (Princeton, NJ: Princeton University Press, 1995), 253-348; Charles Plott and Vernon Smith, *Handbook of Experimental Economics Results*, Vol. 1 (Amsterdam: North-Holland, 2008).

⁸ Tore Ellingsen and Magnus Johannesson, “Promises, Threats and Fairness,” *Economic Journal* 114, No. 495 (2004): 397-420; Gary Charness and Martin Dufwenberg, “Promises and Partnership,” *Econometrica* 74, No. 6 (2006): 1579-601; Christoph Vanberg, “Why Do People Keep Their Promises? An Experimental Test of Two Explanations,” *Econometrica* 76, No. 6 (2008): 1467-80.

⁹ Gurtler and Harbring mentioned the commitment problem in a labor-management context, but the experiment focused specifically on how agent work efforts vary in response to whether the principal can commit to a credible communication of performance information. Oliver Gurtler and Christine Harbring, “Feedback in Tournaments under Commitment Problems: Theory and Experimental Evidence,” *Journal of Economics and Management Strategy* 19, No. 3 (2010): 771-810.

¹⁰ Experimental economists have studied the relationship between cooperation and the shadow of the future, as have scholars of international relations. See, e.g., Pedro Dal Bó, “Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games,” *American Economic Review* 95, No. 5 (2005): 1591-604; John Duffy and Jack Ochs, “Cooperative Behavior and the Frequency of Social Interaction,” *Games and Economic Behavior* 66, No. 2 (2009): 785-812; Tingley, “Dark Side.”

conflict would receive additional payoffs in future periods of interaction. The study found that conflict propensities increase as the continuation probability increases.¹¹ The third strand of literature involves experimental manipulations of bargaining power. These manipulations involve variations in the payoffs for conflict between players, which translate to variations in their bargaining strength.¹² While my experiments also involve variations in bargaining power, I focus on commitment elicitation, enforcement conditions, and intertemporal power shifts, which are not emphasized in this particular literature.

I focus specifically on a commitment problem in the shadow of shifting power, which is believed to be a potent cause of many international wars in history.¹³ To my knowledge, there is no published work in experimental economics that directly isolates the commitment problem in the shadow of shifting power under different information conditions. In this context, the experiments reported here are relatively unique and may contribute to our empirical understanding of rationalist explanations for war.

¹¹ Michael McBride and Stergios Skaperdas, "Conflict, Settlement, and the Shadow of the Future," Working Paper (University of California, Irvine, 2009).

¹² Ken Binmore, Peter Morgan, Avner Snaked, and John Sutton, "Do People Exploit Their Bargaining Power? An Experimental Study," *Games and Economic Behavior* 3, No. 3 (1991): 295–322; Sven Fischer, Werner Güth, and Kerstin Pull, "Is There As-If Bargaining?" *Journal of Socio-Economics* 36, No. 4 (2007): 546–60; Nejat Anbarci and Nick Feltovich, "How Responsive Are People to Changes in Their Bargaining Position? Earned Bargaining Power and the 50–50 Norm," Working Paper (Monash University, 2011).

¹³ Alfred Vagts, *Defense and Diplomacy: The Soldier and the Conduct of Foreign Relations* (New York: King's Crown Press, 1956), chapter 8; Stephen Van Evera, *Causes of War: Power and the Roots of International Conflict* (Ithaca, NY: Cornell University Press, 1999), chapter 4. See also Robert Powell, *In the Shadow of Power: States and Strategies in International Politics* (Princeton, NJ: Princeton University Press, 1999); and Robert Powell, "War as a Commitment Problem," *International Organization* 60, No. 1 (2006): 169–203.

2. Model Predictions

In this section, I revise and simplify Fearon's models to anchor my experimental design.¹⁴

I model the information problem and the commitment problem jointly in their most basic form. My approach is to distill the essence of Fearon's models without the complications and confounds that would be generated through a straightforward replication of those models. My goal is to locate the simplest possible model that combines both the private-information problem and the commitment problem under one structure.

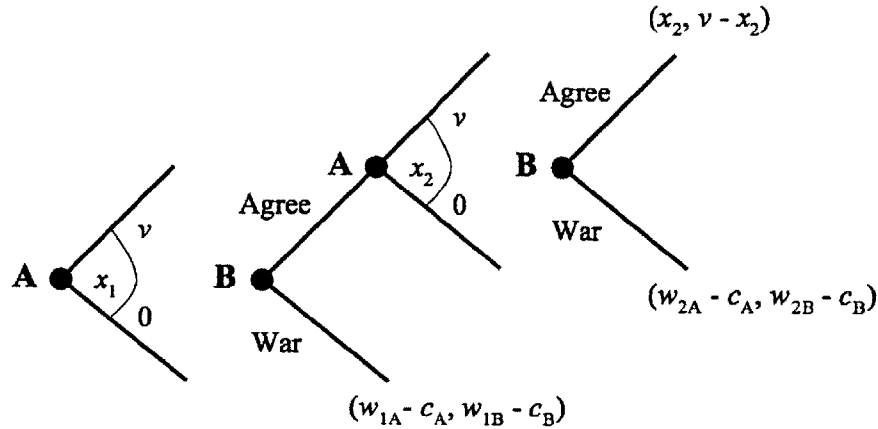
The model has three key features. First, the model combines the information and commitment problems in a two-period model, instead of a single-period model for the information problem and an infinite-period model for the commitment problem, as in Fearon's work. A two-period model provides the simplest possible structure that generates the commitment problem, but avoids the practical difficulties of implementing infinite-period bargaining with human subjects. The second feature involves the use of "perfect enforcement." The enforcement condition is defined such that Player A cannot make any change to its Stage-1 demand during Stage 2. Hence, the commitment problem is completely shut off in the enforcement condition. The use of a perfect-enforcement device eliminates any noise that might arise from the expectation of even a minuscule possibility of renegeing. Finally, I black-box the war technology. Players end up with an

¹⁴ Fearon, "Rationalist Explanations." There are other bargaining models that can also be reformulated to incorporate both private information and the commitment problem, e.g. Michelle Garfinkel and Stergios Skaperdas, "Conflict without Misperceptions or Incomplete Information: How the Future Matters," *Journal of Conflict Resolution* 44, No. 6 (2000): 793-807. I use Fearon's work partly because it is simpler and more general relative to other models; and partly because of its centrality in the rationalist literature on war. The importance of the shadow of shifting power in historical wars is also noted in the previous paragraph.

absolute payoff if war is chosen, without having to deal with expected value calculations involving the probabilities of winning wars. With these changes, the model allows us to focus sharply on the commitment problem in its cleanest form. The experiment becomes tighter in design as well as simpler for the players to play.

Consider a two-stage game in which two players bargain over a prize with the value v . In Stage 1, Player A makes a demand $x_1 \in [0, v]$. Player B observes x_1 and decides whether to agree or to fight. If B fights, the game ends with payoff $(w_{1A} - c_A, w_{1B} - c_B)$, with $w_{1A} = w_{1B}$, where w_{1i} and c_i are Player i 's war payoff in Stage 1 and cost of war respectively. If B agrees, the game enters Stage 2 in which A confirms its demand $x_2 \in [0, v]$. B observes x_2 and decides whether to agree or to fight. If B agrees to x_2 , the game ends with payoff $(x_2, v - x_2)$. If B fights, the game ends with payoff $(w_{2A} - c_A, w_{2B} - c_B)$, where $v = w_{1A} + w_{1B} = w_{2A} + w_{2B}$ and $w_{2A} > w_{1A} > w_{2B} > c$. Thus, A and B have equal bargaining power in Stage 1, but in Stage 2 the bargaining power shifts in A's favor.

Figure 2.1: Crisis Bargaining Game



In the experiment, v is fixed at 10 and c_i is fixed at 2. For the war payoffs, $w_{1A} = w_{1B} = 5$, $w_{2A} = 7$ and $w_{2B} = 3$. These parameter values are chosen to minimize noise in the experiment. The payoffs are set within the integer range of $[0, 10]$ to simplify the mathematics for subjects and to reduce the likelihood of calculation error. The parameter c_i is set at 2 as it is the highest integer value by which all payoffs in the model remain in positive integer domain, given $w_{2A} = 7$ and $w_{2B} = 3$. The values of w_{2A} and w_{2B} are chosen to achieve equidistance from the midpoint. Note that other combinations of integer values will complicate the experiment by either breaking the equidistance or generating potential payoffs that are zero or negative, given the requirements of the model.¹⁵

¹⁵ The exceptions are the combinations of parameter values (1) with $w_{2A} = 6$ and $w_{2B} = 4$, which creates a minimal shift in bargaining power across the two stages; and (2) with c_i set at 0 or 1, which would render war costless or extremely cheap. Future work may investigate the special cases not addressed in this experiment.

Next, I analyze the particular incentive structure implemented in the experiment to derive theoretical predictions for the outcomes.

Public Information with/without Enforcement

Suppose that there is no private information and all payoffs are known to the players. In particular, both players know that $c_A = c_B$.

In the enforcement condition, A cannot change its Stage-1 demand in Stage 2. Hence, $x_1 = x_2$. In the subgame perfect equilibrium A will demand $x_1 = x_2 = v - w_{1B} + c_B$ given B's reservation level ($w_{1B} - c_B$) at Stage 1. Based on the values in the experiment, B will accept the demand $x_1 = x_2$ since $v - x \geq w_{1B} - c_B$.¹⁶ Thus:

Prediction 1: War will be avoided if there is enforcement in the public-information condition.

In the no-enforcement condition, A can change its Stage-1 demand in Stage 2. In Stage 1, B is choosing between x_1 and war. Since there is no enforcement, A's x_1 is not credible and B will only consider x_2 in Stage 1. A will demand $x_2 = v - w_{2B} + c_B$ in Stage 2, based on the amount that makes B just willing to accept. By backward induction, B will choose war in Stage 1 in the subgame perfect equilibrium since $w_{1B} - c_B > v - x_2$ in Stage 2. Thus:

¹⁶ That B will not fight if the utility of war equals the utility of agreement is assumed.

Prediction 2: War will occur with certainty if there is no enforcement in the public-information condition.

Private Information with/without Enforcement

Suppose there is private information on the costs of war c_A and c_B , and that c_A and c_B are drawn from a discrete uniform distribution on integers $[0, v]$.

In the enforcement condition, $x_1 = x_2$ and B will choose war if $w_{1B} - c_B > v - x_1$. Hence, in any equilibrium the probability of war will be $\Pr(w_{1B} - c_B > v - x_1)$. Specifically, B will choose war if $5 - c_B > 10 - x_1$, and in any equilibrium the probability of war will be $\Pr(c_B < x_1 - 5)$. As such, A's expected utility for demanding x_1 is $u_A(x_1) = \Pr(c_B < x_1 - 5)(5 - c_A) + (1 - \Pr(c_B < x_1 - 5))(x_1)$. The experiment restricted c_A and c_B to be integers drawn from a discrete uniform distribution on $[0, 4]$. By direct proof, computing the maximized $u_A(x_1)$ in all combinations of c_A and c_B within $[0, 4]$ shows that the risk of war is always positive. The result and intuition are similar to that in Fearon's take-it-or-leave-it game.¹⁷ A faces a trade-off between the size of its demand x_1 and the risk of war generated by the unknown c_B , with a higher demand leading to a better potential bargain but also a higher risk of war. Thus:

Prediction 3: There is always a positive risk of war in the private-information condition with enforcement.

¹⁷ Fearon, "Rationalist Explanations," 410-11.

In the no-enforcement condition, the equilibrium outcome will converge to the subgame perfect equilibrium outcome analyzed in the public-information case based on the values in the experiment. This is because B, who only considers x_2 in Stage 1, will choose war since the payoff for war in Stage 1 ($5 - c_B = 3$) will never be less than the payoff for agreement ($10 - x_2$) in Stage 2, given that the value of x_2 demanded by A will be based on B's reservation level ($3 - c_B \leq 3$), with $c_B \geq 0$.¹⁸ Thus:

Prediction 4: War will occur with certainty in the private-information condition without enforcement.

3. Experimental Design

The experiment uses a 2x2 factorial design with controlled variation in information and enforcement (Table 2.1). The first treatment variable captures the difference in information conditions; the second treatment variable captures the difference in enforcement conditions.

Table 2.1: Experimental Design

	Public Information	Private Information
Enforcement	p_1	p_2
No Enforcement	p_3	p_4

Notes: p_i = proportion of war outcomes in cell i . Table 2.2 reports the results.

¹⁸ The case of war being avoided due to the possibility of $c_B = 0$ is trivial.

Experimental conditions

- Enforcement: Player A's Stage-1 demand x_1 is directly equated into its Stage-2 demand x_2 . In other words, A chooses x_1 in Stage 1 but has no influence over x_2 in Stage 2 once x_1 is chosen. This is perfect enforcement and it eliminates the commitment problem completely. Note that the shadow of shifting power remains, but the commitment problem disappears.
- No enforcement: A can demand $x_2 \neq x_1$ in Stage 2. Hence, the commitment problem remains in the shadow of shifting power.
- Public information: The cost of war $c_A = c_B = 2$ is known to all players. Each player is told: "You know your Opponent's cost of war. Your Opponent knows your cost of war." Players are informed that their cost of war will be fixed for all rounds in the game.
- Private information: Player i knows its own cost of war ($c_i = 2$), but does not know the cost of war for its opponent. Each player is told: "You do NOT know your Opponent's cost of war. Your Opponent does NOT know your cost of war." Players are told that "[t]o generate the costs of war, the computer will assign one of the values {0, 1, 2, 3, 4} to you. Then, it will assign one of the values {0, 1, 2,

3, 4} to your Opponent.”¹⁹ To control for the effect of c_i by design, all players are assigned with $c_i = 2$.²⁰ Players are informed that once their c_i parameter is assigned in Round 1, it will be fixed for all rounds in the game.

Assignment

The subject pool was randomly divided into the public-information and private-information groups. Subjects in each group were then randomly allocated to either the enforcement condition or the no-enforcement condition. In each round, subjects were randomly assigned as decision-makers for either “Country A” or “Country B” and they played with a randomly assigned opponent within the same information/enforcement conditions. Subjects stayed in the same information condition for Rounds 1-10. A sudden switch in the enforcement condition occurred at Round 6.

Design features

The experiment included three design features. First, I used a crossover design in which half the subject pool played five rounds with enforcement before switching into five rounds with no enforcement; the other half played in the reverse sequence. This provides

¹⁹ In other words, all they know is that the c_i of their opponent will be one of the integers in the uniform distribution $[0, 4]$. Hence, this experiment evaluates the specific case of private information on non-catastrophic costs of war.

²⁰ The instructions explicitly state that the computer will assign the c_i parameter. However, the parameter is *not randomly assigned* but is universally fixed at $c_i = 2$. The purpose is to keep c_i constant for all players, but induce the belief that the opponent’s c_i will be drawn from a uniform distribution based on the model in Section 2. A test question at the end of the instructions confirmed that the subjects were indeed thinking in terms of a uniform distribution. See Test Question 4 in Appendix A1.

a severe test of the experimental results with a sudden exchange of treatments between two treatment groups.²¹ The crossover design allows me to test the robustness of my findings with a higher hurdle than single-trial or constant-treatment experiments. Second, I ran the experiment with both a sequential game structure (rounds 1-10) as well as a game structure that allows for simultaneous moves (rounds 11-15). This design feature provides another stress-test of the results, and it sheds light on whether the treatment effects are generalizable and robust beyond pure sequential bargaining. Third, the random matching of players was subject to a “stranger-matching” requirement that no two players be matched with each other twice in the same scenario. Reciprocity effects that might arise with repeated rounds were thus controlled by design.

Implementation

70 students from Harvard University were recruited through the Harvard Decision Science Laboratory. The experiment was programmed on the z-Tree platform which allowed subjects to interact anonymously through computer terminals.²² The experiment was conducted in three sessions at Harvard University in February 2011. Each session had 22 to 24 subjects. Each subject participated in only one session.

²¹ The crossover design is often practiced in clinical research. The general ideas of “switching replications” and “treatment removal at a known time” are sketched out in Donald Campbell and Julian Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963), 50-7, and clarified in William Shadish, Thomas Cook, and Donald Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Boston: Houghton Mifflin, 2002), 188-95.

²² Urs Fischbacher, “Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments,” *Experimental Economics* 10, No. 2 (2007): 171-8.

Participants began the experiment by reading the instructions on their computer screens. They were placed in the role of a national decision-maker facing an international crisis, in which their country was bargaining with another country for a valuable prize. I framed the situation as one involving crisis bargaining and war to enhance the relevance of the experiment, since my research question deals directly with rationalist explanations for war.²³ The war framing also helps to strengthen the inducement of utilities by reducing the potential other-regarding behavior widely documented in many bargaining experiments.²⁴ This is important since rationalist explanations for war do not assume other-regarding behavior.

Special attention is paid in ensuring that subjects fully understood the game (see Appendix A1). The game was carefully explained, with an emphasis on the payoff difference between Stages 1 and 2. In Stage 1, both countries are equally powerful. If they fight a war, each country will be able to seize 50% of the prize (5 points) for itself. In Stage 2, Country A becomes more powerful than Country B. If they fight a war, Country A will be able to seize 70% of the prize (7 points) and Country B 30% of the prize (3 points). Since war is costly, both countries will also lose points based on its own

²³ Alternatively, one can also frame the experiment to participants as an abstract, context-free game. This may remove the possibility that the experimental results are conditional on subjects thinking about crisis bargaining and war. The tradeoff, however, is the possibility that the results will not hold when decision-makers are indeed thinking in those terms. Currently, there is no consensus among political experimentalists on the problem of abstraction versus contextualization, except that one should choose the option that connects best to the specific research question. See Eric Dickson, "Economics Versus Psychology Experiments: Stylization, Incentives, and Deception," in *Cambridge Handbook of Experimental Political Science*, ed. James Druckman, Donald Green, James Kuklinski and Arthur Lupia (New York: Cambridge University Press, 2011), 58-69.

²⁴ I thank an anonymous reviewer for emphasizing this point.

cost of war. The instructions were written in neutral language, with questions at the end to check subject understanding.

The experiment was sequenced as follows:

- 1) Participants started by playing the crisis bargaining game under a crossover design. Each participant played five rounds in the enforcement condition and five rounds without enforcement. They stayed in the same information treatment for all 10 rounds.
- 2) Thereafter, participants played five rounds of the crisis bargaining game with an option for simultaneous moves.²⁵
- 3) Participants played one round of the game in Part 2 under the no-enforcement condition with timer treatments (see Observation 4 in Section 5).
- 4) The experiment ended with a risk-aversion test in the final round that measured their risk preferences.

Participants were paid solely based on their performance in the game. I used a random payment mechanism in which the computer randomly chose 9 rounds out of 17 rounds, and paid \$0.50 for each point earned in the 9 chosen rounds. The payment scheme was

²⁵ Subjects in Session 1 played three rounds in Part (2) followed by three rounds in Part (3). This design feature allows us to do a robustness check for Observation 4.

designed to increase the saliency of payoffs for each round. Each participant earned \$19.87 on average and each session lasted for an hour. No practice rounds were implemented prior to the experiment. The supplementary materials reproduce the experimental instructions and risk-aversion test (Appendix A1).

4. Results

In this section, I will describe the results with straightforward statistics, followed by robustness checks with different model specifications to estimate the significance of the results.

Result 1: The existence of a commitment problem causes a large increase in the number of wars.

The incidence of war (war outcomes as a percentage of all outcomes) is 59% in the no-enforcement condition (with commitment problem) compared to 20% in the enforcement condition (without commitment problem). Table 2.2 shows the incidence of war across the four experimental conditions in the first ten rounds. Table 2.3 shows the average number of wars per player in each condition. In the public-information condition, the average number of wars per player is more than thrice as high in the no-enforcement group than in the enforcement group. In the private-information condition, the average in the no-enforcement group is more than twice as high as the average in the enforcement group. Both the differences in the number of wars per player are statistically significant at

$p < 0.0001$ (two-tailed t-test: $n = 72$ for public-information condition; $n = 62$ for private-information condition).²⁶ Player average is used as the unit of observation since round outcomes may not be independent. Round outcomes will be analyzed based on different model specifications later in this section.

Table 2.2: Incidence of War Across Conditions

	Public Information	Private Information
Enforcement	0.17	0.23
No Enforcement	0.63	0.55

Note: Maximum possible value for the incidence of war is 1.00 (100%).

Table 2.3: Average Number of Wars Per Player

Session	<i>Public-Information Condition</i>		<i>Private-Information Condition</i>	
	Enforcement	No-Enforcement	Enforcement	No-Enforcement
1	1.00	2.17	1.33	2.67
2	0.83	3.67	1.00	3.20
3	0.67	3.67	1.33	2.33
Overall	0.83	3.17	1.23	2.74

Notes: Maximum possible value is 5.00 for each player. Half the subject pool was randomly assigned to 5 rounds with enforcement followed by 5 rounds with no enforcement; the other half was assigned in the reverse sequence.

²⁶ The differences remain significant at $p < 0.01$ (two-tailed t-test: $n = 36$ for public-information condition; $n = 31$ for private-information condition) if the observations are restricted to only those before the crossover at Round 6.

The results support the model predictions. The behavioral data show strong decisional tendencies towards war when there is a commitment problem in both public and private-information conditions (Predictions 2 and 4), and towards a peaceful bargain when there is enforcement in the public-information condition (Prediction 1).

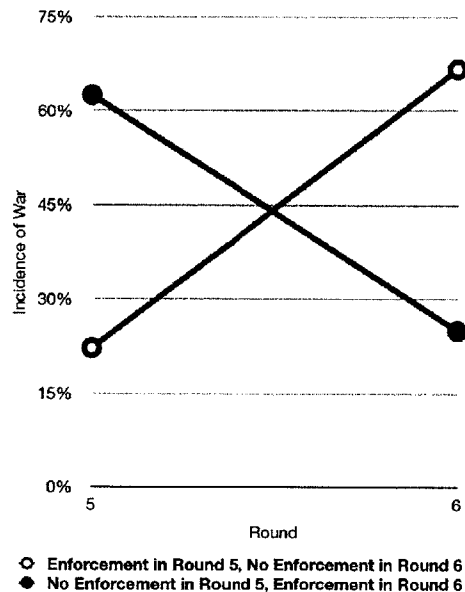
If the causal effects are robust, we should also expect an *increase* in the number of wars when players switch from the enforcement condition to the no-enforcement condition, as well as a *decrease* when they switch from no-enforcement to enforcement. This is indeed what we observe.

Result 2: A sudden introduction of a commitment problem causes a sharp rise in the number of wars. A sudden introduction of enforcement causes a sharp fall in the number of wars.

Figure 2.2 shows the differences in the incidence of war across Rounds 5 and 6, where I introduced a sudden exchange of treatments between the experimental groups. The pattern is sharp. The crossover test reversed the direction of the treatment effects and revealed a striking symmetry in the effects measured. The symmetry is visualized in Figure 2.2. When a commitment problem was introduced, the percentage of war outcomes increased from 22% (Round 5) to 67% (Round 6) ($n = 36$: two-tailed test of proportion, $p = 0.0073$; Mann-Whitney test, $p = 0.0082$). The removal of the commitment problem caused it to fall from 63% to 25% ($n = 32$: two-tailed test of proportion, $p = 0.0325$; Mann-Whitney test, $p = 0.0353$). Splitting the sample by information conditions,

introducing a commitment problem in the public-information condition increased the incidence of war outcomes from 33% to 67%, while the removal of the commitment problem caused it to fall from 56% to 11%.²⁷ In the private-information condition, war incidence increased from 11% to 67% when the commitment problem was introduced, and fell from 71% to 43% when the commitment problem was removed.²⁸ The result provides additional support for Predictions 1, 2 and 4.

Figure 2.2: Incidence of War in Rounds 5-6



²⁷ Between-period change in dyadic outcome with a switch from enforcement to no-enforcement ($n = 18$: two-tailed test of proportion, $p = 0.1573$; Mann-Whitney test, $p = 0.1693$). Between-period change with a switch from no-enforcement to enforcement ($n = 18$: two-tailed test of proportion, $p = 0.0455$; Mann-Whitney test, $p = 0.0519$). Within-subject difference in player war outcome (two-tailed paired t-test, $p = 0.0017$, $n = 36$).

²⁸ Between-period change in dyadic outcome with a switch from enforcement to no-enforcement ($n = 18$: two-tailed test of proportion, $p = 0.0156$; Mann-Whitney test, $p = 0.0188$). Between-period change with a switch from no-enforcement to enforcement ($n = 14$: two-tailed test of proportion, $p = 0.2801$; Mann-Whitney test, $p = 0.2980$). Within-subject difference in player war outcome (two-tailed paired t-test, $p = 0.0008$, $n = 32$).

While the observations agree with Predictions 1, 2 and 4, Results 1 and 2 do not tell us whether the predictions were observationally true because of the equilibria-generating calculations underlying the model, or because of some other systematic factors that were not captured by the model. Hence, I test two observable implications derived from the strategic calculations behind Predictions 1, 2 and 4. If the hypothesized mechanisms are correct, we should expect that in the no-enforcement condition (1) players with a relative power shift *against* their favor will choose to fight in Stage 1; and (2) players with a relative power shift *in* their favor will renege on their Stage-1 agreement in Stage 2. These are indeed what we observe.

Result 3: In the no-enforcement world, players with a power shift in their favor reneged on their agreements 93% of the time.

Figure 2.3 shows the percentage of players who reneged on their agreements across ten rounds in the no-enforcement condition.²⁹ A player is coded as a “reneger” if he or she made a Stage-2 offer that was lower than his or her Stage-1 offer. Figure 2.4 shows the percentage of wars that occurred in Stage 1 without enforcement. Wars in Stage 1 comprised 89% of all wars fought in the no-enforcement condition.³⁰ Together these observations suggest that the results supporting Predictions 1, 2 and 4 were motivated by the strategic calculations underlying the model.

²⁹ Total number of dyadic observations with no war in stage 1 in the no-enforcement treatment = 82. In each round, there was a random matching of the participants into different conflict dyads (within the same treatment condition) subject to the “stranger-matching” requirement. See Section 3.

³⁰ Total number of dyadic observations with a war outcome in the no-enforcement treatment = 103.

Figure 2.3: Percentage of Reneged Offers (No Enforcement)

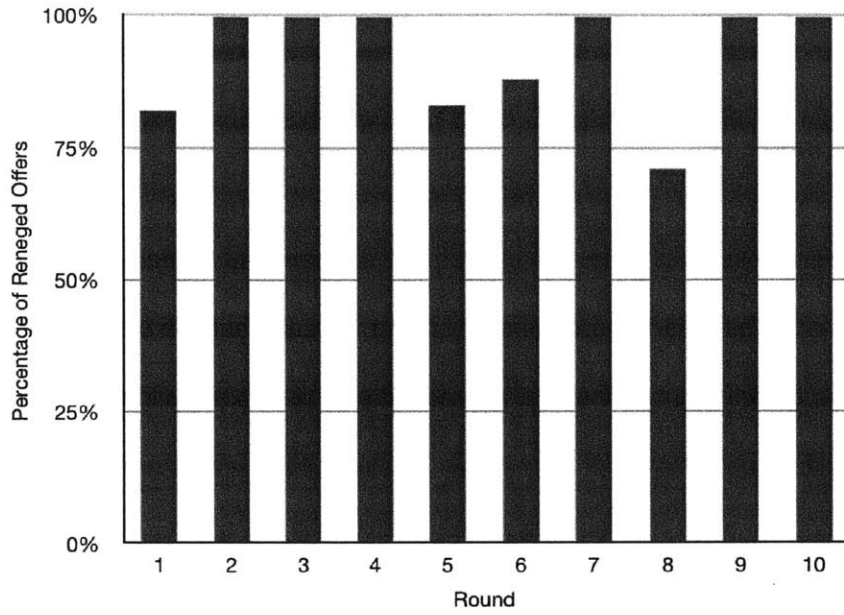
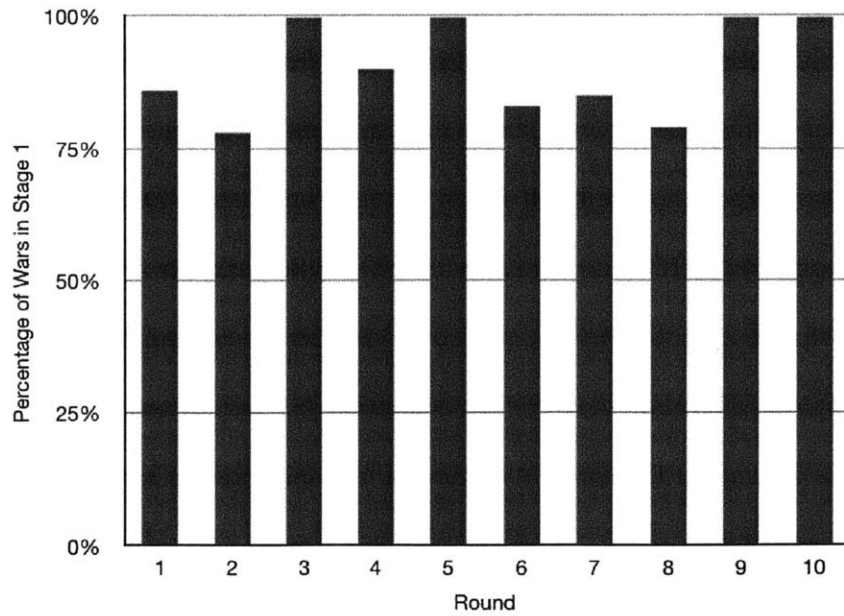


Figure 2.4: Percentage of Wars in Stage 1 (No Enforcement)



Player responses to reneged agreements in Stage 2 also provide an indicative probe of player rationality. Faced with an opponent who reneged on the earlier agreement, players could choose to fight in Stage 2 out of emotions or miscalculation, or they could rationally assess the Stage-2 offer and accept it if it provided a higher payoff than their war payoff (1 point) in Stage 2. Player responses could thus be separated into the non-rationality category in the former and the rationality category in the latter. 88% of player responses fell indicatively into the rationality category, based on the percentage of peaceful responses to reneged offers paying more than the war payoff across 10 rounds. In particular, instances of non-rationality, which occurred in the initial rounds, vanished into full rationality (100%) in the later rounds of the game.³¹

Result 4: The existence of private information has no significant effect on the number of wars in the case of private information on non-catastrophic costs of war.³²

Pooling the data across the enforcement and no-enforcement conditions, the incidence of war in the private-information condition (39%) is almost the same as that in the public-information condition (40%) (two-tailed test of proportion, $p = 0.8374$, $n = 347$). A comparison across the columns in Table 2.2 shows that the incidence of war under private information is about six percentage points higher than that under public information in the enforcement condition (two-tailed test of proportion, $p = 0.3033$, $n =$

³¹ These are Rounds 3-5 before the treatment crossover and Rounds 9-10 after the crossover.

³² In the experiments, players in the private-information condition do not know their opponent's cost of war. However, they know that the opponent's cost of war will be contained within a "non-catastrophic" range represented by the uniform distribution [0, 4].

173). The statistic is about eight percentage points lower in the no-enforcement condition (two-tailed test of proportion, $p = 0.2503$, $n = 174$). The differences remain insignificant in the two-tailed t-test with player averages as observations ($p = 0.0847$ and $p = 0.1970$ respectively; $n = 67$ in each condition). For robustness check, I use logit and linear probability models to analyze the round outcome. Each model uses one dyadic observation per round and session, with robust standard errors corrected for clustering at the subject level. The dependent variable is the decision for war in each round. The treatment variables are binary variables with the value 1 if there is enforcement (for the enforcement treatment) or public information (for the information treatment), and 0 if otherwise. I use a series of different model specifications with and without round and session fixed effects, and with and without an interaction term for the treatment variables, a control for the size of Player A's initial offer, and a control for individual risk preference. Risk preference is measured on a summed score based on the decisions made in the risk-aversion game at the end of the experiment: the higher the score, the greater the individual willingness to take risk. Appendix A1 describes the risk-aversion game.

Across different model specifications, the information treatment has no significant relationship ($p > 0.30$) with the decision for war. By contrast, the coefficient for the enforcement treatment is negative and highly significant ($p < 0.001$) across all model specifications, suggesting a strong negative relationship between enforcement and the decision for war. The negative and significant coefficient ($p < 0.01$) for the initial-offer variable suggests that Player B is less likely to decide for war when Player A makes a higher initial offer. The interaction term for the treatment variables is insignificant ($p \geq$

0.20) across all models. Table 2.4 shows a set of logit estimates based on different controls with round and session fixed effects. Estimations based on the linear probability model yield similar results.

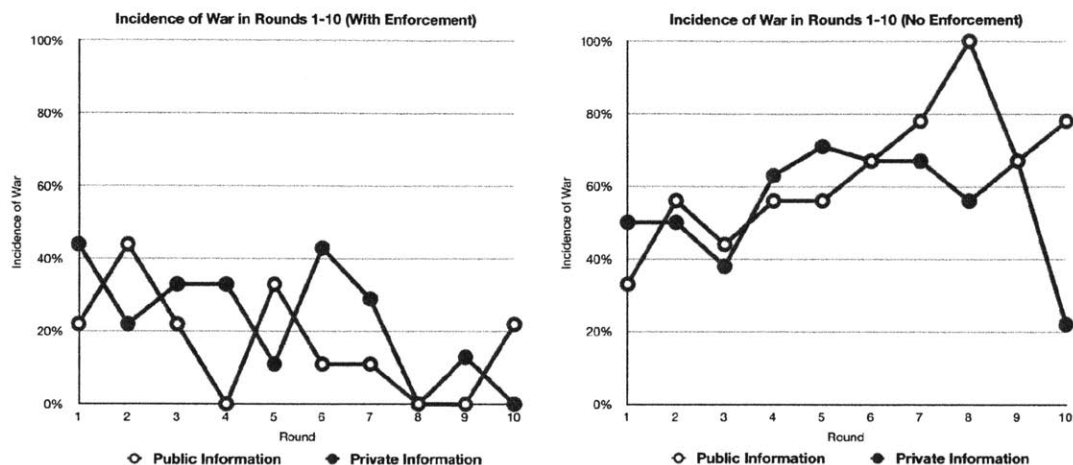
Table 2.4. Logit Estimates of Determinants for the Decision for War (Rounds 1-10)

	(1)	(2)	(3)	(4)
Enforcement	-1.809*** (0.296)	-1.422*** (0.395)	-2.434*** (0.333)	-2.088*** (0.491)
Information	0.044 (0.294)	0.350 (0.397)	0.001 (0.325)	0.270 (0.451)
Enforce*Info		-0.772 (0.601)		-0.653 (0.620)
Initial offer			-0.587** (0.216)	-0.573** (0.216)
Risk preference			0.078 (0.094)	0.075 (0.091)
Constant	0.088 (0.548)	-0.076 (0.567)	2.318 (1.195)	2.125 (1.214)
Round and session fixed-effects	Yes	Yes	Yes	Yes
Prob>Chi ²	0.000	0.000	0.000	0.000
Log-likelihood	-200.600	-199.394	-191.151	-190.333
Pseudo-R ²	0.138	0.143	0.179	0.182
Observations	347	347	347	347

Notes: *** $p \leq 0.001$; ** $p \leq 0.01$; * $p \leq 0.05$. In parentheses are heteroskedasticity-robust standard errors corrected for clustering at the subject level. Round and session dummies are used to control for round and session fixed effects.

In addition, Figure 2.5 shows that differences in information conditions have no systematic association with the number of wars across 10 rounds.

Figure 2.5: Incidence of War (By Enforcement Conditions)



This result suggests the relative non-importance of private information in the shadow of shifting power. But the result does not necessarily disprove the model prediction. The predicted outcomes in the private-information and commitment-problem models differ in kind: Prediction 3 is probabilistic while Predictions 1, 2 and 4 are deterministic. While the models clearly predict war with certainty in the case of the commitment problem, they predict only a *positive risk of war* in the private-information condition with enforcement. This is not unique to my model; comparable private-information models of war, including that in Fearon (1995), generate similar probabilistic predictions. Based on differences in the nature of their model predictions, one would expect the effect of private information to be less decisive than the effect of the commitment problem. Hence, all we can conclude is that the behavioral data provide no clear support for Prediction 3, based on private information on non-catastrophic costs of war in the shadow of shifting

power.³³ This result applies only to this particular form of private information. Future work may investigate if a similar result holds for other types of private information.

Result 5: Over time, the enforcement world remains relatively peaceful despite the shadow of shifting power.

Figure 2.6: Incidence of War in Rounds 1-10

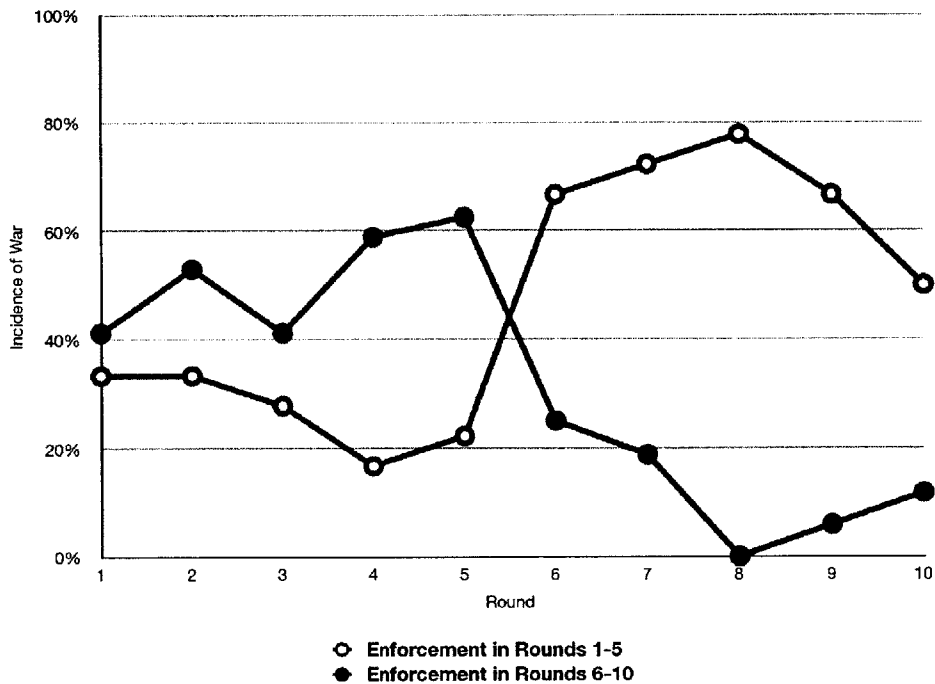


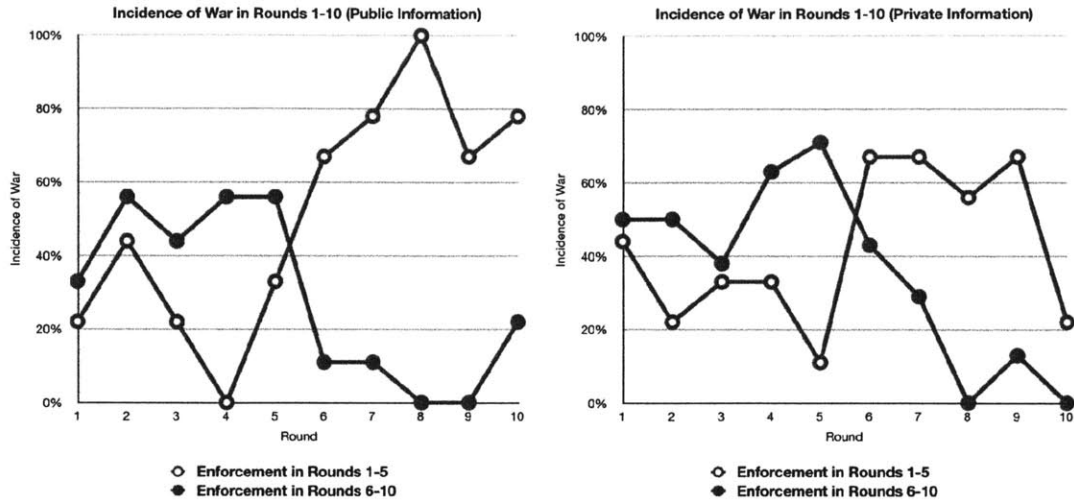
Figure 2.6 shows the evolution in the incidence of war over 10 rounds. The average incidence of war in the final rounds (Rounds 8-10) converged to 6% with an extremely low range of 0% to 12% in the enforcement world. By contrast, in the no-enforcement

³³ Recall that in the private-information condition, Player A does not know c_B and Player B does not know c_A , but they know that the c_i of their opponent exists within a non-catastrophic range represented by the discrete uniform distribution $[0, 4]$.

world, the average incidence of war in the final rounds is 65% with a high range of 50% to 78%. Figure 2.7 shows the trends over time in the different information conditions. The treatment effect is generally sharper in later rounds than in earlier rounds. This observation is consistent with the experimental literature. In game-theoretic experiments with multiple rounds, outcomes in the initial rounds are often more distant from the equilibrium predictions due to the subjects' unfamiliarity with the newly-introduced strategic environment.³⁴ An interesting question is why bargains fail in the enforcement condition in the beginning rounds. It turns out that 67%, 100% and 80% of the bargaining failures in Rounds 1, 2 and 3 respectively were caused by reservation-level offers, without which the treatment effect might have been sharper. In theory, Player A will make an offer to Player B based on the latter's reservation level in Stage 1, and Player B will accept if the offer is equal or greater than its reservation level. Of course, since the reservation level is the same as the war payoff, Player B is actually indifferent between the two options. Hence, reservation-level offers to Player B (3 points) will always run a risk of war. Since subjects are randomly assigned as either Player A or Player B in each round, the practical risks of reservation-level offers will only be apparent to all players after the first few rounds.

³⁴ Practice rounds with the subjects are sometimes used to clean out the unfamiliarity effect. This experiment did not use practice rounds.

Figure 2.7: Incidence of War (By Information Conditions)



Overall, the data show a substantial difference in the focal behavior that had emerged from the experiment: In the enforcement world, the dominant behavior revolved around the peaceful bargain; in the no-enforcement world, the dominant behavior was war. This result corresponds with Predictions 1, 2 and 4, and it suggests that the commitment problem has a causal effect on the incidence of conflict.

The next result shows that the causal effect is generalizable and robust beyond pure sequential bargaining.

Result 6: The treatment effect of the commitment problem is robust with interactions that allow for simultaneous moves.

After Round 10, I modified the game to allow for simultaneous moves. The purpose is to provide a stress-test of the main results by moving beyond pure sequential bargaining. The modified game is motivated by the fact that, in reality, states can always choose to forego the bargaining process if they believe that diplomacy is futile. The game is similar to the public-information variant of the earlier game, but with the option for simultaneous moves: A decides its demand while B decides at the same time whether to wait for the demand (and subsequently accept or reject it) or to wage a war. The model predictions remain unchanged.³⁵ Subjects were randomly assigned to either the enforcement or the no-enforcement condition. They played five rounds under public information (i.e. everyone knows that $c_i = 2$).³⁶

The incidence of war in the no-enforcement condition is 73%, which dramatically exceeds the incidence of 8% in the enforcement condition. The difference in the average incidence of war across players is statistically significant at $p < 0.0001$ (two-tailed t-test, $n = 70$). Regression models analyzing the round outcomes confirm the result. I fitted both logit and linear probability models with the war decision as the dependent variable and

³⁵ In this game, Player B has three options in Stage 1: (1) Agree to go to Stage 2 after seeing Player A's offer, (2) go to war after seeing A's offer, or (3) go to war without seeing A's offer. With enforcement, B is always better off seeing A's offer, which eliminates the rationality of the third option. With no enforcement, A's offer is not credible, and hence the option to see A's offer becomes immaterial to B's decision. Given the elimination of options, the modified model can be solved in the same way as the original model to yield predictions convergent to Predictions 1 and 2.

³⁶ Subjects in Session 1 played three rounds instead of five (see footnote 25). The treatment effect based on the incidence of war across players remains significant at $p < 0.0001$ (two-tailed t-test, $n = 70$) with data based on the three rounds in common (Rounds 11-13).

the enforcement variable as the regressor. I use different model specifications with and without round and session fixed effects, and with and without a control for the size of the initial offer and a control for player risk preference. Across all model specifications, enforcement has a strong negative relationship with war outcomes that is significant at $p < 0.001$. Both the initial-offer and risk-preference variables are insignificant ($p \geq 0.50$). Table B1 (in Appendix B1) displays the logit estimates conditioned on different sets of controls.

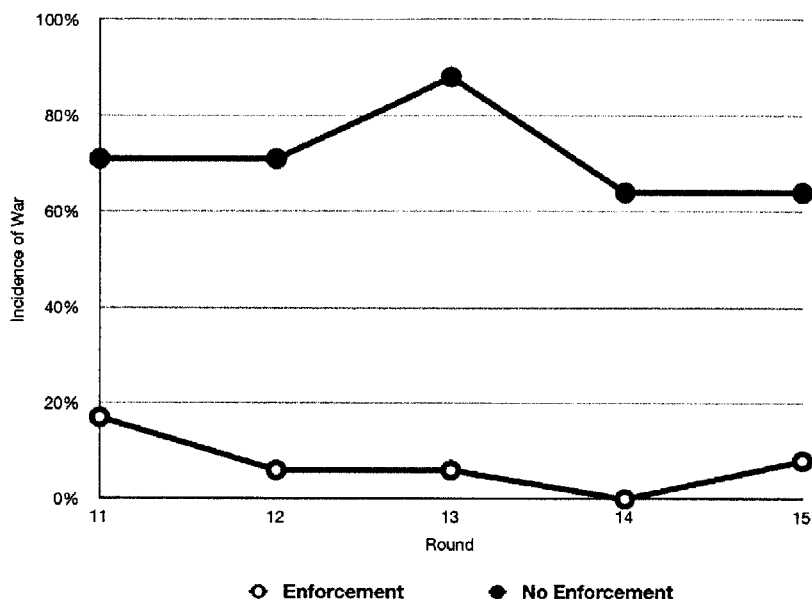
Table 2.5 below summarizes the incidence of war in each session. Figure 2.8 shows dramatic and persistent differences in the incidence of war across the two conditions.

Table 2.5: Incidence of War (Rounds 11-15)

Session	<i>Public-Information Condition</i>	
	Enforcement	No-Enforcement
1	0.00	0.67
2	0.17	0.88
3	0.03	0.63
Overall	0.08	0.73

Note: The maximum value for the incidence of war is 1.00 (100%).

Figure 2.8: Incidence of War (Rounds 11-15)



Taking this result with our earlier observations, we can conclude that the commitment problem has a strong positive effect on the decision for conflict. The treatment effect is large even in an experiment without large stakes. In real-world crises, the stakes are dramatically increased. All else being equal, it is plausible that decision-makers will be more payoff-sensitive and will try to behave more rationally – rather than less rationally – in a high-stake environment than in a small-stake environment.³⁷ In this case, the gravitation to the rational equilibrium – and war – may be even more pronounced.

³⁷ Being “releged upon” in a commitment-problem situation leads to the loss of some money in the context of the experiment, whereas the equivalent loss may be measured in life and death in the context of international politics. If we believe that people are more sensitive to the commitment problem (the treatment variable) in the latter case, then the treatment effect isolated in the laboratory may in fact be an *underestimation* of the true treatment effect in the real world.

5. Case Study Design

I supplement my analysis with a historical case study focused specifically on the commitment problem. I focus on the commitment problem for substantive and methodological reasons.³⁸ The case study evaluates if the logic of the commitment problem operates in a real-world international crisis. The purpose is to probe the historical realism of the experimental findings. I focus on three questions:

- 1) Does the strategic logic in the commitment-problem model operate in a real-world international crisis triggered by an impending power shift?
- 2) If so, how does the logic affect the decision for war?
- 3) Does the case suggest theoretical refinements to the commitment-problem model?

I choose my case to facilitate a fine-grained examination of the decision process, with a focus on these three questions.³⁹ I select the case of the Pacific War for three reasons:⁴⁰

³⁸ I focus on the commitment problem to test whether the main positive result in the experiment also holds in historical crises. Furthermore, it is difficult to isolate the private-information effect: all historical crises have occurred in environments with private information, and it is extremely hard to compare and measure variations across different dimensions of the information environment based on historical data that is necessarily incomplete.

³⁹ This purpose is different from using case studies to generalize whether an explanatory variable has systematic correlation with (or causation on) the outcome variable. Correlation (or causation) was experimentally evaluated in the previous section.

⁴⁰ In principle, any case that satisfies the same three criteria in equal or greater degree will make a good case, and can be used to check the robustness of Observations 1-3 derived from the 1941 case.

- i) The case includes a shock that created a sharp and unambiguous perception of an impending power shift. This allows me to separate out two scenarios across which the magnitude of the explanatory variable was sharply and suddenly changed. The shock as the U.S. imposed an economic blockade on Japan that included a total oil embargo implemented in August 1941. Since Japan depended on foreign oil imports to sustain its war machine, the embargo left Japan with little doubt about its impending decline in relative power.⁴¹ Thus, it would be informative to compare Japan's strategic calculations before and after the oil embargo.

- ii) The case has exceptionally good archival evidence that is relevant, reliable and relatively complete. This is an important reason for selecting this case. In many historical cases, meeting records of top-level decision-making are incomplete: they are either missing or censored for some (or most) of such meetings, which might potentially distort our understanding of the decision-making process. By contrast, the archival evidence in the 1941 case includes the relatively complete records of all liaison and imperial conferences that occurred between July 1941 and the outbreak of the Pacific War.⁴²

⁴¹ The increase in U.S. military preparations during 1941 had also alerted Japan to a potential power shift in the future. However, the U.S. military preparations were continuous in nature rather than a single discrete shock. Thus they did not create in Japan a perception of power shift that was as sudden and as unambiguous as what the oil embargo did in July-August 1941.

⁴² A group of scholars led by Dr. Jun Tsunoda of the National Diet Library found these records in the Military History Archives of the Japanese Defense Agency. As the translator Nobutaka Ike noted: "It is often difficult for an investigator to get reliable information that throws light on the decision-making process. Such information, if recorded at the time of the decision, usually belongs in the top-secret category and is not made public, even many years after the event; in this sense, the documents here translated are probably unique." Nobutaka Ike, *Japan's Decision for War: Records of the 1941 Policy Conferences* (Stanford, CA: Stanford University Press, 1967), xiii.

iii) The case involves a major war of historical importance. If there is an important factor in the case that is related to the commitment problem but omitted in our model, then our model would have missed an insight of substantive importance.

This is an archival analysis of the decision-making process in a particular timeframe, with the inference bounded by explicit evidential boundaries and inferential assumptions. To open the analysis to critical scrutiny, the (1) dependent variable, (2) explanatory variable, (3) study timeframe, (4) evidential boundaries and (5) inferential assumptions will be clearly stated.

(1) My *dependent variable* is not war, but the intermediate decisions or arguments made by top policymakers for or against war. Thus, the case contains multiple observations within my study timeframe, both before and after the oil shock that created a clear perception of an impending power shift. This shock connects to the (2) *explanatory variable*, which is the perception by policymakers of a severe future vulnerability that the opponent cannot commit not to exploit. My focus in the case analysis is on how the explanatory variable is connected (or not) to the dependent variable in the policymakers' deliberations for or against war, in the (3) *study timeframe* between July 1941 just before the oil shock to December 1941 when the declaration of war was delivered. The case study focuses *solely* on the calculations of the Japanese leadership in response to the impending power shift created by the U.S. embargo. My purpose is *not* to explain the Pacific War in entirety or to compare the utility of different theories, but to answer the three specific questions stated above. As my inquiry is of a limited nature, I do not

provide a full narrative of the historical processes leading to the outbreak of the Pacific War.⁴³

I define (4) my *data boundaries* to contain only the records of official meetings between the top decision-makers – the “imperial conferences” that approved key national decisions and the “liaison conferences” that ironed out differences among the cabinet ministers and military chiefs. The data include the four imperial conferences that preceded the Pacific War and the records of the 38th to 74th liaison conferences that occurred between the imperial conferences from July to December 1941. The detailed notes of these meetings have been translated⁴⁴ and are accessible to readers who wish to check accuracy. As they do not match the same standards of reliability and influentiality to the final decision, non-official remarks by top leaders and comments by lower-level officials are excluded, although they seem to be largely consistent with the findings here.

Finally, my conclusions are based on (5) three *inferential assumptions*:

- i) I assume relative constancy in Japan's perception of an impending power shift in the period after the U.S. embargo was imposed and before the Pearl Harbor attack. This is plausible, since the embargo remained in force during this period and the U.S.-Japan negotiations did not achieve notable progress towards the removal of the embargo.

⁴³ For a full historical narrative, see Akira Iriye, *The Origins of the Second World War in Asia and the Pacific* (London: Longman, 1987).

⁴⁴ Ike, *Japan's Decision*.

ii) I assume that all strategic calculations that were influential in Japan's decision for the Pacific War are detectable from the imperial and liaison conferences. Since the imperial conferences made the final decisions for war and the liaison conferences sorted out the differences among top leaders, it would be unusual to have an influential justification for war that did not leave a trace in these meetings.⁴⁵

iii) I also assume that the records are faithful representations of strategic deliberations within the top Japanese leadership. There seems to be no significant incentive for systematic self-misrepresentation in these meetings, as top leaders believed that their decisions were crucial for national survival and that the proceedings would be kept top-secret. A more likely source of bias is the selection effect of the record-taker. It was an army representative who recorded the proceedings for future reference by the army's high command.⁴⁶ Hence, a military bias is likely. However, while the record-taker might be motivated by his organizational emphasis, it is unlikely that he would doctor the facts or omit important parts of the meetings, since that would distort the high command's own internal reference.

At this point, I have made explicit my inferential assumptions and examined their relative plausibility. Insofar as these assumptions are approximated, the inference stands.

⁴⁵ An exception might be calculations based on domestic organizational interests. But even if this exception has a true effect, it does not contradict my purpose, which is to compare the strategic logic in the model with the arguments for war made by the top leaders. See also footnote 39.

⁴⁶ Ike, *Japan's Decision*, xiv.

Findings

Observation 1: Japan's decision for war involved strategic calculations similar to those behind the commitment-problem model.

The strategic calculation in the commitment-problem model (Section 2) can be isolated as follows: *It is better to fight earlier than later, because in the later period the payoffs for war shift against one's favor, and the opponent cannot commit not to exploit its advantage in the future.* The power shift implied in this logic was significantly triggered by the U.S. oil embargo, which took effect on 1 August 1941.

Evidence from the September imperial conference supports Observation 1. The strategic logic was repeatedly articulated to justify a specific deadline for war. In an indication of its importance, the logic fronted the opening statement of the Prime Minister, who commenced the proceedings, as well as the statements made by the second and third speakers at the conference.⁴⁷ For instance, the second speaker, Navy Chief of Staff Nagano Osami, warned: “A number of vital military supplies, including oil, are dwindling day by day. This will cause a gradual weakening of our national defense, and lead to a situation in which, if we maintain the status quo, the capacity of our Empire to act will be reduced in the days to come.... [I]t would be very dangerous for our Empire to remain idle and let the days go by.”⁴⁸

⁴⁷ Records of Imperial Conference, 6 September 1941, in Ike, *Japan's Decision*, 138-41.

⁴⁸ Records of Imperial Conference, 6 September 1941, in Ike, *Japan's Decision*, 139.

There was consensus that Japan's relative military advantage was declining over time. Thus Japan must prepare to fight earlier than later. Otherwise, its political and military situation would only become worse later. At that point, Japan could not expect its enemies to treat it better, now that they had become stronger. Arguments along the same logic appeared repeatedly at the conference. Stemming the relative decline was crucial, since it was clear that the U.S. could not credibly commit not to exploit a declined Japan in the future: "Even if we should make concessions to the United States by giving up part of our national policy for the sake of a temporary peace, the United States, its military position strengthened, is sure to demand more and more concessions on our part; and ultimately our Empire will have to lie prostrate at the feet of the United States."⁴⁹

Distrust of the Western powers appeared frequently in the liaison and imperial conferences. When strategic situations change, even formal agreements could be reneged. For example, after the German-Soviet war started in June 1941, Japan herself had considered reneging on the Soviet-Japanese Neutrality Pact. As President of the Privy Council Hara Yoshimichi told the imperial conference in July: "Some people say that it would be improper for Japan to attack the Soviet Union in view of the Neutrality Pact; but the Soviet Union is notorious for her habitual acts of betrayal. If we were to attack the Soviet Union, no one would regard it as treachery."⁵⁰ Betrayal was not a remote possibility in the spirit of the times. Foreign Minister Toyoda Teijiro highlighted to his colleagues the warning from the German Ambassador: "You will be tricked by the

⁴⁹ Reference Materials for Answering Questions at the Imperial Conference on 6 September Regarding "The Essentials for Carrying Out the Empire's Policies," in Ike, *Japan's Decision*, 152.

⁵⁰ Records of Imperial Conference, 2 July 1941, in Ike, *Japan's Decision*, 87.

United States, and negotiations will be drawn out; you had better break them off and avoid being tricked.”⁵¹ The reference materials for the September imperial conference emphasized: “In short, military force should be used promptly if there is no prospect of diplomatic success. It is expected that the United States and Great Britain will try to delay us with diplomatic negotiations. We must be careful not to be inveigled into this trap.”⁵²

What led to the urgent deterioration in relative power? Director of the Planning Board Suzuki Teichi explained to the conference the current state of Japan's national power. Because Japan depended on foreign trade for its critical resources, “as a result of the present overall economic blockade imposed by Great Britain and the United States, our Empire's national power is declining day by day. Our liquid fuel stockpile, which is the most important, will reach bottom by June or July of next year, even if we impose strict wartime control on the civilian demand. Accordingly, I believe it is vitally important for the survival of our Empire that we make up our minds to establish and stabilize a firm economic base.”⁵³ To the question “Why have we set the last ten days of October as a tentative deadline for war preparations?” the answer was:

We need not repeat that at present oil is the weak point of our Empire's national strength and fighting power As time passes, our capacity to carry on war will decline, and our Empire will become powerless militarily. Meanwhile the naval and air forces of the United States will improve remarkably as time goes on; and defensively, the United States, Great Britain, and the Netherlands will gradually grow stronger in the South. Hence the passing of time not only means that we will face more difficulties in military operations, but also means that the increasing

⁵¹ Records of the 49th Liaison Conference, 30 August 1941, in Ike, *Japan's Decision*, 126.

⁵² Reference Materials for Answering Questions at the Imperial Conference on 6 September Regarding “The Essentials for Carrying Out the Empire's Policies,” in Ike, *Japan's Decision*, 155.

⁵³ Records of Imperial Conference, 6 September 1941, in Ike, *Japan's Decision*, 148.

military preparedness of the United States Navy will surpass the naval power of our Empire after next autumn, and that we will finally be forced to surrender to the United States and Great Britain without a fight.⁵⁴

There was general agreement that the war would be a prolonged one. Precisely so, speed was the essence of the game and decisiveness the key to surviving the war. Navy Chief of Staff Nagano explained that the burdens of a prolonged war were best addressed if Japan could “seize the enemy's important military areas and sources of materials quickly at the beginning of the war, making our operational position tenable and at the same time obtaining vital materials from the areas now under hostile influence.”⁵⁵ Victory depended on the success of the first stage of military operations, which in turn depended on three factors: “first, to decide quickly to commence hostilities in view of the realities of our fighting capacity and theirs; second, to take the initiative rather than to allow them to do so; third, to consider the meteorological conditions in the operational areas in order to make operations easier.”⁵⁶ Delaying the decision for war was equivalent to denying the Japanese military the critical means for winning the war.

Did the imperial conference consider other strategic arguments that contradict Observation 1? It seems that no competing strategic logic was emphasized in the September imperial conference. However, the Army did highlight a second justification for a specific deadline for war along the same logic based on the “Northern Question” (Soviet-Japanese relations). Army Chief of Staff Sugiyama Gen expected a future

⁵⁴ Reference Materials for Answering Questions at the Imperial Conference on 6 September Regarding “The Essentials for Carrying Out the Empire's Policies,” in Ike, *Japan's Decision*, 154.

⁵⁵ Records of Imperial Conference, 6 September 1941, in Ike, *Japan's Decision*, 139-40.

⁵⁶ *Ibid*, 140.

coalition between the Soviet Union and the United States against Japan. But a military offensive in the winter season would be relatively ineffective. Hence, if Japan could “take advantage of the winter season and quickly finish our military operations in the South, [it] would be in a position to deal with any changes in the Northern situation that might take place next spring and thereafter. On the contrary, if we should miss this seasonal opportunity, we will not be able to achieve security in the North during our operations in the South.”⁵⁷ The same justification appeared in the November imperial conference, with Sugiyama warning of the possibility of war in the north with the Soviets. Given this possibility, Japan “must conclude its operations in the South as quickly as possible, and be prepared to cope with this situation.”⁵⁸

The September imperial conference marked the point at which Japan fixed a deadline for the war. But after September, there were two more imperial conferences before the outbreak of war in December. Was the final decision for war influenced by a different set of calculations? To assess this possibility, I examined the records of the two imperial conferences after September and before the Pacific War. I found that the same strategic logic persisted. In the November conference, the arguments for war were connected directly to the same logic. President of the Planning Board Suzuki analyzed Japan's national power and estimated the differences in outcomes between military action and inaction: “In conclusion, it would appear that if we go forward maintaining the present state of affairs, it would be very disadvantageous from the point of view of strengthening

⁵⁷ Ibid, 142.

⁵⁸ Records of Imperial Conference, 5 November 1941, in Ike, *Japan's Decision*, 227.

the material aspects of our national defense, if nothing else.”⁵⁹ Meanwhile, Army Chief of Staff Sugiyama emphasized the operational disadvantages of delaying the decision for war: “[T]he ratio of armament between Japan and the United States will become more and more unfavorable to us as time passes; and particularly, the gap in air armament will enlarge rapidly. Moreover, [American war preparations] will make rapid progress. Also, [the] joint defensive capability of [United States, Great Britain, the Netherlands and China] will be rapidly increased ... Thus it would be very disadvantageous for us to delay; and it is to be feared that it might become impossible for us to undertake offensive operations.”⁶⁰ At the close of the conference, Prime Minister Tojo Hideki said: “Two years from now we will have no petroleum for military use. Ships will stop moving ... I fear that we would become a third-class nation after two or three years if we just sat tight.”⁶¹

In the December imperial conference, however, the proceedings no longer revolved around the arguments for war, since the deadline for war had already been formalized. The die was already cast.⁶² The conference focused on the preparations for a prolonged war, with Navy Chief of Staff Nagano highlighting the military preparations, Prime Minister Tojo highlighting domestic security issues, Minister of Finance Kaya estimating Japan's financial situation during the war, and Minister of Agriculture Ino estimating the

⁵⁹ Records of Imperial Conference, 5 November 1941, in Ike, *Japan's Decision*, 222.

⁶⁰ *Ibid.*, 225-6.

⁶¹ *Ibid.*, 238.

⁶² President of the Privy Council Hara Yoshimichi opened the questions-and-answers session noting that “[w]e are discussing a very grave subject; but it was previously taken up by [the last] Imperial Conference, and every step that could be taken has been taken. Therefore, I have nothing in particular to add.” Records of Imperial Conference, 1 December 1941, in Ike, *Japan's Decision*, 279.

food supplies.⁶³ The conference ended with an imperial sanction for the declaration of war against the United States.

The logic of the commitment problem was activated by an impending power shift triggered by the U.S. oil embargo, without which an important basis for the strategic logic would be removed. As such, we should expect a substantial difference in the strategic deliberations at the imperial conferences *before* and *after* the oil embargo was imposed. I found that this was indeed the case.

At the July imperial conference, the focus was on plans to establish the Greater East Asia Co-prosperity Sphere and settle the war in China expeditiously, as well as the question of war with the Soviet Union. However, “[i]n carrying out the plans outlined above ... our Empire will not be deterred by the possibility of being involved in a war with Great Britain and the United States”; thus Japan should also prepare for that possibility.⁶⁴ Army Chief of Staff Sugiyama argued that “in carrying out various measures for the solution of the Northern Problem [against the Soviets],” Japan should keep to its “basic position of always being prepared for war with Great Britain and the United States, since the attitude of these countries toward Japan cannot be viewed with optimism.”⁶⁵ Navy Chief of Staff Nagano noted that “Great Britain, the United States, and the Netherlands are currently stepping up their pressure against Japan” and that “if our Empire finds itself unable to cope with this, we may, it must be anticipated, finally have to go to war with Great

⁶³ Records of Imperial Conference, 1 December 1941, in Ike, *Japan's Decision*, 271-8.

⁶⁴ “Outline of National Policies in View of the Changing Situation”, 2 July 1941, in Ike, *Japan's Decision*, 78.

⁶⁵ Records of Imperial Conference, 2 July 1941, in Ike, *Japan's Decision*, 81.

Britain and the United States. So we must get ready, resolved that we will not be deterred by that possibility.⁶⁶

At this point, while Japanese leaders acknowledged the possibility of a war with Britain and the U.S., their strategic focus and belligerence did not center on the U.S. The conference heard no specific plan to attack the U.S., and there was no clear articulation of the strategic logic that was to be emphasized in later imperial conferences.⁶⁷ In September, the situation changed. The oil embargo created an impending power decline, and the logic underlying the commitment-problem model took hold in the minds of the decision-makers. Without an impending power decline, that logic was muted and the July imperial conference did not hear any specific proposal for attacking the U.S. With an impending power decline, that logic was activated and the September imperial conference heard repeated justifications for war consistent with the logic. Hence, the commitment-problem model predicts both the content of the speech evidence as well as the *change* in the speech evidence.

⁶⁶ Ibid, 81-2. The conference was aware that the realization of its objectives might provoke Britain and the U.S. into war. The war, however, was seen as a negative scenario which Japan should try its best to avoid. Foreign Minister Matsuoka Yosuke emphasized the need to “maintain a very cautious diplomatic attitude in order to prevent America from entering the European war, and to prevent her from clashing with our country.” According to Matsuoka, “a war against Great Britain and the United States is unlikely to occur if we proceed with great caution.” President of the Privy Council Hara Yoshimichi concluded that “the Government and the Supreme Command are in agreement on this point: that is, we will try our best to avoid a clash with Great Britain and the United States. I believe that Japan should avoid taking belligerent action against the United States, at least on this occasion.” Ibid, 83, 87, 88-9.

⁶⁷ Akira Iriye pointed out that “[a]lthough war with the combined ABCD powers had been envisaged for some time, as of early August there had been no comprehensive master plan.” Iriye, *Origins of the Second World War*, 150.

Could it be that members of the imperial conference had a different strategic logic in mind that they could not articulate to the emperor? How was the agenda set for the September imperial conference? To address these questions, I examined the records of the liaison conferences between July and September. Of major importance was the September 3 conference between Prime Minister Konoye Fumimaro, Foreign Minister Toyoda Teijiro, War Minister Tojo Hideki, and the military chiefs. At this conference, which lasted for seven hours, Japanese leaders debated the policy proposals that were subsequently tabled at the imperial conference on September 6. The justifications for war in the liaison conference were consistent with the logic in the commitment-problem model. Navy Chief of Staff Nagano, whose navy had first drafted the policy proposals, began the meeting with his statement fronted by the same logic: "In various respects the Empire is losing materials: that is, we are getting weaker. By contrast, the enemy is getting stronger. With the passage of time, we will get increasingly weaker, and we won't be able to survive.... Although I am confident that at the present time we have a chance to win a war, I fear that this opportunity will disappear with the passage of time."⁶⁸

Decision-makers emphasized the element of speed in their discussions, as seen in the cited examples. In their reasoning, it was imperative to make a swift decision to attack the U.S. because the situation was changing quickly against Japan's favor. Hence, the September conference emphasized that preparations for war should be completed by late October, and the November conference fixed the deadline for war in early December. It

⁶⁸ Records of Liaison Conference, 3 September 1941, in Ike, *Japan's Decision*, 130-1.

appears that the strategic justifications for war were not driven merely by the impending power shift, but by the perceived *speed* of the power shift. This provides the next finding:

Observation 2: The perceived speed of shift in relative strength was important in Japan's calculations. This factor was omitted in the model.

How does the speed of shift in relative strength affect decisions for war? In an *expected-utility mechanism*, the loss in the expected utility for war accelerates as the speed of decline in relative strength increases. For the declining state, waiting costs increase as the speed of shift increases. As a consequence, rapidly declining states are more likely to choose war than slowly declining states. We can call this the “utility effect” of the power shift. In the case of the Pacific War, evidence of the utility effect can be seen from much of the evidence presented above for Observations 1 and 2. However, the speed of shift in relative strength may also create another effect that is not captured in the expected-utility mechanism. A rapid power shift promotes hastened or truncated diplomacy.⁶⁹ Negotiations may be more likely to fail under the tight time pressure of a truncated bargaining timeframe. This suggests a *behavioral mechanism* that connects from the utility effect in the first mechanism. Due to the utility effect, states truncate the bargaining process with a tight timeline, which increases the risk of bargaining failure. We can call this the “truncation effect” of the power shift.

⁶⁹ Van Evera, *Causes of War*, 83-5.

Observation 3: The speed of shift in relative strength made Japan truncate the bargaining timeframe.

The impending power shift led to the decision for a tight deadline for war, which translated into a tight timeline for diplomacy. The tight timeline severely restricted the possibility of a successful negotiation. This fact was pointed out repeatedly. For example, Foreign Minister Togo stated at the November imperial conference that “the situation is becoming more and more critical every day, and negotiations with the United States are very much restricted by the time element; consequently, to our regret, there is little room left for diplomatic maneuvering. Moreover, the conclusion of a Japanese-American understanding would necessitate great speed in negotiations, partly because of the time required for domestic procedures on the American side. For this reason we have been required to carry on negotiations under extremely difficult circumstances. The prospects of achieving an amicable settlement in the negotiations are, to our deepest regret, dim.”⁷⁰ Remarks that the tight timeline made diplomacy difficult were also made in various liaison conferences between the September and December imperial conferences.⁷¹

Does the truncation effect exist?

The case evidence shows the truncation of the bargaining timeframe due to the perceived speed of shift in relative power. But it does not show if there is a truncation *effect* on the decision for war. Since we cannot observe counterfactuals, we cannot know if a

⁷⁰ Records of Imperial Conference, 5 November 1941, in Ike, *Japan's Decision*, 214.

⁷¹ Records of Liaison Conferences, in Ike, *Japan's Decision*, 167-84, 199-207, 239-47.

hypothetical removal of the tight deadline for diplomacy would have led to a peaceful bargain between the U.S. and Japan. The problem of inference is compounded as the truncation effect theoretically follows the utility effect of power shift, and bargaining truncation often surfaces in conjunction with the utility effect. It is hard to parse out cleanly the truncation effect with historical data to show that the truncation of the bargaining timeframe is not merely incidental, but is in fact causally influential in the decision for war. In other words, we do not know if the behavioral mechanism provides a valid causal model for wars arising from power shifts. This is unfortunate – it is important to know whether the behavioral mechanism holds, because the truncation effect could be more manipulable than the utility effect for states seeking the prevention of war.

To circumvent the limits of the historical data, I replicate the truncation mechanism in a laboratory setting and evaluate its significance experimentally. I begin by incorporating Observations 2 and 3 into the experimental setup. The basic idea is to make the war payoffs time-dependent. I conduct the experiment in the public-information/no-enforcement condition with the interaction modified. In Stage 1 of the game, Player A decides on its demand $x_1 \in [0, 10]$, where x_1 is an integer. At the same time, Player B decides whether to wait for the demand (and subsequently accept or reject it) or to wage a war. If B accepts x_1 , the game enters Stage 2 in which A decides its demand $x_2 \in [0, 10]$. At the same time, Player B decides whether to wait for the demand (and subsequently accept or reject it) or to wage a war. If B accepts x_2 , the game ends with payoff $(x_2, 10 - x_2)$. In both periods, if B chooses to fight before or at the T -th second (where T is a positive integer), the game ends with payoff $(5 - c_A, 5 - c_B)$, but if B chooses to fight after

the T -th second, the game ends with payoff $(7 - c_A, 3 - c_B)$, where c_i is the cost of war for State i . Hence, the payoff shift is avoided if a bargain is made before the T -th second.⁷²

This is a bargaining experiment that probes whether the speed of change in relative bargaining power affects the risk of war. In this experiment, the payoffs for war are time-dependent. In the control condition, the war payoff $(5 - c_A, 5 - c_B)$ applies until the 60th second and the war payoff $(7 - c_A, 3 - c_B)$ applies after the 60th second. In the treatment condition, the duration with war payoff $(5 - c_A, 5 - c_B)$ is reduced to 30 seconds. The specific timings are calibrated based on simulated plays, with 30 seconds providing the time pressure to elicit the truncation effect, and 60 seconds providing the time buffer to hold off the truncation effect. Subjects were randomly assigned to treatment and control conditions. They were randomly assigned as either Player A or Player B and they played with a randomly assigned opponent. The observable implications are straightforward: If the truncation effect does not apply, we should not expect the war outcomes to differ between the treatment and control groups; but if it does, we should.

This was a one-round bargaining experiment (Round 16) that followed the previous experiment, with the same experimental setting and payment scheme. Hence, when the subject reached this experiment, he or she should be familiar with the structure of the bargaining game and should be sufficiently prepared to handle the additional complexity. Before playing, subjects read instructions that highlighted the differences with the

⁷² As such, the predicted outcome differs from that of the earlier public-information model with no enforcement. A's offers will be based on B's reservation level, which is always higher before or at the T -th second than after the T -th second. Hence, B will not wait beyond the T -th second because A will always make a worse offer to B after the T -th second. Knowing this, A will make and confirm its offer based on B's reservation level before the T -th second, and B will accept. War is thus avoided.

previous game (see Appendix A1). Meanwhile, the typical tradeoffs of using the same subject pool were reduced by experimental design: the stranger-matching protocol eliminated potential reciprocity effects; the use of a random payment mechanism and the revelation of earnings at the end of the session reduced potential endowment effects; and the restriction of each session to less than an hour reduced the likelihood of experimental fatigue. Nevertheless, it should be noted that this experiment is a single-trial preliminary study without the full infrastructure for a decisive experimental test. Its purpose is simply to serve as an initial detection probe for the truncation effect.

Observation 4: The truncation effect is likely to exist.

In the treatment group where a tight timeline was imposed to elicit the truncation effect, the percentage of war outcomes was 88% compared to 56% in the control group where the timeline was doubled to buffer the truncation effect. The difference in war outcomes across the two groups is significant ($n = 35$: Mann-Whitney test, $p = 0.0350$; two-tailed test of proportion, $p = 0.0324$) with dyadic binary outcomes as observations.

However, Observation 4 is not a decisive result. I set up a robustness test using Session 1, which ran the experiment for three rounds. The statistical significance behind Observation 4 is based on analysis that combined the single-round outcomes from Sessions 2 and 3 with the first-round outcomes from Session 1. First-round outcomes are most comparable to the single-round outcomes, since they were similarly generated by subjects playing the game for the first time. Hence, combining the single-round outcomes

with the first-round outcomes from Session 1 provides the most comparable combination of observations. Nonetheless, Observation 4 is not robust unless it passes an additional hurdle: that *every* possible combination of outcomes yields statistically significant p -values. This is not the case. When single-round outcomes in Sessions 2 and 3 are combined with second-round and third-round outcomes in Session 1, the differences in war outcomes across the two groups (76% in control group against 67% in treatment group with second-round outcomes in Session 1; and 82% in control group against 61% in treatment group with third-round outcomes in Session 1) are consistent with Observation 4 but statistically insignificant ($n = 35$: Mann-Whitney test, $p = 0.5271$ and $p = 0.1706$ respectively; two-tailed test of proportion, $p = 0.5211$ and $p = 0.1644$ respectively). On the whole, this preliminary probe gives us grounds to suspect that the truncation effect exists. But Observation 4 should be treated as indicative rather than decisive. More experiments are necessary to test the robustness of this particular result.

6. Conclusion

Private information and the commitment problem are central to the rationalist theory of war. But estimating their causal effects is empirically challenging. I use an experiment to test the causal effects. I focus on three questions: (1) Do private information and the commitment problem have causal effects on the incidence of conflict? (2) How large are their effects? (3) Over time, how much more peaceful is a world with credible enforcement or public information compared to a world without? I find six results:

- 1) The commitment problem causes a large increase in the number of wars.
- 2) A sudden introduction of a commitment problem causes a sharp rise in the number of wars. A sudden introduction of enforcement causes a sharp fall.
- 3) In the no-enforcement condition, players with a relative power shift in their favor renege on their agreements more than 90% of the time.
- 4) Private information has no significant effect on the incidence of war in the case of private information on non-catastrophic costs of war in the shadow of shifting power.
- 5) The enforcement world remains relatively peaceful over time despite the shadow of shifting power.
- 6) The treatment effect of the commitment problem is robust with interactions that allow for simultaneous moves.

In my experiment, the commitment problem had a decisive effect on the incidence of conflict in both private- and public-information environments, while private information had no significant effect on the risk of conflict regardless of the enforcement environment. A case study focused on the commitment problem assesses the historical realism of the positive experimental finding. The case study shows how Japan's policy

deliberations leading to the Pacific War were strongly influenced by calculations resembling the strategic logic I isolated in the model and experiment. The study also reveals how the perceived speed of shift in relative strength made Japanese leaders push for a tight deadline for war. The deadline truncated the bargaining timeframe and reduced the prospects of a diplomatic settlement.

Convergent results from the formal, experimental and historical analysis make us more confident to conclude: The commitment problem has a strong positive effect on the incidence of conflict. Despite the costs, decision-makers with a commitment problem gravitate towards direct conflict. Hence, impending power shifts can be extremely dangerous. In general, power shifts are perilous when there is no enforcement structure to suppress the incentive to renege on prior agreements. The existence of this incentive dissolves the willingness to trust and makes costly conflict more likely. I have focused specifically on interstate conflict, but the strategic form of the commitment problem isolated in this chapter is extremely general. As a result, the findings may have potential implications for other forms of conflict – ranging from civil conflict to contentious politics – that involve the commitment problem in a similar form.

Chapter 3

Decisions, Processes and War: Evidence from the Sino-Japanese Wars

The Sino-Japanese wars that started in 1931 and 1937 were two of the most consequential wars in modern history. What caused these wars? What was the sequence of decisions and interactions that led to each war? Are the historical details explained by the theories in Chapter 1? Are there gaps in those theories?

Chapter 3 addresses these questions. It has six sections. Section 1 describes the research setup and the value of using two-sided archival evidence to study the causes of war. Sections 2 and 3 focus on the 1937 case. Section 2 summarizes the predictions derived from the theories in Chapter 1, and highlights which predictions fit or contradict the case evidence. Section 3 traces the decision-making process based on archival records from both sides of the conflict. Two-sided archival evidence allows us to see the interactive decisions and perceptions that drove each crisis into war. Sections 4 and 5 focus on the 1931 case. Section 4 mirrors Section 2 (theoretical tests) while Section 5 mirrors Section 3 (historical analysis). Section 6 highlights policy lessons from the history of these wars.

1. Two-Sided Archival Evidence

This chapter tries to get as close as possible to the decision processes that drove the crises into wars. I explore whether the mechanisms tested experimentally in Chapters 2 and 4 had influenced decision-making in actual historical crises – and if so, how so?

My case selection is based on the criterion of *two-sidedness* – that archival records of decision-making on *both* sides of the conflict are accessible and reliable.¹ The criterion of two-sidedness allows for a more accurate and unbiased understanding of the origins of war. It allows us to map out the actions and reactions generated by both sides of the conflict. It reveals the interactive perceptions that characterized a crisis and motivated the decisions for war.

War is a dyadic outcome. Conceptually, we know that one-sided evidence cannot fully explain a dyadic outcome. Practically, however, we are limited by the fact that two-sided archival evidence is rarely available. We are forced to work with what we have.² In my case universe, five wars satisfy the criterion of two-sidedness: The Second Sino-Japanese War, the Third Sino-Japanese War, the Pacific War, and (to some extent) the Russo-Japanese War and the Korean War. I chose the Sino-Japanese wars for three reasons. First, they are the most under-studied cases among the five. Hence, a detailed analysis of

¹ For example, the case of the Vietnam war is not selected due to its archival one-sidedness: We have access to good primary evidence on top-level decision-making on the American side, but not on the Vietnamese side.

² By implication, archival two-sidedness is an objective criterion for case selection that is largely out of the researcher's control. In some sense, it is the cases that select themselves; this reduces the possibility of researcher bias in case selection.

the decision process in these cases makes a useful documentary contribution. Also, as the existing literature offers limited prior information on these cases, I had chosen the cases without clear prior expectations on whether they would fit or contradict the rationalist mechanisms.³ Finally, this is a pair of cases that involve the same dyad in two different time periods. This allows us to make intertemporal comparisons between the two cases.

Definition of Study Variables; Delineation of Data Boundaries

This chapter consists of two case studies. They are micro-level archival studies of the decision process over a specified timeframe, with inference bounded by clearly-defined data boundaries. To make the analysis transparent and to facilitate the replicability of the findings, the (a) outcome variable, (b) explanatory variables, (c) study timeframe and (d) evidence boundaries will be defined.

My *outcome variable* is not the war outcome, but the decisions and proposals made by key policymakers for and against escalation or war. Thus, each case contains multiple observations. There is substantial variation on the outcome variable *within* each case. The variation reflects the complexities that pave the road to war – all these information are lost if we use a dataset with a dependent variable coded “war = 1” and “no war = 0”. Indeed, historical contingencies and “accidents” figured prominently in 1931 and 1937.⁴

³ This reduces the researcher’s ex-ante bias in case selection, since the results are not clear until *after* the archival research is done.

⁴ Just before the Mukden crisis that led to the Second Sino-Japanese War in 1931, the General Staff in Tokyo sent Major-General Tatekawa Yoshitsugu with a direct order to stop the Ishiwara-Itagaki plot. They chose the wrong man for the mission. While he had arrived hours before the Mukden explosion, Tatekawa went drunk and fell asleep in a Japanese restaurant. During the July crisis in 1937, the Army

But the proposals and arguments made by leaders for or against escalation are far from random. They are based on specific calculations and beliefs. They reveal in detail what drives pro-war or anti-war behavior. They give a more grounded and fine-grained understanding of how war occurs. Hence, they are a more useful dependent variable.

My *explanatory variables* are based on the mechanisms specified in Chapter 1; they are summarized in the next section. The *study timeframe* is fixed at September 1931 (the Mukden crisis) for the Second Sino-Japanese War and July 1937 (the Marco Polo Bridge crisis) for the Third Sino-Japanese War.

This is a limited study focused on the outcome variable of interest (the decisions and proposals made by key policymakers for and against escalation), rather than a complete historical account of the events and individuals involved in the long road to war. The latter is beyond the scope of this chapter. Hence, while this study describes how crises evolve into wars, it does not explain how preexisting situations evolve into crises. Thus,

HQ at Tokyo sent Colonel Shibayama to do a firsthand evaluation of the situation in Tianjin. Tokyo had intended to wait for the evaluation before it decided whether military mobilization was necessary. According to Crowley, Shibayama arrived at a consensus with the field army HQ that “the settlement of July 11 should be implemented at all costs and that reinforcements were not to be sent to North China.” “However, whether because of inclement weather or because of a side trip to the headquarters of the Korean Army, Shibayama did not arrive in Tokyo until July 20, a delay of more than forty-eight hours [In the meantime Tokyo received reports] that Nanking had ordered the mobilization of all troops north of the Yangtze River” Jonathan Spence's textbook introduced the Third Sino-Japanese War by highlighting the fateful confluence of separate events: “Premier Hayashi's government failed to get its economic policies through the Japanese parliament, and was replaced by a government headed by the influential but indecisive Prince Konoe. Japan's commanding general in north China suffered a heart attack, and had to be replaced by a less experienced subordinate. And Chinese troops in the vicinity of the “Marco Polo Bridge” (Lugouqiao) decided to strengthen some shore-line defenses on the banks of the Yongding River.” Sadako Ogata, *The Making of Japanese Foreign Policy, 1931-1932* (University of California Press, 1964), 58-9; James Crowley, *Japan's Quest for Autonomy: National Security and Foreign Policy, 1930-1938* (Princeton, NJ: Princeton University Press, 1966), 333; Jonathan Spence, *The Search for Modern China* (New York: W. W. Norton, 1999), 419.

long-term theories of war such as those related to systemic-level factors fall outside my explanatory scope.

Analytically, attention is focused on key policymakers whose calculations and decisions directly impacted how the crisis evolved. Empirically, emphasis is placed on contemporary records of correspondences, meetings and discussions available in the archives. Inevitably, the analysis has omitted the thoughts and actions of officials outside the decision-making circles, as well as the private calculations and ambitions of individuals that could not be observed from the records. Factional splits were important as well, as we will see. However, private ambitions and motives may be under-reported in the available evidence, as they are less likely to be captured on record.⁵ With this caveat, I ground my analysis on contemporary records at the decision-making level.⁶ Such evidence is more reliable from a historiographic standpoint, and more tightly connected to my dependent variable at the decision-making level from an analytical standpoint. Clearly demarcated evidence boundaries also allow for replicability of results.

My evidence boundaries are demarcated by the subset of primary sources I use. On the Chinese side, I use the archival collections produced by the Central National Archive, Second National Archive, and two provincial archives of the People's Republic of China, as well as the archival collections produced by the Kuomintang Party History Committee in the Republic of China on Taiwan. These sources, and their abbreviations, are footnoted

⁵ Much more can be said about individual ambitions and rivaling factions if we loosen our boundaries of evidence to include postwar memoirs or the recollections and speculations of those outside the decision-making circles.

⁶ Evidence that is not firsthand or contemporary will be indicated.

here.⁷ On the Japanese side, I use Japanese archival records translated into Chinese⁸ by the Central National Archive, Second National Archive, the Tianjin Municipal Daihon'ei Rikugunbu Translation Team, and Fudan University in the People's Republic of China.⁹

These sources are not exhaustive. But they represent a relatively comprehensive and authoritative subset of accessible archival materials that reflect the deliberations of top

⁷ Central National Archive of China, Second National Archive of China, and the Social Science Institute of Jilin Province, ed., *Jiu Yi Ba Shibian* [The September 18 Incident] (Beijing: Zhonghua Shuju, 1988). Henceforth: "JYB". Liaoning Provincial Archive, ed., *Jiu Yi Ba Shibian Dangan Shiliao Jingbian* [September 18 Incident: Essential Archival Collections] (Shenyang: Liaoning Renmin Chubanshe, 1991), Vol. 5-6. Henceforth: "JYBD". Second National Archive of China, ed., *Kangri Zhanzhen Zhengmian Zhanchang* [Anti-Japanese War of Resistance Battlefield] (Jiangsu Guxi Chubanshe, 1987). Henceforth: "KRZZ". Second National Archive of China, ed., *Zhonghua Minguo Shi Dangan Ziliao Huibian* [Archival Collections on the History of Republican China] (Jiangsu Guxi Chubanshe, 1991), Vol. 5, Part 1. Henceforth: "ZMSD". Second National Archives of China, ed., "Luguoqian Shibian Hou Guomingdang Zhengfu Junshi Jiguan Zhangguan Huibao 1-15 Huiyi Jilu" [Minutes of the 1st to 15th Joint Meeting of the Military High Command After the Marco Polo Bridge Incident], *Republican Archives* (1987): 3-17. Henceforth: "JMRA". Li Yong and Zhang Zhongtian, ed., *Jiang Jieshi Nianpu* [Chronological Records of Chiang Kaishek] (Beijing: Central Party History Publication, 1995). Henceforth: "JJNP". Zhang Youkun, Qian Jin and Li Xuequn, ed., *Zhang Xueliang Nianpu* (Beijing: Shehui Kexue Wenxian Chubanshe, 2009). Henceforth: "ZXNP". Kuomintang Party History Committee, ed., *Xian Zongtong Jianggong Sixiang Yanlun Zongji* [Overall Collection of the Thoughts and Speeches of Late President Chiang] (Taipei: Kuomintang Party History Committee, 1985). Henceforth: "XZJS". Qin Xiaoyi, ed., *Zongtong Jianggong Dashi Changbian Chugao* [Preliminary Draft of President Chiang's Chronological Biography] (Taipei, 1978-2004), by the Kuomintang Party History Office. Henceforth: "ZJDC".

⁸ There is potential bias as Chinese translations of Japanese archives may deliberately make the Japanese appear more hawkish and aggressive than is the "truth". But in fact, as Section 3 shows, the translated materials provide a rather *un*-hawkish depiction of Japanese decision-makers in the 1937 July crisis.

⁹ Central National Archive of China, Second National Archive of China, and the Social Science Institute of Jilin Province, ed., *Huabei Shijian: Riben Diguo Zhuyi Qinhua Dangan Ziliao Xuebian* [North China Incident: Collected Archival Sources on the Japanese Imperialist Invasion of China] (Beijing: Zhonghua Shuju, 2000). Henceforth: "HBSJ". Faculty of History at Fudan University, ed., *Riben Diguo Zhuyi Duiwai Qinvue Shilian Xuanbian* [Collected Historical Materials on Japanese Imperialist Invasions] (Shanghai: Shanghai Renmin Chubanshe, 1983). Henceforth: "RDZDQ". Zhang Bofeng and Zhuang Jianping, ed., *Kangri Zhanzhen* [Anti-Japanese War of Resistance] (Chengdu: Sichuan University Press, 1997), by the Chinese Association for Historical Studies and the Contemporary History Research Unit of the Chinese Institute of Social Science. Henceforth: "KZ". Tianjin Municipal Translation Team for the *Daihon'ei Rikugunbu*, ed., *Riben Diguo Zhuyi Qinhua Ziliao Changbian: Da Ben Ying Lujunbu Zhaiyi* [Extensive Sources on the Japanese Imperialist Invasion of China: Selected Translations of *Daihon'ei Rikugunbu*] (Sichuan Renmin Chubanshe, 1987). Original Source: Japan Defense Agency War History Office, *Daihon'ei Rikugunbu* [Imperial Headquarters, Army], 10 vols. Henceforth: "DHR".

decision-makers in the respective crises. My findings are conditional on the veracity of these sources. As such, I have tried to define the boundaries of my sources as explicitly as possible. This opens the evidential foundation of my interpretations to direct scrutiny. It also facilitates external assessment of the extent to which the cited evidence in my analysis is representative of the archival evidence as a whole. By doing so, I hope that the gaps and biases which I failed to avoid in my interpretations are at least traceable and rectifiable by fellow scholars. Further, as more evidence becomes available in the future, the robustness of the archival findings can be tested.

THIRD SINO-JAPANESE WAR (1937)

World War II started in China.¹⁰ The Third Sino-Japanese War started in 1937 and merged into the Pacific War in 1941. The war lasted eight years. It killed four million soldiers.¹¹ It killed 20 to 30 million people.¹² It changed the history of East Asia and the world.

But the English-language scholarship has not paid much attention to the Marco Polo Bridge crisis that started the war in July 1937. The last significant archival study of the crisis that appeared in English was Hata Ikuhiko's essay translated by David Lu and

¹⁰ Scholars of Asian history have made this point. Jonathan Spence, for instance, called the Japanese attack on Wanping city "the first battle of World War II". Spence, *Search for Modern China*, 421.

¹¹ Chinese and Japanese battle deaths in the merged Sino-Japanese War and Pacific War, based on Meredith Reid Sarkees, "The Correlates of War Data on War: An Update to 1997," *Conflict Management and Peace Science* 18 (2000): 123-44.

¹² Diana Lary, "Introduction: The Context of the War," in *China at War: Regions of China, 1937-1945*, ed. Stephen Mackinnon, Diana Lary, and Ezra Vogel (Stanford, CA: Stanford University Press, 2007), 1.

edited by James Morley.¹³ It is a major contribution. However, it drew all its conclusions from Japanese archival sources. The same limitation characterized its predecessors.¹⁴ Hence, scholars could only rely on the Japanese part of the story, which allows for a very thin understanding of what actually happened on the Chinese side. On the other hand, Chinese scholarship, while enormous, is often characterized by a strong patriotic antipathy against the Japanese, whose official historical treatment of the war (and the Nanjing Massacre) China has yet to forgive.¹⁵ In Chinese historiography, the dominant consensus is that the 1937 war was the product of a deliberate Japanese imperialist

¹³ Hata Ikuhiko, "The Marco Polo Bridge Incident, 1937," in *The China Quagmire: Japan's Expansion on the Asian Continent, 1933-1941*, ed. James Morley (New York: Columbia University Press, 1983), 243-86.

¹⁴ James Crowley, *Japan's Quest for Autonomy: National Security and Foreign Policy, 1930-1938* (Princeton, NJ: Princeton University Press, 1966); John Hunter Boyle, *China and Japan at War, 1937-1945: The Politics of Collaboration* (Stanford, CA: Stanford University Press, 1972); Mark Peattie, *Ishiwara Kanji and Japan's Confrontation with the West* (Princeton, NJ: Princeton University Press, 1975). See also the descriptions of the Marco Polo Bridge crisis in Michael Barnhart, *Japan Prepares for Total War: The Search for Economic Security, 1919-1941* (Ithaca, NY: Cornell University Press, 1987), 84-90; Barbara Brooks, *Japan's Imperial Diplomacy: Consuls, Treaty Ports, and War in China, 1895-1938* (Honolulu: University of Hawaii Press, 2000), 180-5; and Hatano Sumio and Shoji Junichiro, "The Sino-Japanese War of 1937-45: Japanese Military Invasion and Chinese Resistance," in *Japan-China Joint History Research Report: Modern and Contemporary History, Vol. 1*, ed. China Joint History Research Committee (Governments of Japan and the People's Republic of China, 2011), 128-33. None of these works consulted Chinese sources. There are relevant studies that worked with Chinese archival sources, including: John Garver, *Chinese-Soviet Relations, 1937-1945: The Diplomacy of Chinese Nationalism* (New York: Oxford University Press, 1988); Parks Coble, *Facing Japan: Chinese Politics and Japanese Imperialism, 1931-1937* (Cambridge, MA: Harvard University Press, 1991); Sun Youli, *China and the Origins of the Pacific War, 1931-1941* (New York: St. Martin's Press, 1993); Hans Van de Ven, *War and Nationalism in China 1925-1945* (London: Routledge, 2003); and Jay Taylor, *The Generalissimo: Chiang Kai-shek and the Struggle for Modern China* (Cambridge, MA: Harvard University Press, 2009). However, Garver's work dealt with Sino-Soviet wartime relations. Coble's work did not cover the 1937 July crisis. Sun and Taylor touched on it only marginally. Van De Ven's work, while focused on the theme of nationalism and war, is more comprehensive. None of these works worked with Japanese archival sources.

¹⁵ Diplomatic quarrels over the history of the 1937 war persist into the present day. See, e.g., "Chinese City Severs Ties After Japanese Mayor Denies Massacre," *New York Times*, 22 February 2012; "Tokyo Governor Backs Nanjing Massacre Denial," *The Straits Times*, 24 February 2012. See also Parks Coble, "China's 'New Remembering' of the Anti-Japanese War of Resistance, 1937-1945," *China Quarterly* 190 (2007): 394-410.

strategy aimed towards the subjugation of China.¹⁶

Unsurprisingly, our existing knowledge on the 1937 case is partial and occasionally incorrect. On the one side, the landmark contribution by Hata concluded: “On China’s part, its political unification and economic reconstruction were near completion, and, for the time being, it was necessary to avoid any final break in its relations with Japan. The Nationalist government approached the issue cautiously, fearing that extension of the conflict would provide new opportunities for the Chinese Communists.”¹⁷ This is a reasonable conjecture, but evidence from the Chinese archives shows that it is inaccurate. On the other side, Immanuel Hsu concluded: “Once hostilities began, Japanese reinforcements from Manchuria and the home islands poured into North China, occupying all the strategic points outside [Beijing]. Obviously, the Marco Polo Bridge incident was but the beginning of a much larger design.”¹⁸ This conclusion fits our militarist stereotype of Imperial Japan, but evidence from the Japanese archives suggests a contrary interpretation. The general point is clear: Without archival materials that reflect decisions and calculations on *both* sides of a conflict, it is difficult to interpret accurately the dyadic interactions that led to war.

¹⁶ For instance, the authoritative multi-volume history recently commissioned by the Chinese Institute of Social Science opened the volume on the Third Sino-Japanese War as follows: “[The 1937 Japanese invasion] was definitely not a historical tragedy that evolved out of an inadvertent incident. It was the concrete implementation of a strategy to invade China that was long-dreamed by Japanese imperialists, as well as the inevitable product of Japan’s China policy.” Li Xin, ed., *Zhonghua Minguo Shi* [History of Republican China], Vol. 9, Part 1 (Beijing: Zhonghua Shuju, 2011), 1.

¹⁷ Hata, “Marco Polo Bridge Incident,” 244.

¹⁸ Immanuel Hsu, *The Rise of Modern China* (New York: Oxford University Press, 2000), 583. Running for six editions, this book remains for decades the most authoritative general history of modern China in the English language by a Chinese historian.

This case study may be the first detailed analysis of the 1937 July crisis in the English language that intersects archival evidence from both Chinese and Japanese sources. It reconstructs the sequence of calculations and decisions made in Nanjing and Tokyo during the 1937 crisis. It examines the actions and reactions generated by both sides of the conflict. Its focus is on the interactive perceptions that had characterized the crisis and motivated the decisions for war.

Background

In 1937, China's 29th Army and Japan's China Garrison Army were stationed around the Beijing area. The 29th Army, under General Song Zheyuan, was formally subordinate to the central government in Nanjing under Chiang Kaishek. Chiang did not have full control over the regional army. Song controlled the army, but he had to balance the political pressure from Nanjing and from the Japanese.¹⁹ Japan's China Garrison Army was formed in 1901 after the Boxer War to defend Japanese interests in North China. (Note that the China Garrison Army was a different unit from the Kwantung Army that triggered the Second Sino-Japanese War in 1931.)²⁰ The Army General Staff was the guiding influence in Tokyo during the crisis.²¹ In July 1937, the Army Chief of Staff was Prince Kanin, but his leadership was largely symbolic. The Vice Chief of Staff,

¹⁹ See Marjorie Dryburgh, "Regional Office and the National Interest: Song Zheyuan in North China, 1933-1937," in *Chinese Collaboration with Japan, 1932-1945: The Limits of Accommodation*, ed. David Barrett and Larry Shyu (Stanford, CA: Stanford University Press, 2001), 38-55.

²⁰ Kazuki Kiyoshi, who arrived as the new commander of the China Garrison Army in July 1937, was highly dissatisfied with the dovish attitude he found in his HQ. "[F]rom the moment of his arrival ... he regarded his staff officers scornfully as "a bunch of cowards.'" Hata, "Marco Polo Bridge Incident," 256.

²¹ David Lu, *From the Marco Polo Bridge to Pearl Harbor: Japan's Entry into World War II* (Washington, DC: Public Affairs Press, 1961), 7.

Lieutenant-General Imai Kiyoshi, was on sick leave. Major-General Ishiwara Kanji, Chief of the Operations Division, played a leading role in the General Staff when the crisis erupted.²²

On July 7, a Japanese regiment held a night exercise near the Marco Polo Bridge in the outskirts of Beijing. The regiment heard mysterious gunshots. Then it found one soldier missing.²³ The Japanese demanded to search the nearby Wanping city but the Chinese denied their demands. A clash broke out and sparked the Marco Polo Bridge crisis.²⁴

2. Theoretical Predictions

I present the theoretical tests (Section 2) and the historical narrative (Section 3) separately. In the narrative, I provide a descriptive reconstruction of the decision process based on the archival records, avoiding a selective and theory-laden treatment of the evidence.²⁵ My theoretical interpretations are summarized and kept separately in the headings (under “Observation”) in Section 3; these headings will highlight where the evidence fits with or contradicts the theoretical predictions. I adopt this approach to avoid chopping up the chronological flow of the historical narrative and to preserve the

²² Peattie, *Ishiwara Kanji*, 292; Hata, “Marco Polo Bridge Incident,” 251.

²³ The missing soldier soon came back by himself. But it was too late.

²⁴ For a Japanese account of the incident, see Hata, “Marco Polo Bridge Incident,” 245-9. For a Chinese account, see Zhang Xianwen, *Zhonghua Minguo Shi III: Riben Quanmian Qinhua* [History of Republican China Part III: The Japanese Invasion of China] (Nanjing University Press, 2005), 1-3.

²⁵ Ideally, the findings should be replicable by future scholars operating within the same boundaries of evidence. There is admittedly a touch of Rankean historiographic idealism here. I do not take a position on whether the Rankean ideal is achievable or naive, except to note its usefulness in disciplining the documentary contributions I wish to make.

documentary contributions of this study. Pointers on names, translations, and Romanization are footnoted here.²⁶

Theories and Observable Implications

Theory 1 (Commitment Theory of War)

- **Predictions: (T1a)** Decisions or proposals for escalation or war are more likely when leaders expect a significant increase in their future vulnerability that can be exploited by their opponent. We should also observe **(T1b)** leaders justifying their proposals by arguing that their opponent cannot be trusted to fulfill a peaceful agreement or maintain the status quo after becoming stronger in the future.

Mechanism 1 (Exogenous Enforcement)

- **Predictions: (M1a)** Decisions or proposals against escalation are more likely when leaders expect an external actor to enforce a bargain or status-quo arrangement. We should also observe **(M1b)** leaders justifying their decisions by highlighting the efficacy of the external enforcer.

²⁶ All Chinese names are in *hanyu pinyin* for the rest of this dissertation. I also use the Asian convention of placing the last name first (i.e. “Deng Xiaoping” rather than “Xiaoping Deng”). To minimize confusion, names well known in their Wade-Giles form are kept unchanged, if their *hanyu pinyin* are rarely used in the English-language literature. Hence, “Chiang Kaishek” does not become “Jiang Jieshi” and the “Kwantung Army” does not turn into the “Guandong Army”. In 1937, Beijing was called “Beiping”. In this chapter, I call Beiping “Beijing”.

Mechanism 2 (Inadvertent Enforcement)

- **Predictions: (M2a)** Decisions or proposals against escalation are more likely when leaders are wary of a third-party rival who may take advantage of the war. We should also observe **(M2b)** leaders justifying their decisions by emphasizing the threat from their third-party rival in the event of war.

Mechanism 3 (Endogenous Enforcement)

- **Predictions: (M3a)** Decisions or proposals against escalation are more likely when one's potential vulnerability is reduced through measures undertaken by the opponent. We should also observe **(M3b)** leaders negotiating for the reduction of their future strategic vulnerability and **(M3c)** leaders justifying their proposals against (or for) war based on their belief that the vulnerability reduction could be achieved through negotiations (or not). The breakdown of endogenous enforcement leads to a conflict spiral that is the reverse image of Prediction M3a.

Theory 2 formulates the private-information explanation for war into a general hypothesis: *War is more likely when private information cannot be signaled credibly.* Since we can never know every piece of private information held by each actor (as they are *private*), Theory 2 is extremely hard to test with observational data. It is more useful, therefore, to focus on the observable signals transmitted in a crisis. I break Theory 2 into three specific signaling mechanisms and test them separately:

Mechanism 4 (Costly Signaling)

- **Predictions: (M4a)** Decisions or proposals for escalation are more likely when the opponent does not incur significant cost in signaling its resolve to fight. We should also observe **(M4b)** leaders justifying their decisions by arguing that the opponent's signals are not credible because they are not costly in themselves.

Mechanism 5 (Costly Implementation)

- **Predictions: (M5a)** Decisions or proposals for escalation are more likely when the opponent's threats are costly to implement. We should also observe **(M5b)** leaders justifying themselves by citing the opponent's costs of implementing the threat.

Mechanism 6 (Contradictory Signaling)

- **Predictions: (M6a)** Decisions or proposals for escalation are more likely when leaders have detected a contradiction in the signals of resolve sent by their opponent. We should also observe **(M6b)** leaders justifying their decisions by citing the contradiction in their estimation of the opponent's resolve.

Sino-Japanese War in 1937

Do the theoretical predictions agree with the archival evidence? I address this question by studying if the decisions and proposals made by key decision-makers matched the theoretical predictions. The results are in Table 3.1:

Table 3.1: Predictions vs. Evidence from the Third Sino-Japanese War

Predictions	Empirical Support
T1 (Commitment Theory) Prediction T1a ** Prediction T1b **	Direct Support: Despite progress in the peaceful negotiations at the local level, both Tokyo and Nanjing suspected that the other would not keep its word. Tokyo feared that the agreement would become mere scraps of papers in the future. Nanjing moved the Central Army to North China, believing that “once their reinforcements arrive, [the Japanese] would betray all trust and finish off the [local] 29th Army in one swoop.” ²⁷
M1 (Exogenous Enforcement) Prediction M1a * Prediction M1b	Indirect Support: Unlike in 1931 when Nanjing entrusted the League of Nations to enforce the peace, Nanjing did not see external enforcement as a feasible option in 1937.
M2 (Inadvertent Enforcement) Prediction M2a ** Prediction M2b **	Direct Support: Tokyo initially tried to contain the crisis in the belief that fighting a war in China would create an advantage for the Soviet Union, its rival in the North.
M3 (Endogenous Enforcement) Prediction M3a ** Prediction M3b * Prediction M3c *	Direct Support: Negotiations at the local level focused on the mutual withdrawal of troops to maintain the military status quo between both sides. But the peaceful settlement was broken even before it was finalized. The commitment problem intensified as each side sent reinforcements and increased the potential vulnerability of the other.
M4 (Costly Signaling) Prediction M4a # Prediction M4b	Contradiction: Costly signals through military deployments on both sides increased the perceived vulnerability in both Nanjing and Tokyo. This promoted hawkish reactions on both sides, which increased rather than decreased the risk of war.
M5 (Costly Implementation) Prediction M5a * Prediction M5b	Indirect Support: Pro-war Japanese leaders argued that China could not fight a protracted war with Japan, given the relative military balance. Dissenters argued the opposite.
M6 (Contradictory Signaling) Prediction M6a ** Prediction M6b **	Direct Support: Efforts to reach a peaceful settlement at the local level were contradicted by the troop reinforcements authorized by Nanjing and Tokyo, making each doubt the other’s commitment towards a peaceful resolution.

Notes: ** Direct support for the prediction. * Indirect support for the prediction. # Prediction contradicted.

²⁷ Telegram from Defense Minister He Yingqin to General Song Zheyuan, 15 July 1937, KRZZ, 187.

The 1937 case supports the commitment theory of war and the three enforcement mechanisms. The evidence is relatively strong. The mechanisms make unique predictions, many of which matched both the observed outcomes as well as the details in the speech evidence from top leaders as they deliberated for and against war. But the evidence for the signaling mechanisms is mixed. In particular, the evidence does not support Mechanism 4 (costly signaling). In fact, the 1937 case suggests that war becomes more and not less likely when an action that transmits the costly signal (e.g. military mobilization) also increases the vulnerability of the opponent. This suggests that our theoretical expectations should be recalibrated in an environment where sending the costly signal also changes the military balance endogenously.

The next section provides the historical analysis. It is a descriptive reconstruction of the decision processes leading to war, based on the defined limits of my archival evidence. The reconstruction focuses on what was discussed by the key decision-makers. I separate my archival observations from my theoretical interpretation of those observations. My interpretations are encapsulated in the headings (“Observation”), which indicate where the evidence fits or contradicts the theoretical predictions.

3. Historical Analysis

Observation 5: Key policymakers in Tokyo sought the non-expansion of hostilities when the crisis broke out, in view that fighting a war in China would create a strategic advantage for the Soviet Union. This finding supports Mechanism 2 (inadvertent enforcement).

Observation 5 makes two claims. Claim (1) is that key policymakers in Tokyo sought the non-expansion of hostilities when the crisis erupted. Claim (2) is that the policymakers believed that fighting a war in China would create a strategic advantage for the Soviet Union.²⁸ Both claims must hold for Observation 5 to hold.

To establish Claim (1), I investigate Tokyo's reactions when the crisis broke out. What were Tokyo's military orders and political decisions in the first three days of the crisis? A key piece of evidence is Order No. 400 issued by the Army General Staff on 8 July and undersigned by Chief of Staff Prince Kanin. The order went straight to the point: "to prevent the incident from escalating, [the field army must] avoid the further use of military force."²⁹

On the same day, a plan was drafted at the Army General Staff for dealing with the crisis. According to the draft plan, it was necessary to limit the crisis within the Beijing-Tianjin

²⁸ Claim (2) implies that the perception was influential at the start of the crisis, but it does not imply that all policymakers throughout the crisis agreed that the strategic threat from the Soviet Union would be best addressed by avoiding a war in China. See Observation 11.

²⁹ Order No. 400 of the Japanese Army Chief of Staff, 8 July 1937, HBSJ, 831.

area and to maintain security within the area; thus the army should proceed according to the basic policy of not expanding hostilities. However, the draft plan also indicated that if the Chinese side adopted a provocative attitude, it would be necessary to dispatch forces to reinforce the field army and to expel Chinese forces from the Beijing-Tianjin area.³⁰ Also on the same day, Army Minister Sugiyama Gen ordered divisions west of Tokyo to suspend the discharge of infantry soldiers originally scheduled on 10 July.³¹

At 8:50 am on 9 July, the Cabinet held an emergency meeting. In the meeting, Army Minister Sugiyama told the Cabinet that “in view of the troop numbers and the antagonistic attitude of the 29th Army, [I am] considering reinforcements from Japan of a strength amounting to about three divisions.”³² However, most of the Cabinet did not concur. In the same morning, the field army reported that a ceasefire agreement had been reached. As such, the Cabinet decided to hold off action. At 11 am on the same day, a five-minister meeting was convened for further deliberations. The meeting decided to stick to the policy of non-expansion, with the hope that the crisis could be quickly resolved. But if China did not back down and the conflict escalated to a dangerous level, Japan would have to take “appropriate measures.”³³

Was the field army keen on expanding the crisis? It turned out that the China Garrison Army HQ was moving towards a localized peaceful settlement. A ceasefire agreement was reached in the morning of 9 July. On 11 July, the China Garrison Army and the

³⁰ DHR, 305-6.

³¹ DHR, 300.

³² DHR, 305.

³³ Ibid.

Chinese 29th Army successfully reached a local settlement on the basis of mutual withdrawal of their troops.³⁴ But on 12 July, Lieutenant-General Kazuki Kiyoshi, a hardliner, arrived as the new commander of the China Garrison Army, as the previous commander was seriously ill. Dissatisfied with the dovish attitudes in his HQ, he sent a hawkish recommendation to Tokyo on 13 July.³⁵ Tokyo responded by sending two senior officers with instructions for Kazuki to observe the government's non-expansion policy. In the end, Kazuki was persuaded by the arguments of his chief of staff and a private letter from the Emperor's chief aide-de-camp indicating the Emperor's desire for a peaceful solution.³⁶ On the whole, while policymakers in Tokyo did not eliminate the possibility of military action, the evidence suggests that the non-expansion of hostilities formed the policy baseline when the crisis broke out (Claim 1).

The most direct evidence for Claim (2) comes from Major-General Ishiwara Kanji, who was a leading influence in the Army General Staff when the crisis erupted. Ishiwara explained his thoughts in an interview with Prince Takeda at the Army General Staff's War History Section:

Prince Takeda: [What was] the reason behind the non-expansion policy at the start of the incident?

Ishiwara: Because there should not be a war between Japan and China; because such a war could not be ended quickly; [and because of] the view that no matter what, this obstacle [incident] must be overcome. So I was determined to insist on the non-expansion policy. But a key factor in my determination was the consideration over the war with the Soviet Union. That is, in the event of a

³⁴ Hata, "Marco Polo Bridge Incident," 250.

³⁵ DHR, 316.

³⁶ Hata, "Marco Polo Bridge Incident," 256.

protracted war [with China], if the Soviet Union attacks, Japan at this point would not be ready to handle it.³⁷

...

Prince Takeda: At that time [July 1937] were there completely no plans for a full-scale war [with China]?

Ishiwara: Not possible to have one. Because we did not even have sufficient military strength to deal with the Soviet Union.³⁸

The interview was done in 1939, two years *after* the 1937 crisis. If the Soviet threat was critical to the High Command, it should be reflected in the military doctrines written *before* the crisis erupted. This is indeed the case. The “Outline of National Defense Policy” by the Army General Staff in June 1936 explicitly stated: “The first task is to concentrate our full strength to deal with the Soviet Union.”³⁹ In the Army’s outline of its five-year production plan in May 1937, the High Command emphasized the need to complete its preparations for a Soviet-Japanese war in five years time and to avoid any conflict in China.⁴⁰ In conjunction, the Japanese government made a remarkable shift in its China policy. At the four-minister conference in April 1937, the government decided to adopt “a fair attitude towards the Nanjing government and its efforts in the national unification of China”, so as to “eliminate the root causes of the anti-Japanese attitudes in China”.⁴¹ Japan’s policy in North China was to create an anti-communist zone

³⁷ Record of Interview between Prince Takeda and Ishiwara Kanji, 1939, DHR, 334.

³⁸ *Ibid*, 336.

³⁹ Outline of National Defense Policy by the Japanese Army General Staff, 30 June 1936, DHR, 281.

⁴⁰ DHR, 295.

⁴¹ Strategic Policy on China, 16 April 1937, RDZDQ, 215.

sympathetic to Japan's interests.⁴² On 18 June, the Army HQ reiterated to the field army its strict prohibition against adventurist activities in China.⁴³ In the Cabinet meeting on 6 July, Foreign Minister Hirota Koki reported that Japan's policy towards China remained unchanged, but it would be hard to achieve Sino-Japanese friendship given the strong anti-Japanese sentiments on the ground. While Japan was dissatisfied with the present situation, there was no better alternative than to continue resolutely with the existing policy.⁴⁴ The Cabinet concurred. The Marco Polo Bridge crisis erupted on the next day.

Observation 6: Chiang Kaishek distrusted Japan's commitment to a localized settlement and made preparations for a Japanese invasion.⁴⁵ This finding supports Theory 1 (commitment theory of war).

Chiang also feared that the Japanese would sway General Song Zheyuan to break off North China from Nanjing's control. This supports Theory 1 through a domestic-level channel.

⁴² To achieve this goal, Japan would "rely primarily on economic means and not promote any measures that would create political fragmentation in China"; Japan would use its efforts "to eliminate China's fears and suspicions of Japan on the one hand, and to facilitate economic cooperation and cultural understanding on the other." Ibid.

⁴³ DHR, 297.

⁴⁴ Ibid.

⁴⁵ The origins of Nanjing's deep distrust lie beyond the scope of my analysis; they must be traced (at least) from the 1931 Manchuria incident. Suffice to say that what the Japanese army did in China since 1931 had inspired little faith. The Tangu truce, the He-Umezumi agreement, the Doihara-Qin agreement, and the establishment of the East Hebei autonomous council and the Hebei-Chahar political council had consecutively threatened China's sovereignty and territorial integrity.

“The Japanese fear the Chinese and the Chinese distrust the Japanese. If these two types of mentalities are not removed, then each side will be suspicious of the other, and all kinds of tragic consequences will ensue.”

– Chiang Kaishek, January 1936 ⁴⁶

On 8 July 1937, Chiang Kaishek was informed of the clash at Marco Polo Bridge.⁴⁷ Chiang reacted quickly.⁴⁸ He ordered Song Zheyuan, the Commander of the 29th Army, to “resolutely defend Wanping city and not to withdraw.” “The entire [29th Army] should be activated to prepare for the [potential] escalation of the incident. I am prepared to provide reinforcement immediately.”⁴⁹ In a follow-up message to Song, Chiang announced his intention to send four divisions as reinforcement.⁵⁰ On the same day, Chiang started deploying the divisions and ordered the Military Commission at Nanjing

⁴⁶ Dialogue with Chiang Kaishek (“The Path to Resolving the Issues Between China and Japan”), 13 January 1936, XZJS, Vol. 38, 33.

⁴⁷ Telegram from Yan Kuan to He Yingqin, 8 July 1937, KRZZ, 178; Telegram from He Yingqin to Chiang Kaishek, 8 July 1937, KRZZ, 179.

⁴⁸ What caused Chiang’s strong reaction to the crisis? The evidence suggests that Chiang was reacting directly to the Japanese threat. A confluence of factors might have motivated his reaction: a deep distrust of the Japanese, an increased confidence in the capabilities of the Central Army, and the importance of defending Beijing. Van De Ven highlighted the strategic significance of the Marco Polo Bridge, which was near the junction of the Beijing-Hankou and Beijing-Suiyuan railroads: “If the Japanese seized it, Beijing’s defences could not be reinforced and the Japanese could then use these railroads to fan out across north China and advance towards the north-west.” Peattie argued that Chinese victory in the 1936 Sino-Japanese clash in Suiyuan “played a major part in China’s eventual determination to undertake ... a major military confrontation with Japan.” A connection has also been drawn between the Xian Incident of December 1936 and Chiang’s reaction to the crisis in July 1937. For instance, Wong argued that although Chiang “certainly could have disregarded” the anti-Japanese pledge he made to Zhang Xuqiang when he was held hostage by the latter in Xian, Chiang’s “sudden rise in popularity upon his release, due to his identification now with the rising tide of anti-Japanese nationalism, made it more than ever difficult for him to oppose resistance in favor of continuing the civil war ... There was indeed no longer any room in which Chiang could maneuver, but he was still determined to make it appear that he himself had brought about the change in policy, not the Xi’an Incident.” There is no direct support for the Xian connection based on the records within my evidential boundaries. Of course, it is rather difficult to prove the connection with archival evidence in the first place, as the last sentence in the previous quote suggests. Van De Ven, *Nationalism and War*, 190; Peattie, *Ishiwara Kanji*, 279; Wong Young-tsu, “The Xi’an Incident and the Coming of the War of Resistance,” in *China in the Anti-Japanese War, 1937-1945: Politics, Culture, and Society*, ed. David Barrett and Larry Shyu (New York: Peter Lang, 2001), 24.

⁴⁹ “8 July 1937,” ZJDC, Vol. 4, Part 1, 1114.

⁵⁰ *Ibid.*

to prepare for a full mobilization.⁵¹ On 9 July, Chiang informed Song that two divisions were already dispatched.⁵² At the same time, Defense Minister He Yingqin was summoned back to Nanjing to prepare the armed forces for a Japanese attack.⁵³

Chiang implemented a series of emergency measures on 10 July. He ordered one hundred divisions to be organized for combat and eighty divisions for support. The stockpile of ammunition would be distributed to separate locations. In addition, Chiang ordered the preparation of one million troops in reserve as well as six months of battle provisions.⁵⁴ He also asked Song to hasten the completion of works along the defense frontline.⁵⁵ In a separate telegram, Chiang told Song: “The defense of our land calls for a spirit of active readiness and a determination to fight to the death. As for negotiations [with the Japanese], you must be especially wary of their usual treacherous tactics, and abide by the principle that not even an inch of our sovereignty can be compromised.”⁵⁶

What were the calculations behind Nanjing’s strong and decisive stance? From the beginning, Chiang was highly distrustful of Japan’s intentions and its commitment towards a peaceful resolution. Chiang wrote in his personal diary on 8 July: “[The Japanese thieves] have triggered a conflict at Marco Polo Bridge. Are they trying to force us into submission by exploiting a period in which our preparations [against Japan] are

⁵¹ “8 July 1937,” ZJDC, Vol. 4, Part 1, 1115.

⁵² Telegram from Chiang Kaishek to Song Zheyuan, 9 July 1937, KRZZ, 179; Telegram from Chiang Kaishek to Sun Lianzhong, 9 July 1937, KRZZ, 179.

⁵³ “9 July 1937,” JJNP, 248.

⁵⁴ “10 July 1937,” JJNP, 248-9.

⁵⁵ Telegram from Chiang Kaishek to Song Zheyuan, 9 July 1937, KRZZ, 181.

⁵⁶ Telegram from Chiang Kaishek to Song Zheyuan, 9 July 1937, KRZZ, 180.

not yet completed? Or are they trying to sway Song Zheyuan and promote the independence of North China?"⁵⁷ Chiang also asked himself: "[The Japanese thieves] have challenged us. Is this the time to accept the challenge with a determined resistance?"⁵⁸ Judging from Chiang's actions, the answer was yes. China must preempt Japan's malign intentions quickly and decisively. In a telegram to the field army on 11 July, Chiang said: "Unless our forces are fully prepared and show the determination to fight to the death, there is absolutely no hope for a peaceful settlement."⁵⁹ On 12 July, he told Song: "The Japanese would launch a full-scale attack on the fifteen (of July). This intelligence is highly accurate. I hope you are not deceived by their delaying tactics."⁶⁰ On 13 July, Chiang told Song: "The Marco Polo Bridge incident certainly cannot be resolved peacefully The central government has decided to commit its full efforts towards a military resistance. We will choose honorable sacrifice over dishonorable self-preservation [To achieve victory] you should stand together with the central government and avoid being deceived by the enemy."⁶¹ Defense Minister He Yingqin also warned Song that while the crisis might appear to have calmed down at this point, "once their reinforcements arrive, [the Japanese] would betray all trust and finish off the 29th Army in one swoop."⁶² In the meantime, divisions from the Central Army were moving towards North China to forestall the Japanese.

⁵⁷ Excerpt of Chiang Kaishek's Diary, in "8 July 1937," ZJDC, Vol. 4, Part 1, 1115.

⁵⁸ Ibid.

⁵⁹ "11 July 1937," ZJDC, Vol. 4, Part 1, 1121.

⁶⁰ "12 July 1937," ZJDC, Vol. 4, Part 1, 1123.

⁶¹ "13 July 1937," JJNP, 249.

⁶² Telegram from He Yingqin to Song Zheyuan, 15 July 1937, KRZZ, 187.

Observation 7: Contradictory signals coming from Japanese military activities increased Nanjing’s distrust. Nanjing feared that Japan would renege on a peaceful settlement reached at the local level, once the military balance tilted in Japan’s favor. The evidence connects to Theory 1 (commitment theory), Mechanism 3 (endogenous enforcement) and Mechanism 6 (contradictory signaling).

Observation 8: Song Zheyuan tried to avoid war with the Japanese. The war was likely to decimate his troops and threaten his political position in North China. This suggests a domestic-level manifestation of Mechanism 2 (inadvertent enforcement).

A field report on 9 July claimed that the Marco Polo Bridge incident was clearly part of a premeditated ploy by the Japanese army; but the report also highlighted that the Japanese had initiated contact with the Chinese authorities and “appeared to have the intention to make a compromise with our side and prevent the incident from escalating.”⁶³ A separate report reaffirmed that the Japanese had expressed their desire not to expand the conflict; but the report also warned that the Japanese gesture could be a tactical prelude to an offensive attack.⁶⁴ A subsequent telegram suggested that “the Japanese has the habit of exploiting tactical opportunities, and it appears to be the case in view of the current situation.”⁶⁵

⁶³ Telegram from Yu Feipeng to He Yingqin et al., 9 July 1937, KRZZ, 164.

⁶⁴ Telegram from Yan Kuan to He Yingqin, 9 July 1937, KRZZ, 180.

⁶⁵ Telegram from Yan Kuan to He Yingqin, 10 July 1937, KRZZ, 181.

Japan's negotiation demands were reported as follows: (i) the withdrawal of Chinese troops from the Marco Polo Bridge area; (ii) the punishment of perpetrators in the incident; (iii) the full suppression of anti-Japanese activities in North China; and (iv) joint cooperation against communism.⁶⁶ These demands were communicated in a field report on 11 July. At the same time, the report highlighted the details of a "finalized" offensive plan by the Japanese to invade major parts of North China, and concluded that "the incident seemingly cannot be resolved in a simple way."⁶⁷

But while divisions from the Central Army moved towards the north, local authorities in North China moved towards a peaceful settlement. Beijing Mayor Qin Dechun informed Nanjing on 10 July: "The current situation is heading in a positive direction. If the central authorities have not yet completed its preparations for a major battle, or if it is concerned [that mobilizing the divisions] would have repercussions that would escalate [the crisis], may I request a temporary halt of the northward troops at their current positions? If there is a need [for them] later, I will request accordingly."⁶⁸ On 12 July, Qin informed Nanjing that the Japanese had withdrawn their troops and expressed their desire to avoid the recurrence of similar incidents.⁶⁹ Qin, however, left a caveat that despite the positive gestures, "I cannot say for sure whether the Japanese will keep their word."⁷⁰

⁶⁶ Telegram from Yan Kuan to He Yingqin, 11 July 1937, KRZZ, 182.

⁶⁷ Ibid.

⁶⁸ Telegram from Qin Dechun to Qian Dajun, 10 July 1937, KRZZ, 165.

⁶⁹ Telegram from Qin Dechun to Qian Dajun, 12 July 1937, KRZZ, 166.

⁷⁰ Telegram from Qin Dechun to Chiang Kaishek, 12 July 1937, KRZZ, 166.

On the very same day, negative gestures were observed. The Chinese sighted Japanese troops, tanks, ammunition and other combat provisions moving towards the Beijing-Tianjin area.⁷¹ Nanjing feared that Japan's Kwantung Army might strike from the north and warned Qin Dechun to hasten his battle preparations.⁷² Subsequent reports between 13 and 14 July alerted Nanjing that Japanese planes had commenced their air maneuvers in the area and that troop reinforcements were moving in continuously from the north.⁷³ Intelligence sources suggested that the Japanese were making a major effort to turn Song Zheyuan against Nanjing.⁷⁴

However, Nanjing did not stop the local authorities from continuing their negotiations with the Japanese.⁷⁵ In fact, some commanders in the High Command felt that China should be careful not to expand the crisis, as the Central Army was not fully prepared for war.⁷⁶ While highly suspicious of the Japanese, the central government did not eliminate the possibility of a diplomatic solution. Chiang told party members: "We are preparing for war, but we are definitely not asking for war." "In the final second before peace becomes impossible, we will still hope for peace. We hope that the incident can be

⁷¹ Telegram from Qin Dechun to Qian Dajun, 12 July 1937, KRZZ, 183.

⁷² Telegram from Qian Dajun to Qin Dechun, 12 July 1937, KRZZ, 183.

⁷³ Telegram from Yan Kuan to He Yingqin, 13 July 1937, KRZZ, 184; Telegram from Xiong Bin to Chiang Kaishek, 14 July 1937, KRZZ, 185-6; Telegram from Yan Kuan to He Yingqin, 14 July 1937, KRZZ, 186-7; Telegram from Qin Dechun to Qian Dajun, 14 July 1937, KRZZ, 187.

⁷⁴ Minutes of the 4th Joint Meeting of the Military High Command After the Marco Polo Bridge Incident, 14 July 1937, JMRA, 6.

⁷⁵ Minutes of the 2nd Joint Meeting of the Military High Command After the Marco Polo Bridge Incident, 12 July 1937, JMRA, 4-5.

⁷⁶ Minutes of the 4th Joint Meeting of the Military High Command After the Marco Polo Bridge Incident, 14 July 1937, JMRA, 7.

peacefully resolved through diplomatic means.”⁷⁷ Defense Minister He Yingqin told Song that he “should not abandon the possibility of peace on the one hand, but should also make his military preparations secretly on the other hand.”⁷⁸ At Nanjing, the Foreign Vice-Minister was tasked to meet with the Japanese embassy on 12 July to probe the Japanese sincerity for a peaceful resolution. If the Japanese “are indeed sincere, we may be able to negotiate with them.”⁷⁹

But the intelligence received by Nanjing cast doubt on Japan’s commitment towards a peaceful solution. Pondering over the intelligence reports, Chiang wrote in his diary on 12 July: “The Kwantung Army has entered the Tianjin area. Yesterday the [Japanese] Cabinet held an emergency meeting. All parties and industries in Japan have pledged their support to the Cabinet. In my view, the conflict will definitely expand. We must definitely make active preparations.”⁸⁰ The next day, Chiang told Song: “The central government is fully prepared to declare war and has no qualms to live or die together with your troops. Overall, victory or defeat will depend on your unity with the central government. Whether it is war or peace, you must certainly not act on your own.”⁸¹ On the same day, Chiang ordered He Yingqin to reinforce the 29th Army with weapons and ammunition, and to construct storage and air defense facilities in the area.⁸² On 15 July, Chiang cabled the local authorities at Jinan and Qingdao that two Japanese divisions were

⁷⁷ Excerpt of Chiang Kaishek’s Speech at the Lushan Conference, in “17 July 1937,” JJNP, 250.

⁷⁸ Telegram from He Yingqin to Song Zheyuan, 17 July 1937, KRZZ, 188-9.

⁷⁹ Minutes of the 3rd Joint Meeting of the Military High Command After the Marco Polo Bridge Incident, 13 July 1937, JMRA, 5.

⁸⁰ Excerpt of Chiang Kaishek’s Diary, in “12 July 1937,” ZJDC, Vol. 4, Part 1, 1123.

⁸¹ “13 July 1937,” ZJDC, Vol. 4, Part 1, 1124.

⁸² Telegram from Chiang Kaishek to He Yingqin, 14 July 1937, KRZZ, 185.

moving towards their area, and that Jinan and Qingdao must prepare themselves quickly.⁸³

Over the next two days, Nanjing heard news of further military deployments by the Japanese. By 17 July, Nanjing was informed that Japan had mobilized five divisions from the home islands of Japan plus one division from the Korea Army.⁸⁴ At the same time, the Japanese Embassy met the Chinese foreign minister to express Japan's desire not to escalate the conflict. China was also told that the conflict could be peacefully resolved as long as the authorities in North China fulfill its commitment to the local agreement signed on 11 July. But Nanjing was doubtful of Japan's intentions. Defense Minister He Yingqin felt that Japan was using a tactic to promote the autonomy of North China from the central government.⁸⁵ Judging from Japan's military deployments, He felt the Japanese were using a delaying tactic to disarm their opponent before using superior force to wipe out the 29th Army in North China.⁸⁶

On 17 July, Chiang declared that a peaceful solution to the crisis must satisfy four conditions: (i) It cannot affect the sovereignty and territorial integrity of China; (ii) it cannot create any illegal political change in North China; (iii) it cannot remove any official in North China appointed by the central government; and (iv) it cannot impose

⁸³ "15 July 1937," ZJDC, Vol. 4, Part 1, 1125.

⁸⁴ Telegram from He Yingqin to Song Zheyuan, 17 July 1937, KRZZ, 188-9.

⁸⁵ Minutes of the 7th Joint Meeting of the Military High Command After the Marco Polo Bridge Incident, 17 July 1937, JMRA, 9.

⁸⁶ Telegram from He Yingqin to Song Zheyuan, 17 July 1937, KRZZ, 188-9.

any constraint on the 29th Army.⁸⁷ Subsequently, Defense Minister He met with the Japanese emissary, who warned that if the Chinese reinforcements did not withdraw, the conflict would escalate and Japan would react with a serious resolve. Defense Minister He rebutted that the Chinese reinforcements were purely defensive and Nanjing would consider their withdrawal only after the withdrawal of the Japanese reinforcements.⁸⁸ The meeting ended in disagreement. On 19 July, Chiang informed Defense Minister He that the decision to fight a war of resistance had been made, and that the minister should start making plans to incorporate the Chinese Communist Party in the war effort.⁸⁹

Suddenly, a peaceful solution was in sight. Song Zheyuan visited Japan's China Garrison Army HQ on 18 July to iron out a settlement.⁹⁰ On 19 July, Song informed Nanjing that "both sides hope to revert promptly to the situation that had existed before 8 July" and he requested for the Defense Minister's "utmost patience" while he resolve the matter.⁹¹ In effect, Song had taken his own initiative without consulting the central government. Song withheld the details of his settlement from Nanjing, and Nanjing found itself in a bind.⁹²

⁸⁷ "17 July 1937," JJNP, 249.

⁸⁸ "19 July 1937," ZJDC, Vol. 4, Part 1, 1132.

⁸⁹ "19 July 1937," JJNP, 250.

⁹⁰ Telegram from Qin Dechun to Qian Dajun, 19 July 1937, KRZZ, 190; Telegram from Xiong Bin to He Yingqin, 19 July 1937, KRZZ, 171.

⁹¹ Telegram from Song Zheyuan to He Yingqin, 19 July 1937, KRZZ, 170.

⁹² "19 July 1937," ZJDC, Vol. 4, Part 1, 1133. See also: Excerpt of Chiang Kaishek's Diary, in "23 July 1937," ZJDC, Vol. 4, Part 1, 1134.

But the Japanese army continued its military maneuvers on land and in air, with episodic violence breaking out around the frontline.⁹³ On 21 July, Wanping city was shelled by Japanese fire.⁹⁴ Defense Minister He warned Song: “The Japanese are plotting against us. Their ambitions remain unchanged. Judging from their moves, we cannot gauge their true intentions.”⁹⁵ On 22 July, a field report told Nanjing that the Chinese troops were ready to fight if the Japanese intruded, and warned: “These two days are critical.”⁹⁶ On 24 July, Zhang Zhizhong, Mayor of Tianjin, tried to meet General Kazuki to check on the withdrawal of Japanese troops based on the earlier agreement on 18-19 July. Zhang was told that Kazuki was sick and could not meet him. Kazuki’s “illness” aroused deep suspicions.⁹⁷ At the same time, the Japanese troops continued to increase in numbers.⁹⁸ On the same day, Chiang alerted Song that Japan had made major deployments around North China in the last two days. Chiang predicted: “There will certainly be a large-scale operation within a week. I hope you remain prepared at all times ... Do not be deceived.”⁹⁹ Chiang also cabled Ambassador Kong Xiangxi in London: “[T]he Japanese will soon make a surprise move or serve an ultimatum, or do something like that ... If Britain and the United States can send a serious joint warning [to Japan before it made its move], we may be able to avoid the coming disaster.”¹⁰⁰

⁹³ Telegram from Xiong Bin to Chiang Kaishek, 20 July 1937, KRZZ, 192; “21 July 1937,” ZJDC, Vol. 4, Part 1, 1134.

⁹⁴ Telegram from Song Zheyuan to Chiang Kaishek et al, 27 July 1937, KRZZ, 196-7.

⁹⁵ Telegram from He Yingqin to Song Zheyuan, 20 July 1937, KRZZ, 171-2.

⁹⁶ Telegram to He Yingqin, 22 July 1937, KRZZ, 193.

⁹⁷ “24 July 1937,” ZJDC, Vol. 4, Part 1, 1136.

⁹⁸ Telegram from Yan Kuan to He Yingqin, 24 July 1937, KRZZ, 194.

⁹⁹ “24 July 1937,” ZJDC, Vol. 4, Part 1, 1136. Around this time, Song cabled the High Command to ask for the details of its war plans against the Japanese. Minutes of the 15th Joint Meeting of the Military High Command After the Marco Polo Bridge Incident, 25 July 1937, JMRA, 17.

¹⁰⁰ “24 July 1937,” ZJDC, Vol. 4, Part 1, 1136-7.

The crisis exploded the very next day with a reported attack by Japanese troops in the Langfang area.¹⁰¹ On 26 July, the field army reported that the Japanese had launched an assault at Guangan Gate.¹⁰² General Kazuki served an ultimatum to Song on the same day. Song rejected it.¹⁰³ A field report told Nanjing: “There is no more hope for peace.”¹⁰⁴ On 28 July, Japanese forces attacked Beijing. Beijing fell in a day. On 31 July, Chiang proclaimed: “Peace is impossible now. We can only resist to the end. We must bear all sacrifices and fight the [Japanese] thieves to death.”¹⁰⁵ The eight-year Sino-Japanese War had commenced.

Observation 9: Contradictory signals coming from Chiang Kaishek’s military movements increased Tokyo’s distrust in a localized peaceful settlement. The evidence connects to Theory 1 (commitment theory), Mechanism 3 (endogenous enforcement) and Mechanism 6 (contradictory signaling).

In this section, I map the moves and calculations made by Tokyo since 10 July.

¹⁰¹ Telegram from Song Zheyuan to He Yingqin, 26 July 1937, KRZZ, 195; Telegram from Yan Kuan to He Yingqin, 26 July 1937, KRZZ, 196. Separate evidence from the Nanjing archives and the Tokyo archives lead to contradictory interpretations of the Langfang and Guangan incidents. See the corresponding account based on Japanese archives in the next section.

¹⁰² Telegram from Song Zheyuan to He Yingqin, 26 July 1937, KRZZ, 195; Telegram from Yan Kuan to He Yingqin, 26 July 1937, KRZZ, 196.

¹⁰³ Telegram from Song Zheyuan to Chiang Kaishek et al, 28 July 1937, KRZZ, 199.

¹⁰⁴ Telegram from Yan Kuan to He Yingqin, 27 July 1937, KRZZ, 198-9.

¹⁰⁵ Excerpt of “Proclamation to All Soldiers in the War of Resistance,” in “31 July 1937,” JJNP, 252.

Despite the ceasefire agreement on 9 July, intelligence reports to the High Command suggested that the crisis was escalating. Most crucial was the intelligence on 10 July that Chiang had mobilized the air force of the Central Army and ordered four divisions to advance northwards.¹⁰⁶ These developments strengthened the arguments of the hawkish elements within the Army General Staff.¹⁰⁷ The High Command was concerned that the 4,000-strong field army would not be able to defend itself and the 12,000 Japanese citizens in the area. According to a report from the Army General Staff on 10 July, the field army was surrounded by Chinese troops and trapped in a precarious situation. To defend against the 29th Army and the advancing Central Army, the mobilization of three divisions from Japan was necessary.¹⁰⁸ After an internal discussion, the General Staff agreed that it would send representatives to understand the situation in the field and to prevent the crisis from expanding; and that if forced to the corner, the army would have to fight a quick and decisive battle to prevent negative repercussions on a potential war with the Soviet Union.¹⁰⁹

On 11 July, an emergency Cabinet meeting was convened. The Cabinet concluded that the incident “was undoubtedly an anti-Japanese military action planned by the Chinese”.¹¹⁰ The Cabinet agreed that the Army should prepare to mobilize three

¹⁰⁶ DHR, 308. As Crowley pointed out, the Central Army’s move into North China had diplomatic and military complications as it contradicted the He-Umezu accord. According to Peattie, hawkish elements in the Army General Staff used this information to push for troop mobilization. Crowley, *Japan’s Quest*, 329; Peattie, *Ishiwara Kanji*, 296.

¹⁰⁷ Hawkish elements in Tokyo included Army Minister Sugiyama and the key leaders in the Operations and China sections of the Army General Staff. DHR, 298; Hata, “Marco Polo Bridge Incident,” 251. Observation 11 describes the views of the hawkish faction.

¹⁰⁸ DHR, 307.

¹⁰⁹ DHR, 311.

¹¹⁰ Cabinet Decision on the Marco Polo Bridge Incident, 11 July 1937, HBSJ, 831-2.

divisions from Japan to North China.¹¹¹ The military preparations, however, would cease if Japan received a satisfactory reply to its demands.¹¹² The government proclaimed that “despite the agreement by the 29th Army for a peaceful settlement”, the Chinese had continued their attacks, made preparations for war, and mobilized the Central Army. These actions showed that China “had no sincerity to conduct peaceful negotiations and had fully rejected the possibility of a local settlement.”¹¹³ On the same day, the Army and Navy agreed on a set of operational principles in North China: “(1) To do the utmost to restrict the combat area within the Beijing-Tianjin area, and to avoid military force in Central and South China as a matter of principle (2) Army and Navy should coordinate their combat operations. (3) To do the utmost to prevent triggering any incidents with a third country in the implementation of the [North China] operations.”¹¹⁴ Meanwhile, the Army General Staff released the mobilization orders.¹¹⁵

In the same afternoon, good news arrived: The local authorities in North China informed Japan’s China Garrison Army that it would “fully accept the requests made by Japan on 10 July.”¹¹⁶ The Chinese 29th Army would apologize to Japan, punish those responsible for the crisis, and prevent similar incidents from recurring. Anti-Japanese societies would

¹¹¹ According to Hata, Navy Minister Yonai had opposed the deployment, and Prime Minister Konoe, Foreign Minister Hirota, and Finance Minister Kaya were unenthusiastic about it; but “with little discussion the army’s proposal was accepted, with the proviso that the principle of nonextension of the conflict be observed and that following mobilization, actual dispatching of troops could still be cancelled if there were no necessity for it.” Hata, “Marco Polo Bridge Incident,” 252.

¹¹² *Ibid.*

¹¹³ Government Proclamation on Troop Deployment, 11 July 1937, HBSJ, 833-4.

¹¹⁴ Excerpt of “On the Army-Navy Agreement Regarding Military Operations in North China,” 11 July 1937, in DHR, 311-2.

¹¹⁵ Order No. 56 of the Japanese Army General Staff, 11 July 1937, HBSJ, 832-3; Order No. 57 of the Japanese Army General Staff, 11 July 1937, HBSJ, 833.

¹¹⁶ DHR, 312.

be suppressed; and Chinese troops would not be stationed around the Japanese base in the vicinities of Wanping city and the Longwang temple.¹¹⁷ In a quick reversal, the High Command froze the mobilization orders and terminated the procedures to obtain imperial sanction. The General Staff regained confidence that a local settlement would resolve the crisis. Not all reservations were dispelled, however. In a move with important consequences, the General Staff had maintained the orders to activate the Kwantung Army and Korea Army as a contingency plan.¹¹⁸ An informal comment by the Army Ministry spokesman suggested: "It is rumored, according to dispatches from North China, that the Chinese have accepted all the conditions presented by our Garrison Army.... But if we, on the basis of a mere oral agreement, should trust the Chinese, we would only be deceived by them again."¹¹⁹ The Tokyo Broadcasting Station also cast doubt on whether the peaceful settlement was made out of sincerity, fearing that it would become mere "scraps of papers" in the future.¹²⁰

The reservations resurfaced on 12 July, when the Nanjing government informed the local authorities in North China that any settlement with Japan must first be approved by Nanjing.¹²¹ In response, the Japanese Embassy at Nanjing warned the Chinese Foreign Minister: "To contain the crisis, it is most important that China fulfill the agreement

¹¹⁷ DHR, 313.

¹¹⁸ DHR, 312. It seems that intelligence from Military Attache Okido Sanji in Nanking had influenced this decision. According to Okido, Nanjing had mobilized its air force and amassed four Central Army divisions along the northern border of Henan province. Hata, "Marco Polo Bridge Incident," 252.

¹¹⁹ Hata, "Marco Polo Bridge Incident," 252. Hata pointed that because of "the government's statement of 11 July and its campaign to unify public opinion, it appeared to observers that Japan was now prepared to embark on an aggressive war in China."

¹²⁰ DHR, 313.

¹²¹ DHR, 315.

faithfully. Nanjing's stubborn obstruction of the process and its orders to mobilize the Central Army towards the north will escalate the incident and lead to tragic consequences."¹²² In the evening of 12 July, mobilization orders were issued by Nanjing to the divisions of the Central Army located around the Yangtze and Yellow Rivers. ¹²³

In the 13 July Cabinet meeting, Army Minister Sugiyama informed that the Chinese side was not making progress in implementing a local settlement and that China's Central Army was also advancing to the north. According to Sugiyama, "The sincerity of the Nanjing government is dubious. Besides, Song Zheyuan's sincerity towards a peaceful settlement remains in doubt. Even if Song is sincere, anti-Japanese sentiments are growing strong among his subordinates in the 29th Army. The present situation provides no cause for optimism."¹²⁴ A subsequent meeting between the Army Minister and the Army General Staff produced a new directive on the North China incident. The Army "will continue to maintain its non-expansion policy on the basis of a localized settlement, and endeavor to preventing any action that would lead to a full-scale war."¹²⁵ On the issue of mobilizing troops from Japan, the Army HQ would monitor the developments before making a decision. However, "if the Chinese side disregarded the terms of the earlier settlement and showed no sincerity towards implementing [a settlement], or if Nanjing continued to transfer its troops northwards for offensive purposes, the China

¹²² Diary Entry on the Marco Polo Bridge Incident, Intelligence Bureau of the Japanese Foreign Ministry, HBSJ, 862.

¹²³ Ibid.

¹²⁴ DHR, 316.

¹²⁵ Excerpt of "Guidelines for the Solution of the North China Incident," 13 July 1937, in DHR, 317-8.

Garrison Army should take decisive action,” but subject to the Army HQ’s prior approval.¹²⁶

In the five-minister meeting on 16 July, Army Minister Sugiyama explained that “the negotiations are making no progress because the local authorities in North China lack sincerity”, and proposed a deadline of 19 July for achieving a local settlement.¹²⁷ Foreign Minister Hirota highlighted that “the immediate issue is to get the Central Army to stop its northern advance.”¹²⁸ The meeting set 19 July as the date for troop mobilization. Accordingly, the High Command ordered the field army to terminate local negotiations and commence operations if the demands were not met by the deadline, with the operations limited to North China.¹²⁹ According to the General Staff directive issued on 17 July, the China Garrison Army should begin its operations “based on a confirmation that [the other party] has no sincerity in implementing the terms of settlement” and “exercise superior force against the 29th Army” to achieve “a decisive strike that would cause the Central Army to abandon its determination to join the battle.”¹³⁰ The Army HQ would also choose an appropriate time to deploy troops to Manchuria to defend against a potential Soviet attack.¹³¹

¹²⁶ Ibid.

¹²⁷ DHR, 322.

¹²⁸ Ibid.

¹²⁹ Ibid, 322-3.

¹³⁰ “Guiding Directive for the Use of Military Force in North China” by the First Bureau of the Army General Staff, 17 July 1937, HBSJ, 843.

¹³¹ Ibid, 844.

Then a twist occurred. On 18 July, General Song Zheyuan visited the China Garrison Army HQ and offered his apologies for the incident. The next day, Song provided the details of implementing the terms of settlement as requested by Japan.¹³² In yet another reversal of decisions, Major-General Ishiwara stopped the implementation of the 19 July mobilization order at the Army General Staff.¹³³ The Cabinet meeting on the following day revealed reservations towards a military solution:

Navy Minister Yonai: Non-expansion [of conflict] is our fundamental policy, and a local settlement is being finalized at this moment. If we mobilize our troops, what would happen?

Army Minister Sugiyama: Your opinions are correct, but we have not seen any effort by the Chinese to implement the terms of settlement. Also, the Central Army and other troops are amassing in strength towards the Beijing-Tianjin region. For protection of our overseas citizens and the self-defense of our troops in North China, there is an urgent need for military reinforcements. Furthermore, the mobilization of our troops is already agreed upon in our earlier Cabinet meeting.

Navy Minister Yonai: On the issue of self-defense, I completely understand. If we need to defend ourselves and protect our overseas citizens, shouldn't we also deploy our army to Shanghai as well?

Army Minister Sugiyama: I understand your opinion. The future implications would require further study. On the current problem in North China, the Cabinet had earlier decided to mobilize five divisions. As the situation is urgent, I hope that the Army High Command be allowed discretion to decide when to mobilize the troops. If we miss the timing, the delay would create an irreversible situation.

Prime Minister Konoye: I understand the Army Minister's thoughts. Although it is not very appropriate, I think we should hold off the mobilization of our troops until we know the outcome of the meeting between our ambassador and the Chinese foreign minister.

Army Minister Sugiyama: Yes, let us do that.^{134 135}

¹³² Diary Entry on the Marco Polo Bridge Incident, Intelligence Bureau of the Japanese Foreign Ministry, HBSJ, 867; DHR, 326.

¹³³ DHR, 328.

¹³⁴ Excerpt of Cabinet Minutes, 20 July 1937, DHR, 328.

But on 25 July the crisis rapidly escalated. The igniting spark came when Tokyo was informed that a maintenance detachment of Japan's China Garrison Army had met with a heavy assault by Chinese forces in the Langfang area.¹³⁶ The news pushed Ishiwara, the chief advocate of the non-expansion policy in the General Staff, to jettison his earlier policy. On 26 July, Ishiwara made an urgent call to the Military Section of the General Staff at 1 o'clock in the morning: "There is no choice but to mobilize. A delay would ruin everything. Speed ahead with the mobilization."¹³⁷ Later in the day, it was reported that a Chinese division had attempted to surround and assault Japanese troops at Guangan Gate. Japan's China Garrison Army informed the Army General Staff that in view of the latest atrocities at Langfang and Guangan, it would commence military operations against the Chinese troops within the Beijing-Tianjin area on 27 July.¹³⁸ On 26 July, the Army HQ issued Order No. 418 signed by Chief of Staff Prince Kanin: "In view of the current situation, the Commander of the China Garrison Army must abandon Order No. 400 and apply military force as necessary."¹³⁹ The Cabinet approved the mobilization order on 27

¹³⁵ In Hata's account, the Cabinet met again later in the day and approved Sugiyama's proposal after hearing news that Chinese troops had fired on Japanese forces (Navy General Staff documents related to the Kiri Project). But on July 22, the High Command cancelled the mobilization order for the third time (Diary of Admiral Shimada Shigetaro). This occurred after Colonel Shibayama's return from his North China mission with the report that the Chinese local authorities were implementing the agreement. According to Major Takashima Tatsuhiko from the General Staff War Guidance Section, the China Garrison Army had also recommended to Tokyo on July 22 that the mobilization order be cancelled. Hata, "Marco Polo Bridge Incident," 259.

¹³⁶ DHR, 329-30. The Nanjing archives and the Tokyo archives contradict each other in their accounts of the Langfang and Guangan incidents. See the account based on Chinese archives in the previous section.

¹³⁷ DHR, 330.

¹³⁸ Ibid.

¹³⁹ Order No. 418 by the Japanese Army Chief of Staff, 26 July 1937, HBSJ, 845.

July. The order was implemented in the evening, with 209,000 soldiers mobilized from Japan.¹⁴⁰ With that, Japan entered the eight-year Sino-Japanese War.

Observation 10: External third-party enforcement was neither salient nor credible in the 1937 case. This provides indirect support for Mechanism 1 (exogenous enforcement).

In 1937, no great power or international organization had displayed the willingness and credibility to enforce a peaceful settlement between Nanjing and Tokyo. Nanjing's appeal to the League of Nations and the great powers during the 1931 Manchuria incident was an instructive failure. The League of Nations lacked military power and enforcement capabilities. The non-committal attitude of Britain and the isolationist policy of the United States were apparent.¹⁴¹ While Nanjing sought external diplomatic and military support during the 1937 crisis, I found no evidence that Nanjing believed any great power or international organization would commit itself to enforce a peaceful bargain between China and Japan.¹⁴²

¹⁴⁰ DHR, 331-3; Diary Entries on the Marco Polo Bridge Incident, Intelligence Bureau of the Japanese Foreign Ministry, HBSJ, 846-51.

¹⁴¹ According to Taylor, Chiang appealed to the League of Nations and the Western signatories of the Nine-Power Pact on 12 July. But "as Chiang no doubt fully expected, the United States, United Kingdom, France, and Italy only timidly urged restraint on Tokyo. To avoid irritating Japan, Secretary of State Cordell Hull even stopped a shipment of bombers purchased by the Chinese government. American missionaries promoted an embargo of strategic materials to Japan, but Congress refused to act." Taylor, *The Generalissimo*, 146.

¹⁴² A Nine-Power Conference finally convened in Brussels from 3 to 24 November 1937. The end product from the conference was simply "a resolution encouraging both China and Japan to avail themselves of assistance by other countries to bring the conflict to an early end." John Garver, "Chiang Kai-shek's Quest for Soviet Entry into the Sino-Japanese War," *Political Science Quarterly* 102 (1987), 305.

Anomalies

Reality is often more complex than theory. This sub-section catalogues the theoretical anomalies derived from the 1937 case. There are two important anomalies that surfaced from the historical evidence. The first relates to Mechanism 2 (inadvertent enforcement). The second relates to Mechanism 4 (costly signaling).

Anomaly 1

The evidence shows that the strategic threat presented by the Soviet Union was salient to the Japanese top leadership; yet there was a powerful faction in the army that argued for war with China. Does it undercut Observation 5? Does it contradict Mechanism 2?

To answer these questions, we must examine the argument made by the hawkish faction.¹⁴³ The core of the argument is as follows: A Soviet-Japanese war was impending. The war would create an offensive advantage for China and encourage it to attack Japan from the rear. This would weaken Japan in its fight with the Soviet Union. Hence, as explained by Peattie, “China could and should be knocked out in a swift and decisive campaign (*sokusen-sokketsu*, literally rapid war, rapid settlement) in order to eliminate a threat to the Japanese rear. Then, with China out of the way, Japan could

¹⁴³ Peattie identified members of the hawkish faction in Tokyo to include “almost all” of the Intelligence Division in the Army General Staff, as well as Army Minister Sugiyama, Vice Minister Umezu and Colonel Muto, Ishiwara's deputy in the Operations Division. Peattie, *Ishiwara Kanji*, 291-3.

more easily take on the Soviet Union.”¹⁴⁴ A joint assessment by the Japanese Embassy and the China Garrison Army HQ told the Army General Staff in March 1937: “In harmonizing our relations with China, we must at least achieve Chinese neutrality in our war against the Soviet Union. If this cannot be achieved, we should attack China before the Soviet war begins and undercut the foundation of the Chiang regime.”¹⁴⁵ The Kwantung Army was a strong advocate of the hawkish position. In June 1937, a proposal co-signed by General Tojo Hideki (then at the Kwantung Army) recommended: “Viewing the China situation from the perspective of our military preparations for the Soviet war, we believe that if our military power permits, [Japan should] strike first at the Nanjing regime and remove the threat from our rear. This is the most superior strategy.”¹⁴⁶

As the evidence shows (under Observation 5), this argument did not shape Tokyo’s decisions at the start of the crisis. Tokyo opted for a local settlement to deny a strategic advantage to the Soviet Union. While the idea did exist in the Army High Command, it did not dominate Tokyo’s decisions. Instead, Tokyo was more concerned about its potential vulnerability as Chiang Kaishek’s Central Army advanced towards North China.¹⁴⁷ On balance, the evidence favors Observation 5 and Mechanism 2. But the evidence also highlights the following finding:

¹⁴⁴ Peattie, *Ishiwara Kanji*, 286.

¹⁴⁵ DHR, 293.

¹⁴⁶ Proposal on the War Strategy against Soviet Union and China by the Kwantung Army, 9 June 1937, RDZDQ, 231.

¹⁴⁷ See the analysis organized under Observation 9.

Observation 11: A hawkish faction believed that a future Soviet-Japanese war would create an offensive opportunity for China to attack Japan; thus the faction argued that Japan should neutralize China before fighting the Soviet Union. This suggests a conditional refinement to Mechanism 2 (inadvertent enforcement).

This is an existence claim: that such a faction *did* exist. It is not a causal claim; it does not claim that the existence of this faction caused Tokyo to decide for war. But Observation 11 is informative in two ways. First, it provides indirect support for Theory 1, which argues that war is more likely when there exists a potential vulnerability that an opponent cannot commit not to exploit. Upon inspection, the strategic logic in Observation 11 is reducible to the commitment problem highlighted in Theory 1. Second, while the causal-claim version of Observation 11 was not realized in this empirical case, it is theoretically realizable in other (or future) cases. The “early-neutralization” argument in Observation 11 has a clear strategic logic. The logic motivated the pro-war recommendations made by the Kwantung Army. While the recommendations were not taken by Tokyo, the neutralization logic had touched on strategic calculations for war and peace in a real way.

The neutralization logic assumes that the costs of neutralizing the minor enemy are sufficiently low, such that the neutralization effort (i.e. war) would not create a significant strategic advantage to the major enemy. Otherwise, the effects of inadvertent enforcement (Mechanism 2) will kick in and dampen the incentive for war. Hence, in the 1937 case, advocates of the neutralization argument emphasized the ease of *sokusen-*

sokketsu (rapid war, rapid settlement).¹⁴⁸ Those who opposed the argument raised their objections against the *sokusen-sokketsu* assumption. Dissenters felt that a China war would be costly and protracted. The disagreement over the *sokusen-sokketsu* assumption reduced into a disagreement over China's resolve and strength: Was China willing and able to resist Japan by fighting a long and costly war?

Dissenters to the *sokusen-sokketsu* assumption took China's signals of resolve – its verbal threats and military deployments – seriously. The leading dissenter, Major-General Ishiwara, argued that Nanjing's determination to resist "cannot be underestimated" and that if war occurred, "it cannot be controlled."¹⁴⁹ Hence, Ishiwara pushed hard for a non-expansion policy. Advocates of *sokusen-sokketsu* underestimated China's resolve. They argued that China was unable and unwilling to fight a long and serious war with the Japanese army.¹⁵⁰ Japan's military successes in China since the 1931 war in Manchuria, as well as China's acquiescence in many of those conflicts, provided the evidence for their assessment. In Ishiwara's view, his hawkish colleagues opted for war because they thought "the China incident would be easily settled [by military force] just like the

¹⁴⁸ According to Boyle, the Chief of the General Staff China Section had argued: "We need only dispatch more troops and take [Beijing], and the rest of the country will be on its knees." The Chief of the General Staff Intelligence Division argued that Chinese resistance would collapse with a single decisive strike. Army Minister Sugiyama told the Emperor that "the China Incident will all be over in a month." Boyle, *China and Japan at War*, 50-3.

¹⁴⁹ Record of Interview between Prince Takeda and Ishiwara Kanji, 1939, DHR, 334. See also the excerpt of the minutes of meeting between Ishiwara and Army Minister Sugiyama on July 18, in DHR, 326-7.

¹⁵⁰ According to Hata, the China Section and War Guidance Section of the Army General Staff "tended to minimize China's war-making capabilities and concluded optimistically that it could be completely subdued by a small military force deployed for a short period of time." On the other hand, dissenters led by Ishiwara argued that a conflict in China would involve "(a) the simultaneous mobilization of fifteen divisions, (b) immediate mobilization of at least half of the resources needed for war, (c) extension of the war to the entire region north of the Yellow River and even to Shanghai, (d) military operations lasting for at least six months, and (e) a war expenditure of 5.5 billion yen." Hata, "Marco Polo Bridge Incident," 251.

Manchuria incident [of 1931].”¹⁵¹ Hence, they opposed the non-expansion policy and pushed for a quick war in China.¹⁵² On the whole, those who found *sokusen-sokketsu* attainable argued for war. For those who didn’t, they argued against war.

Anomaly 2

Observation 12: Costly signaling of resolve from each side made both sides insecure and promoted hawkish counter-responses. This finding contradicts Mechanism 4 (costly signaling).

Costly signaling of resolve did not seem to encourage either side to back down. In fact, the 1937 case suggests that as one side steadily increased its costs and risks of war with military deployments, the other side pushed harder with their own deployments. This anomaly suggests an important tradeoff involved in the costly signaling of resolve:

Observation 13: When signals of resolve directly increase the vulnerability of the opponent, they deepen the commitment problem and make war more likely. This suggests a realistic limitation to Mechanism 4 (costly signaling).

¹⁵¹ Record of Interview between Prince Takeda and Ishiwara Kanji, 1939, DHR, 334. See also Ishiwara’s remarks in DHR, 335.

¹⁵² Indeed, at the eruption of the Marco Polo Bridge incident, the Kwantung Army informed Tokyo that it had prepared its forces and were ready “to set off immediately.” However, Tokyo sent explicit orders not to expand the conflict. DHR, 299; Order No. 400 of the Japanese Army Chief of Staff, 8 July 1937, HBSJ, 831; DHR, 305.

Revisiting the decision-making processes described under Observations 7 and 8, it is clear that while military deployments acted as signals of resolve, they also acted to increase the perceived vulnerability of the other side. They deepened the commitment problem and made war more likely by increasing the opponent's potential vulnerability (Theory 1). The problem might be moderated if defensive deployments were differentiable from offensive deployments.¹⁵³ This did not apply in the 1937 case. Policymakers on each side reacted by treating the opponent's deployments as potentially offensive. Theoretically, this finding suggests an important but under-explored connection between costly signaling theory and Robert Jervis's spiral model. Practically, the finding suggests that when signals of resolve increase defensive vulnerability, it can be extremely dangerous to tune up the resolve level in the signals sent.

The processes leading to the Third Sino-Japanese War in 1937 are a striking contrast to those leading to the Second Sino-Japanese War in 1931. But the two wars are intimately connected. Let us now retrace our steps to September 1931 when a group of conspirators in the Japanese army engineered a crisis at Mukden (Shenyang).

¹⁵³ Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics* 30 (1978): 167-214.

SECOND SINO-JAPANESE WAR (1931)

The Second Sino-Japanese War started with the Mukden crisis on 18 September 1931. While the war caused fewer casualties than the Third Sino-Japanese War, it is no less significant historically. After the Mukden crisis, Japan turned into the road that led to the Third Sino-Japanese War and the Second World War; Japanese militarism dominated politics and ended an era of party government; and the League of Nations crumbled as a credible institution for world peace.¹⁵⁴ The Second Sino-Japanese War sowed the seeds for the Third Sino-Japanese War, the Pacific War, and the eventual collapse of Imperial Japan.

The 1931 Mukden crisis is a focal event from which we derive many of our generalizations and stereotypes about Japanese militarism.¹⁵⁵ Planned by Lieutenant-Colonel Ishiwara Kanji and Colonel Itagaki Seishiro at the Kwantung Army, the crisis was designed to provide a pretext for invading Manchuria. The existing literature emphasizes a defiant Kwantung Army taking the lead to secure Manchuria under Japanese control. Conspirators at the Kwantung Army believed that Manchuria would

¹⁵⁴ Akira Iriye, "Introduction: The Extension of Hostilities, 1931-1932," in *Japan Erupts: The London Naval Conference and the Manchurian Incident, 1928-1932*, ed. James Morley (New York: Columbia University Press, 1984), 233. The significance of the Mukden crisis in 1931 is almost never in doubt among scholars of Japanese history. Marius Jansen argued, for example, that the consequences were "far-reaching in all respects": International opprobrium united domestic public opinion in Japan; military insubordination had demonstrated its success; the military became much stronger and ready for future adventures; and the link between domestic reform and military expansionism was forged. See Marius Jansen, *Japan and China: From War to Peace, 1894-1972* (Chicago: Rand McNally, 1987), 384-5.

¹⁵⁵ Iriye pointed out that "[t]he place of the military in Japanese decision making, the role of junior officers in the military hierarchy, the independence of the field army, army-navy rivalry, civilian ultranationalists – all these concepts have been examined with reference to the Mukden Incident It is no exaggeration to say that much generalization on prewar Japanese history has been based on the study of the Manchurian crisis." Iriye, "Introduction: The Extension of Hostilities," 233.

provide a solution for Japan's deteriorating international position and its domestic economic and political crises.¹⁵⁶ Leaders in Tokyo turned out to be either unwilling or unable (or both) to stop the Kwantung Army. The civilian government was politically weak and could not stand up to the military.

Given the overwhelming numerical superiority of Chinese troops in Manchuria, it was arguably China's failure to stand up against the Kwantung Army that allowed the Mukden crisis to expand so rapidly into a large-scale invasion. But while there is a rich literature on the Japanese role in the Mukden crisis, there are few archival studies of Chinese reactions in the English language. The most relevant works are two studies by Park Cobles and Sun Youli on Chinese foreign policy in the 1930s. Both works connected Nanjing's non-resistance policy to Chiang Kaishek's delicate position vis-à-vis his domestic enemies that included the communist insurgents in Jiangxi and the rebel regime in Guangdong.¹⁵⁷

¹⁵⁶ See, e.g., Ogata, *Making of Japanese Foreign Policy*; Akira Iriye, *After Imperialism: The Search for a New Order in the Far East, 1921-1931* (Cambridge, MA: Harvard University Press, 1965); Crowley, *Japan's Quest for Autonomy*; Peattie, *Ishiwara Kanji*; Seki Hiroharu, "The Manchurian Incident, 1931," in *Japan Erupts: The London Naval Conference and the Manchurian Incident, 1928-1932*, ed. James Morley (New York: Columbia University Press, 1984), 139-230; Shimida Toshihiko, "The Extension of Hostilities, 1931-1932," in *Japan Erupts: The London Naval Conference and the Manchurian Incident, 1928-1932*, ed. James Morley (New York: Columbia University Press, 1984), 241-335; Barnhart, *Japan Prepares For Total War*; Tobe Ryōichi, "The Manchurian Incident to the Second Sino-Japanese War," in *Japan-China Joint History Research Report: Modern and Contemporary History, Vol. 1*, ed. China Joint History Research Committee (Governments of Japan and the People's Republic of China, 2011), 95-127. In the English-language historiography, Ogata and Seki provided the most detailed studies on the Japanese role in the genesis of the Mukden crisis. Shimida described in close detail how the hostilities extended across North China. Peattie provided a valuable study of Ishiwara and his role in the Second and Third Sino-Japanese wars. Barnhart showed how Manchuria dominated the strategic thinking of various Japanese leaders.

¹⁵⁷ Chiang Kaishek might also have been influenced by the belief that Tokyo would successfully rein in the Kwantung Army, or the belief that the League of Nations and other great powers would intervene on China's behalf. Coble, *Facing Japan*, 18.

While this generalization is accurate, neither of these studies had focused on the Mukden crisis. In this section, I analyze the processes leading to the Second Sino-Japanese War, focusing specifically on Japanese and Chinese calculations in the Mukden crisis based on archival records from both sides of the conflict.

Background

The leased Kwantung Territory in Manchuria was served by the South Manchuria Railway controlled by the Japanese government. Japan held extraterritorial rights of taxation and security in the railway areas, which included parts of major cities including Mukden (Shenyang). Japan's Kwantung Army was responsible for the security of these areas.¹⁵⁸ In September 1931, General Honjo Shigeru commanded the Kwantung Army, with Colonel Itagaki Seishiro and Lieutenant-Colonel Ishiwara Kanji on his general staff. Prime Minister Wakatsuki Reijiro headed the civilian government in Tokyo with Shidehara Kijuro as his Foreign Minister. Meanwhile, Zhang Xueliang dominated Northeast China, including Manchuria, as the regional warlord and Vice Commander-in-Chief of the armed forces under the Nanjing government headed by Chiang Kaishek.¹⁵⁹ Zhang had about 20,000 troops around Mukden and a total of about 250,000 troops in the province; the Kwantung Army had only 10,000 troops.¹⁶⁰

¹⁵⁸ Jansen, *Japan and China: From War to Peace*, 372-3. For more background about the origins of the Kwantung Army, see Alvin Coox, *Nomonhan: Japan against Russia, 1939* (Stanford, CA: Stanford University Press, 1985), chapters 1-2.

¹⁵⁹ Coble, *Facing Japan*, 11-16.

¹⁶⁰ Seki, "Manchurian Incident," 144; Peattie, *Ishiwara Kanji*, 106; Coble, *Facing Japan*, 12.

In the summer of 1931, tensions were high between China and Japan. The Wanbaoshan incident – a clash between Chinese and Korean farmers over irrigation ditches – had triggered anti-Chinese riots in Korea as well as anti-Japanese demonstrations in China.¹⁶¹ Nanjing and Tokyo were in the midst of a diplomatic crisis over the Nakamura incident, in which Manchurian soldiers killed a Japanese army captain on a reconnaissance mission. The incident inflamed public sentiments in both China and Japan. While Nanjing and Tokyo were still figuring out a settlement to the incident, the Mukden crisis erupted.

At around 10 pm on 18 September, conspirators at the Kwantung Army blasted a section of the South Manchuria Railway tracks at Liutiaokou near Mukden. Immediately thereafter, troops from the Kwantung Army attacked Chinese defense installations around Mukden. Commander Honjo authorized a general assault on Chinese forces shortly after midnight.¹⁶² The Mukden crisis had begun. The success of the Ishiwara-Itagaki conspiracy depended on whether the operations in Mukden would expand into a full invasion of Manchuria.¹⁶³ That would depend, in turn, on Chinese reactions and Tokyo's acquiescence.

¹⁶¹ See Liu Jie, "Sino-Japanese Diplomacy during Cycles of Mutual Antagonism: On the Eve of the Manchurian Incident," in *Toward a History Beyond Borders: Contentious Issues in Sino-Japanese Relations*, ed. Yang Daqing, Liu Jie, Hiroshi Mitani and Andrew Gordon (Cambridge, MA: Harvard University Asia Center, 2012), 81-120.

¹⁶² For a detailed narrative of the Mukden incident, see Seki, "Manchurian Incident," 225-30.

¹⁶³ Jansen, *Japan and China: From War to Peace*, 381.

4. Theoretical Predictions

My outcome variable of interest consists of the decisions and proposals for and against escalation or war. These are found in the arguments made by the key players to trigger, expand, or avoid military conflict. The key players included, on the Japanese side, the decision-makers in Tokyo and the conspirators in the Kwantung Army led by Ishiwara and Itagaki; and, on the Chinese side, Chiang Kaishek and Zhang Xueliang. My focus is not merely on *why* the Mukden crisis occurred, but also on *how* it expanded into a large-scale invasion. Chinese reactions in the Mukden crisis are crucial for understanding the latter – but they are often bracketed or treated in general terms in existing studies on the causes of the Second Sino-Japanese War.

Do the theoretical predictions in Chapter 1 fit the decision processes leading to the Second Sino-Japanese war? The results are summarized in Table 3.2:

Table 3.2: Predictions vs. Evidence from the Second Sino-Japanese War

Predictions	Empirical Support
T1 (Commitment Theory) Prediction T1a ** Prediction T1b	Direct Support: Ishiwara argued that Japan's position in Manchuria would be highly vulnerable in the future, given that China would become strong once it achieved national unification, and that China's nationalist movement to recover its sovereign rights was inevitable.
M1 (Exogenous Enforcement) Prediction M1a ** Prediction M1b **	Direct Support: Nanjing entrusted the League of Nations to enforce the status quo, and assumed that a non-military position would facilitate international intervention in resolving the crisis. Conspirators at the Kwantung Army, however, believed that the League was ineffectual, given that the U.S., Britain and France were trapped in economic crisis and had competing interests in East Asia.
M2 (Inadvertent Enforcement) Prediction M2a ** Prediction M2b **	Direct Support: Chiang's impulses towards war were checked by his acute sense of domestic vulnerability, given China's internal turmoil and his domestic rivals at the rear. Tokyo's strategic fear of Soviet intervention also motivated its more prudent approach to the crisis compared to the conspirators at the Kwantung Army, who believed that the Soviets faced "major constraints in the use of military force in the Far East."
M3 (Endogenous Enforcement) (N.A.)	
M4 (Costly Signaling) Prediction M4a * Prediction M4b	Indirect Support: Instead of sending costly signals to deter the Japanese, Nanjing's policy of non-resistance kept its military costs to the minimum. The Kwantung Army escalated the crisis into a large-scale invasion, despite its numerical inferiority vis-à-vis the Chinese forces.
M5 (Costly Implementation) Prediction M5a * Prediction M5b	Indirect Support: Ishiwara expected that China with its internal turmoil would not be able to react strongly to the Japanese invasion of Manchuria.
M6 (Contradictory Signaling) Prediction M6a * Prediction M6b	Indirect Support: Nanjing's verbal insistence on upholding its territorial integrity was contradicted by its policy of military non-resistance.

Notes: ** Direct support for the prediction. * Indirect support for the prediction. # Prediction contradicted.

The 1931 case provides direct support for the commitment theory of war and the mechanisms of exogenous and inadvertent enforcement. Support for the signaling mechanisms, however, is indirect. This is because we have little evidence on the credibility assessments of specific signals sent during the crisis. A lack of resolve was reflected by the Chinese avoidance of military costs during the crisis (Mechanism M4) and the contradiction between Chinese non-resistance on the ground and its expressed determination to maintain its territorial integrity (Mechanism M6). But we do not know how these information had influenced the expectations and decisions on the Japanese side. Of course, such evidence can be hard to find in the first place, since the Kwantung Army operated in a conspiratorial setting. Nonetheless, there is only indirect evidence for the signaling mechanisms.

The next section provides a detailed reconstruction of the calculations behind the Mukden crisis and the spread of hostilities. The timeframe is 18 September to the end of September 1931: this was the crucial period when the crisis expanded into a large-scale invasion of Manchuria. I focus on the deliberations and decisions made by three sets of key players. I begin with (1) *Chinese reactions* to the Mukden incident, which were largely determined by Chiang Kaishek and Zhang Xueliang. Next I turn to (2) *Tokyo's reactions* to the Mukden incident and its deliberations on whether to support or suppress the Kwantung Army's operations. Finally, I examine (3) the calculations and arguments that had motivated *the conspirators in the Kwantung Army*. Each part is a descriptive reconstruction of the decision processes based on what is found in the archival records. As before, the headings (under "Observation") indicate where the evidence fits with or contradicts the theoretical predictions.

Chinese Reactions to the Crisis

Observation 14: Chiang Kaishek's impulses towards war were checked by his acute sense of domestic vulnerability, given China's internal turmoil and his domestic rivals at the rear. This finding supports Mechanism 2 (inadvertent enforcement).

Observation 15: Nanjing entrusted the League of Nations and assumed that a non-military position would facilitate international intervention in resolving the crisis. This finding supports Mechanism 1 (exogenous enforcement).

Even before 18 September, Zhang Xueliang was careful to avoid military conflict with the Japanese. By August 1931, Zhang had understood from intelligence sources that a pro-war mood was gathering steam in Japan.¹⁶⁴ The situation was becoming delicate and Zhang was keen to avoid a clash. On 6 September, he instructed Zang Shiyi (Liaoning Provincial Government Council Chairman) and Rong Zhen (Northeast Border Defense Army Chief of Staff): "It is desirable that we do our best to maintain composure and circumspection as we deal with all matters [arising from the increasingly sensitive situation with Japan]. Regardless of how the Japanese seek trouble, our side should exercise utmost forbearance so as to prevent clashing with [the Japanese] and leading to complications. This secret order should be immediately transmitted to the relevant

¹⁶⁴ ZXNP, 400-1. On 8 September, Zhang received intelligence that Japan's Korea Army had been conducting bridging exercises in its waters. Zhang replied on 10 September that he was "already apprised of the information." "8 September 1931," ZXNP, 402.

subordinates and be given serious emphasis.”¹⁶⁵ On the same day, he instructed Brigade Commander Wang Yizhe in the field: “The Sino-Japanese situation is currently very serious. Our troops should exercise special caution with regard to Japanese troops. Regardless of any kind of provocation [from the Japanese], [our troops] should exercise forbearance. Conflict is prohibited so as to prevent incidents.”¹⁶⁶

Chiang held a similar view.¹⁶⁷ Chiang’s private diary shows that while he recognized the growing threat from Japan, he was beset by urgent domestic problems that he had to resolve first. Chiang wrote: “The Japanese initially retreated but later advanced [their position] in the Wanbaoshan incident; [they] publicized the Nakamura incident in a massive way. Their invasion tendencies have become more threatening by the day. Alas, our nation has yet to pacify its internal [disturbances]; and natural calamities and floods have caused the people to suffer. In this situation, what should I do? ... If our China can quickly wake up to the need for self-strengthening through unification, what is there [about the Japanese] to fear?”¹⁶⁸ On 16 September, Chiang wrote that he was trapped in a

¹⁶⁵ Telegram from Zhang Xueliang to Zang Shiyi and Rong Zhen, 6 September 1931, KZ, Vol. 1, 150.

¹⁶⁶ “6 September 1931,” ZXNP, 402.

¹⁶⁷ Some historians have claimed that Chiang had ordered Zhang to adopt a non-resistance policy in a secret telegram on 16 August 1931: “From now on, regardless of how Japanese troops attempt to provoke incidents in Northeast [China], our side should respond with non-resistance and try our best to avoid a clash. [You] should definitely not react on the spur of the moment in disregard of our country and people.” Attributed Telegram from Chiang Kaishek to Zhang Xueliang, 16 August 1931, KZ, Vol. 1, 150. However, the authenticity of this source remains controversial. The actual telegram has never been found in any existing archive. See Yang Tianshi, *Xunzhao Zhenshi de Jiang Jieshi: Jiang Jieshi Reji Jiedu II* [In Search of the Real Chiang Kaishek: Deciphering the Diaries of Chiang Kaishek II] (Beijing: Huawen Chubanshe, 2010), 57. See also the summary of meeting between Chiang and Zhang on 12 September 1931 in which Chiang had supposedly told Zhang to adopt a non-resistance stance. This is reported in the chronological records of Chiang produced by the PRC Central Party History Committee (“12 September 1931,” JJNP, 195). However, the same claim does not appear in the chronological records by the Kuomintang Party History Committee in Taiwan.

¹⁶⁸ Chiang Kaishek’s Diary, 1 September 1931, ZJDC, Vol. 2, 380-1. On 23 July 1931, Chiang proclaimed that “internal pacification is necessary before external resistance” and that “if we do not

very dangerous situation: “The red bandits [communists], Guangdong rebels and Japanese bandits are coming in mutual succession. This is indeed the most treacherous time for [our] Party and Nation.”¹⁶⁹

The Mukden crisis broke out two days later in the night of 18 September. Zhang, who was in Beijing, ordered his deputies to ensure that their subordinates “strictly maintain their calm towards this incident to prevent undesirable complications to the matter. [This is] of utmost importance!”¹⁷⁰ The prohibition on the use of force was duly transmitted to Brigade Commander Wang Yizhe by Chief of Staff Rong Zhen.¹⁷¹ Wang’s troops reportedly “thundered with anger” and begged their commander for permission to fight, but the non-resistance order was enforced, despite many “clutching their rifles in tears and pounding the wall with their fists.”¹⁷² The order was firmly maintained in the days that followed. Zhang ordered his troops to retreat into safety and avoid clashing with the enemy.¹⁷³ Local officials were ordered to “calmly await the resolution of the matter” and to “continue to protect overseas Japanese and other foreign citizens in accordance to our

eliminate the red bandits [the communists], we cannot protect ourselves from external humiliation. If we cannot achieve [national] unification, we cannot achieve external resistance.” Proclamation to the Nation on Internal Pacification and External Resistance, 23 July 1931, JJNP, 149.

¹⁶⁹ “16 September 1931,” ZJDC, Vol. 2, 383.

¹⁷⁰ Telegram from Zhang Xueliang to Zhang Zuoxiang et al., 19 September 1931, JYBD, 255.

¹⁷¹ “I reported [the Japanese attack] to Vice-Commander Zhang [Xueliang] in Beijing over the telephone and requested for his instructions. [Zhang instructed that we would] respect the peaceful principles of the League of Nations and avoid a clash. Therefore [I] transmitted the instruction to Seventh Brigade Commander Wang Yizhe, ordering him against [military] resistance.” Incident Report by Chief of Staff Rong Zhen, KZ, Vol. 1, 153.

¹⁷² Incident Report by Chief of Staff Rong Zhen, KZ, Vol. 1, 155-6. The report corroborated with Brigade Commander Wang Yizhe’s comments to reporters in October 1931. “7 October 1931,” ZXNP, 408.

¹⁷³ “22 September 1931,” ZXNP, 405.

obligations, and ... not discriminate against them.”¹⁷⁴ At this point, Zhang did not expect the Mukden incident to expand into a large-scale invasion.¹⁷⁵

Chiang was away from Nanjing when he learned of the incident in the night of 19 September.¹⁷⁶ Chiang expressed helplessness in his diary that evening: “The [Japanese] bandits are indeed exploiting the internal fragmentation created by the Guangdong rebels to invade our eastern provinces. Alas! Our internal turmoil is unending and the rebels show no signs of repentance Natural calamities and lawlessness are prevalent. Our national strength is at its lowest ebb. [I] want to stand firm and protect [China] from humiliation, but [our national] strength does not suffice!”¹⁷⁷ On 21 September, Chiang arrived back at Nanjing and called a meeting to discuss the Japanese aggression. Chiang proposed to take the matter to the League of Nations.¹⁷⁸ The meeting agreed to suspend the military campaigns against the communists and the Guangdong government; to divert troops to assist in the defense of North China; to persuade the Guangdong

¹⁷⁴ Rehe Provincial Government Order to Local Officials, 21 September 1931, JYBD, 256.

¹⁷⁵ His subsequent report to Nanjing noted: “the Japanese maneuver was perceivably nothing beyond their usual acts of provocation; and to prevent the incident from escalating, [we] have absolutely adhered to the principle of non-resistance. [The wanton aggression of the Japanese forces subsequently] gave our troops no choice but to retreat from their barracks.” Telegram from Zhang Xueliang to Chiang Kaishek and Wang Zhengting, 24 September 1931, KZ, Vol. 1, 160. Zhang might have perceived the Mukden incident as bait for China to provide a pretext for Japanese militarists. Zhang explained in an oral interview in 1990: “I cannot shift the responsibility [for non-resistance] to the Central [government]. I personally did not want to escalate the incident. I did not realize that Japan would come [into China] as a whole [in a major way]. I judged that [the Japanese] wanted to provoke [us]. [That is] to find an incident to provoke China, and then provide an excuse for the military. At that time China had many problems, such as the Nanjing incident and the Jinan incident – they all turned from big incidents into small incidents, [and from] small incidents into nothing. If I had known that this incident could not be resolved so easily, then [things] would have been different.” NHK Interviews with Zhang Xueliang, 1990, KZ, Vol. 1, 163.

¹⁷⁶ Yang Tianshi, *Xunzhao Zhenshi de Jiang Jieshi*, 45-7.

¹⁷⁷ “19 September 1931,” ZJDC, Vol. 2, 386-7.

¹⁷⁸ KZ, Vol. 1, 158; “21 September 1931,” ZJDC, Vol. 2, 387. On appealing to the League of Nations, see also ZMSD, Vol. 5, Part 1, 390.

government to unite with Nanjing; and to ask the public to “maintain their calm and forbearance, unite as one, prepare for self-defense, and entrust the League of Nations to resolve the crisis with justice.”¹⁷⁹

To Chiang, the best response to the crisis was to rely on international intervention and domestic unification. Speaking to the party on 22 September, Chiang said: “Now that international opinion has reached a consensus on the unreasonable nature of Japanese [actions], our people must share the same stand [to] meet power with reason, barbarism with pacifism; [to] exercise forbearance despite the pain and anger; [to] temporarily adopt a passive attitude while awaiting the judgment of international justice.”¹⁸⁰ But if all else fails and “peace becomes impossible,” the Central Government has the “final resolve and preparation ... to defend the survival of the people and the integrity of the nation.”¹⁸¹ At this point, the League of Nations did not disappoint. It passed a resolution calling for the withdrawal of Japanese troops and the immediate termination of all actions that might lead to the expansion of hostilities.¹⁸²

¹⁷⁹ KZ, Vol. 1, 158; “16 September 1931,” ZJDC, Vol. 2, 387. On the same evening, Chiang met Song Ziwen and a League of Nations representative. In response to the representative’s optimism that the crisis would be quickly diffused, Chiang replied: “You should understand that since the Japanese troops have occupied Shenyang, [they] would definitely not withdraw easily. I know the Japanese mentality towards Northeast [China]. [They] would rather cause the total destruction of Tokyo and the three main islands of Japan than to withdraw from Northeast [China] voluntarily. I hope you remember firmly what I have said and not view the situation lightly.” “16 September 1931,” ZJDC, Vol. 2, 387. On 22 September, Chiang observed: “The actions of the Japanese invaders have just marked the start of the Second World War. Not sure if the people of different countries can see this.” “22 September 1931,” ZJDC, Vol. 2, 390.

¹⁸⁰ “22 September 1931,” ZJDC, Vol. 2, 389.

¹⁸¹ *Ibid.*

¹⁸² *Ibid.*, 390.

On 23 September, Zhang Xueliang's deputy, Wan Fulin, arrived at Nanjing. Chiang told Wan that "the diplomatic situation suggests that there remains justice [in the international sphere]. The territorial integrity of the eastern provinces must be kept intact. [Zhang] should definitely not engage in bilateral negotiations [with the Japanese] and sign unequal treaties that humiliate Chinese sovereignty."¹⁸³ In Chiang's view, "the [Japanese] thieves are cunning and unreasonable. It will not be easy to resolve the matter quickly. It is better to entrust the matter to the collective judgment of the League of Nations as that may provide some hope of recovering [the lost territories]. Otherwise [I] will rather fight a war with the [Japanese] bandits, [and attain] glory despite defeat."¹⁸⁴

The Nanjing government's national proclamation on 24 September emphasized military non-resistance to facilitate international intervention in resolving the crisis: "The Government has now presented the case to the League of Nations, pending a just resolution [to the incident], and therefore [the Government] has strictly ordered all troops in the nation to avoid confrontations with Japanese troops. The same warning applies to all civilians, who must maintain an attitude of seriousness, composure and calm."¹⁸⁵ The proclamation highlighted Chiang's hope for domestic unification in the face of crisis: "All conflicts within the nation should now be resolved. All our people should abandon their personal prejudices and unite as one under the Nationalist Government, in pursuit of the security and independence of our nation. All citizens should recognize that unless they support the unification of the nation, there is nothing [we can do] in dealing with our

¹⁸³ "23 September 1931," ZJDC, Vol. 2, 390.

¹⁸⁴ *Ibid.*, 390-1.

¹⁸⁵ Proclamation to the People by the Nanjing Government, 23 September 1931, KZ, Vol. 1, 159.

external [troubles].”¹⁸⁶ Chiang sent his envoys to Hong Kong to persuade the Guangdong government to reunite with Nanjing.¹⁸⁷ The Foreign Ministry also lobbied influential groups in Shanghai to exert pressure on the League of Nations.¹⁸⁸

Meanwhile, Zhang was informed by Japanese diplomatic sources that the Japanese army would cease its advances, and that the troops heading to Jilin would retreat in a matter of days. Zhang noted that while the information was hard to believe fully, its tone of delivery sounded sincere.¹⁸⁹ On 25 September, Zhang told representatives from anti-Japanese resistance groups that he would never sell out China, but the non-resistance policy remained the best option in the current situation.¹⁹⁰

However, Japan rejected international intervention by the League of Nations and called for a bilateral settlement with the Chinese, which was in turn rejected by Nanjing. Chiang’s diary entries explained his thinking: “In the Nakamura incident, Japan did not seek a bilateral settlement with us. Now it has invaded the eastern provinces and refuses international intervention. The cunning, divisive tactics of the Japanese are extremely lethal.”¹⁹¹ “If [we conduct] direct or localized negotiations [with the Japanese], the

¹⁸⁶ Ibid.

¹⁸⁷ “24 September 1931,” ZJDC, Vol. 2, 391.

¹⁸⁸ Telegram from the Ministry of Foreign Affairs to the Shanghai Administrative Office, 24 September 1931, ZMSD, Vol. 5, Part 1, 535.

¹⁸⁹ “25 September 1931,” ZXNP, 406.

¹⁹⁰ Ibid.

¹⁹¹ “25 September 1931,” ZJDC, Vol. 2, 391-2.

outcome would definitely be unfavorable [to us]. Rather than to perish without a fight, why not fight and perish, so as to preserve the integrity of the Chinese people?”¹⁹²

At this point, Chiang faced strong domestic pressure from public opinion and his political rivals at Guangdong. The public was clamoring for military resistance. On 28 September, Beijing saw a 200,000 strong pro-resistance demonstration.¹⁹³ On the same day, furious students burst into Foreign Minister Wang Zhengting’s office and beat him up.¹⁹⁴ The Guangdong government was also capitalizing on the crisis.¹⁹⁵ Negotiations between the Nanjing and Guangdong governments were obstructed by the latter’s insistence on Chiang’s resignation and the reorganization of the national government. Chiang was frustrated at how the Guangdong government was creating trouble without paying heed to the larger picture, yet “now I can only take things in my stride and tolerate the humiliation to fulfill the larger purpose, so that there is some small hope of salvaging [the situation].”¹⁹⁶ “The invasion and threat from Japanese bandits are urgently increasing step by step; can the Guangdong side wake up?”¹⁹⁷

Chiang’s diary entries in October 1931 further reveal his vacillation between thoughts of war and hopes for resolution through the League of Nations, with an inverse relationship

¹⁹² Entry in Chiang Kaishek Diary, 26 September 1931, in Yang Tianshi, *Xunzhao Zhenshi de Jiang Jieshi: Jiang Jieshi Reji Jiedu I* [In Search of the Real Chiang Kaishek: Deciphering the Diaries of Chiang Kaishek I] (Taiyuan: Shanxi Renmin Chubanshe, 2008), 201.

¹⁹³ “28 September 1931,” ZXNP, 407.

¹⁹⁴ “28 September 1931,” ZJDC, Vol. 2, 392.

¹⁹⁵ At the start of the crisis, the Guangdong regime proposed three measures to resolve the crisis: (1) Chiang’s resignation from office; (2) the dissolving of the Guangdong government; and (3) the creation of a unified Chinese government after a conference on national unification. “21 September 1931,” JJNP, 195.

¹⁹⁶ “30 September 1931,” ZJDC, Vol. 2, 393.

¹⁹⁷ “2 October 1931,” ZJDC, Vol. 2, 394.

between the two.¹⁹⁸ But the idea of war, however intense, was always checked by Chiang's realization of vulnerability, given domestic turmoil and his rivals at the rear.

Tokyo's Reactions to the Crisis

How did the Japanese government respond to the crisis? Here I focus on the sequence of moves and decisions made by Tokyo in the first week of the crisis, which was when the key developments took place. There was as a tug of war between three distinct groups: between the civilian government and the Army HQ in Tokyo, and between the Army HQ and its army in the field.

19 September 1931: Tug of war between Cabinet, Army HQ and Field Army

Tokyo received a report of the incident in the early morning of 19 September, which stated: "At about 10 pm on the 18th [September], to the west of our northern barracks at Fengtian, there was a violent commotion in which [anti-Japanese] troops damaged the Northern Manchurian railroad, attacked our defense guards and clashed with

¹⁹⁸ See ZJDC, Vol. 2, 394-402. Chiang's personal vacillation between his impulses for war and his realization of domestic vulnerability continued throughout the end of 1931. For example, on 27 November, Chiang reportedly told a group of students that he would declare a war of resistance in three days time. But on 30 November, at the appointment ceremony of Foreign Minister Gu Weijun, he announced in his speech that "the value of invisible diplomatic battles goes beyond any type of battle" and emphasized that "to deal with our external [threat] we must first stabilize our internal [domestic situation]; unification [is necessary] before we can protect [ourselves] against [external] humiliation. There has never been a country that is unable to unify itself and yet win a victory over its external [enemy]." "27 November 1931" and "30 November 1931," JJNP, 199.

reinforcements from our defending units.”¹⁹⁹ A later report from the Kwantung Army stated: “Our Army has mobilized our main forces to clear out Chinese troops along the Manchurian rail lines.”²⁰⁰

At 7 am, an urgent meeting was held at the Army General Staff. Vice-Chief of Staff Ninomiya Harushige described the reports received from the field and highlighted that “the purpose of this meeting is to lay a firm foundation for determined [action] within the Army General Staff.” Military Affairs Bureau Chief Koiso Kuniaki emphasized that “[t]he current actions by the Kwantung Army are appropriate reactions to the situation that are completely in accordance with its [field] mission.” The meeting converged with the view and unanimously agreed to send reinforcements to the Kwantung Army. Preparations were made to table the proposal for Cabinet approval.²⁰¹ Around 8 am, the General Staff received a telegram from the Korea Army Commander Hayashi Senjuro. Hayashi informed that he had activated aerial and reconnaissance units to reinforce the Kwantung Army. In addition, a mixed brigade from the 20th Division was preparing to move out. Hayashi had also ordered the 19th Division to prepare for mobilization.²⁰²

At 10 am, an emergency Cabinet meeting was held. Just before the meeting, Prime Minister Wakatsuki Reijiro asked Army Minister Minami Jiro, “Regarding this action by

¹⁹⁹ Report from Kwantung Army HQ to Vice-Chief of General Staff (Tokyo), 19 September 1931, DHR, 192.

²⁰⁰ DHR, 193.

²⁰¹ Ibid. The General Staff had devised a plan in April 1931 with three options for solving the Manchurian problem, from replacing Zhang Xueliang with a pro-Japanese leader to an outright occupation of Manchuria. Shimida suggested that at the outbreak of the Mukden Incident, “the supreme command had the choice either of treating it as a minor incident to be settled immediately or of interpreting it within the larger framework.” Shimida, “Extension of Hostilities,” 252.

²⁰² DHR, 193; see also Hayashi’s 19 September telegram in DHR, 195.

the Kwantung Army, is it certain that [the Kwantung Army] was forced into defensive action due to violence from Chinese troops? Can we believe this?" Minami replied that it was definitely the case.²⁰³ However, at the formal session of the meeting, Foreign Minister Shidehara Kijuro produced intelligence reports showing that the incident was very likely a premeditated plot by the Kwantung Army.²⁰⁴ In this context, Minami could not push for the proposal to deploy reinforcements from the Korea Army to the Kwantung Army. The Cabinet decided to adopt a policy calling for the non-expansion of the Mukden incident.²⁰⁵

Around the same time, Vice-Chief of Staff Ninomiya, Military Affairs Bureau Chief Koiso, and Military Education Inspector-General Araki Sadao met to discuss the situation. They agreed that the current incident was an opportunity to achieve a satisfactory solution to the problems faced in Manchuria.²⁰⁶ But at the 2 pm meeting between Army Minister Minami, Army Chief of Staff Kanaya Hanzo and Inspector-General Araki, Minami expressed his agreement with the Cabinet's policy not to escalate the crisis. Subsequently, Kanaya instructed the General Staff that the Government's policy was to resolve the incident quickly and return to the status quo.²⁰⁷

However, the instructions were transmitted by the Army General Staff to the Kwantung Army in an ambiguous fashion: "(1) The determination and actions of the Kwantung

²⁰³ DHR, 194.

²⁰⁴ See, e.g., Telegrams No. 625 and No. 630 from the Consulate at Fengtian to Foreign Minister Shidehara, 19 September 1931, KZ, Vol. 1, 146; and Shimida, "Extension of Hostilities," 245-6.

²⁰⁵ DHR, 194.

²⁰⁶ Ibid.

²⁰⁷ Ibid.

Army Commander following the night of 18 September were highly appropriate, and [the Army General Staff] strongly believes that they have enhanced the prestige of the Imperial Army. (2) In view of the situation and the attitude of the Chinese side after the incident, the Cabinet has decided that the management of the incident should not exceed the appropriate limits. As such, the [Kwantung] Army's actions should follow this guideline accordingly."²⁰⁸ The ambiguity in the order might have allowed the Kwantung Army to interpret the order optimistically by emphasizing the first part and deemphasizing the second.

In the evening of 19 September, Chief of Staff Kanaya instructed Korea Army Commander Hayashi to hold his troops; a chaser was sent two hours later asking the Korea Army to confirm if Tokyo's order was implemented. In response, the Korea Army argued: "We still think it is necessary to dispatch a mixed brigade in view of the Kwantung Army's request for reinforcements and other general conditions, especially the need to guard the Antung-Mukden Railway. We would like to know, for our future reference, what special consideration has caused you to order us to postpone sending the brigade."²⁰⁹ A telegram in Kanaya's name was sent to Hayashi: "Your unilateral action this morning is not necessarily objectionable, in view of then-existing conditions. However, as the situation around Mukden has somewhat improved, I have given my consent to the cabinet decision not to extend hostilities for the time being, unless

²⁰⁸ DHR, 195. According to Crowley, "[War Minister] Minima cabled the gist of the cabinet's "localization" decision to General Honjo, the commanding officer of the Kwantung Army. Parallel with this message, however, the head of the operations division of the general staff, Colonel Imamura, also advised Honjo that the viewpoint of the cabinet was not necessarily binding for the army, because the Manchurian crisis involved the "right of supreme command."'" Crowley, *Quest for Autonomy*, 124.

²⁰⁹ Shimida, "Extension of Hostilities," 244.

unexpected conditions should develop. Your army's movement out of Korea must wait upon imperial approval. It is necessary to gather and hold the mobilized troops at Shingishu until further notice. Please telegraph immediately the present whereabouts of the mixed brigade."²¹⁰ At this point, the Korea Army deferred to Tokyo's authority and halted its troops at Shingishu.²¹¹

20 September 1931: The expansionists pushed forward

In the morning of 20 September, Vice-Chief of Staff Ninomiya, Army Vice-Minister Sugiyama Gen and Inspector-General Araki met and reaffirmed that the Army HQ should not make the Kwantung Army return to the status-quo.²¹² Army Minister Minami and Chief of Staff Kanaya also seemed to be converging to a similar position. Meeting in the afternoon, Minami, Kanaya and Araki discussed the Korea Army's reinforcement of the Kwantung Army. The meeting noted that since the Cabinet had already decided on a non-expansion policy, it would be inconvenient for the Army Minister to ask for troop reinforcements. Hence, the meeting decided to seek the Prime Minister's in-principle understanding that the Army might act without Cabinet approval if there was a "sudden change" in situation that did not permit time for ex-ante approval from Tokyo.²¹³

That evening, the Korea Army continued to press the Army HQ:

²¹⁰ Shimida, "Extension of Hostilities," 245.

²¹¹ Ibid.

²¹² DHR, 196.

²¹³ DHR, 195.

Regarding the Kwantung Army's urgent appeal to our Army to provide reinforcements, we deeply understand the [Army HQ's] honorable instruction on the necessity to await higher-level orders before mobilizing the forces. However, if higher-level orders are issued, that would start a war between Japan and China, and the situation would become worse. To localize the crisis, [the Korea Army] deeply believes that it would be better to activate the troops autonomously [without Cabinet sanction]. As such, we sincerely hope that the [Army HQ] vest the authority for us to move a part of our forces across the border according to the need of the situation.²¹⁴

21 September 1931: The Korea Army broke loose

By 21 September, the Kwantung Army had occupied Mukden, Changchun, Andong, Fenghuangcheng and Yingkou.²¹⁵ At 10 am, the Cabinet met to discuss the need for the Korea Army to reinforce the Kwantung Army. Only Army Minister Minami and Prime Minister Wakatsuki felt that there was such a need. The rest of the Cabinet including the Navy Minister disagreed, and the matter remained inconclusive.²¹⁶ At 11:30 am, the Army Minister and Chief of Staff sent a joint telegram to the Kwantung Army Commander that “the actions of your Army will definitely not become a meaningless sacrifice” and that the Army HQ will surely work towards the enhancing the “prestige and authority of the nation and the Japanese military.”²¹⁷

Later in the day, Commander Hayashi telegraphed Tokyo: “The Kwantung Army has already commenced operations in Jilin and [its] troops are evidently insufficient. We have repeatedly received urgent appeals [from the Kwantung Army] calling for

²¹⁴ DHR, 196.

²¹⁵ Jilin also succumbed on 21 September. Unrest instigated by the Kwantung Army provided a pretext for the Kwantung Army to march to the north. Ogata, *Making of Japanese Foreign Policy*, 67.

²¹⁶ DHR, 196.

²¹⁷ *Ibid.*

reinforcement by the Korea Army. [Their urgent appeals] cannot be ignored. I have ordered the brigade at Shingishu to move across the border. I am terribly apologetic for contradicting the [Army HQ's] earlier order."²¹⁸

22 September 1931: The Cabinet acquiesced

Apologetic or not, this was a fait accompli. At the Cabinet meeting on 22 September, no one voiced any active support for or opposition against the unauthorized border crossing by the Korea Army brigade.²¹⁹ In view that troops had already crossed the border, the Cabinet proceeded to release the funding for the operations. At 4:20 pm, the Emperor granted an audience to Chief of Staff Kanaya, who reported that the Korea Army's autonomous crossing into Manchuria was "truly an action forced by the necessity of the circumstances including the disparity in troop strength between the [Kwantung Army] and the enemy."²²⁰ The Emperor reportedly showed his displeasure when he approved the deployment, cautioning that the Army should "behave discreetly in the future."²²¹

Meanwhile, the Kwantung Army was pushing the Army HQ to authorize an expedition to "save" Harbin. At night on 22 September, Vice-Chief of Staff Ninomiya reminded the

²¹⁸ Ibid.

²¹⁹ DHR, 197.

²²⁰ Ibid.

²²¹ Ogata, *Making of Japanese Foreign Policy*, 66. According to the diary of Harada Kumao, who liaised between Emperor Hirohito and Prince Saionji, the Emperor had told Wakatsuki earlier: "The government's policy of non-enlargement is most proper. Strive to achieve that goal." On 22 September, Wakatsuki reported to the Emperor that the Cabinet had no choice but to approve the fait accompli by the Korea Army. "Swayed by Wakatsuki's opinions and out of meticulous respect for the authority of the prime minister's office, the Emperor then reluctantly approved Kanaya's request" Stephen Large, *Emperor Hirohito and Showa Japan: A Political Biography* (London and New York: Routledge, 1992), 47.

Kwantung Army that military moves to Harbin without Tokyo's approval was "strictly forbidden."²²² The next morning, Ninomiya, Army Vice-Minister Sugiyama, Military Affairs Chief Koiso and Inspector-General Araki jointly agreed on a memorandum, "Principles on the Limits of Military Occupation," which endorsed the occupation of a wide area along the South Manchuria Railway, including Qingdao, but excluding North Manchuria (with Harbin) and west of the Liao River towards Rehe.²²³

23 September 1931: Tokyo reasserted its authority

However, the Cabinet meeting on 23 September decided firmly that "even if the situation deteriorates in Qingdao and Harbin, no troops will be sent there to protect overseas [Japanese] citizens."²²⁴ After its acquiescence to the Korea Army's fait accompli, the Cabinet now decided to adopt a firm position. According to Crowley, Prime Minister Wakatsuki had reached a compromise with Army Minister Minami: the Cabinet would accede to the Army's "right of supreme command"; the Army General Staff would restrain the Kwantung Army; and Tokyo could exploit the incident to achieve a favorable treaty with Nanjing.²²⁵ When presented with the memorandum by Vice-Chief of Staff Ninomiya and others, Minami rejected it and persuaded Chief of Staff Kanaya to restrain the Kwantung Army within the South Manchuria Railway zone, with the already-

²²² Shimida, "Extension of Hostilities," 263.

²²³ Ibid, 264. North Manchuria was close to the Soviet Union; the area west of the Liao River was part of Zhang Xueliang's power base.

²²⁴ DHR, 199.

²²⁵ Crowley, *Quest for Autonomy*, 126.

occupied Jilin as the only exception.²²⁶ Army General Staff officers tried to persuade Minami the next day, but he remained firm with Kanaya on his side.²²⁷ On 24 September, the Army General Staff issued the following instructions to the Kwantung Army:

The chief of staff ... has accepted [the army minister's] view that it will be best to endorse and conform to the government's policy, so long as it does not interfere with the fundamental objective of settling the Manchurian question. Accordingly, now that the first stage of military action has been completed, we consider it advantageous for the Kwantung Army to recombine small units outside the railway zone, avoid permanent occupation of [Chengjiadun, Xinmin, and Dunhua], withdraw from [Jilin] if the situation permits it, and preserve intact the main force of the army in the South Manchuria Railway zone. These steps will help reveal the army's fair attitude, enable the army minister to act efficiently in the cabinet, and be of advantage from a general point of view as we try to achieve the ultimate objective. We are sending Major-General Hashimoto [Toranosuke] to you to convey our intent in detail.²²⁸

Ninomiya telegraphed the Kwantung Army to explain Tokyo's decision to evacuate citizens from Harbin and Qingdao if the situation worsened. The decision had been reported to the Emperor and would not be changed.²²⁹ This time, Tokyo was determined. The Kwantung Army had no choice. It cancelled the Harbin and Qingdao expeditions.

On 24 September, the Government released a proclamation on the Mukden crisis:

The Imperial Government had called for an emergency cabinet meeting on 19 September and decided to commit great effort to prevent the [Manchurian] incident from expanding. The Army [Minister] had transmitted the orders to the Commander of the [Kwantung Army]. On 21 September, although a portion of the troops had moved towards Jilin, [the deployment] was merely to protect the

²²⁶ Shimida, "Extension of Hostilities," 264.

²²⁷ *Ibid.*, 265.

²²⁸ *Ibid.*, 266.

²²⁹ *Ibid.*

Manchurian Railway from its surrounding threat and was not a military occupation. Upon completion of its mission, most of the deployed troops must immediately withdraw back to Changchun. In addition, in view of the unstable situation along the Manchurian Railway, a combined brigade with 4,000 troops from the Korea Army has been placed under the command of the Commander of the [Kwantung Army]. However, the total number of troops remains within the limits of [earlier] treaties. Hence, with regard to diplomatic relations, [the deployment] cannot be deemed an expansion of the incident.²³⁰

At this point, Tokyo had reined in the expansionists in the field. But the rein was loose.²³¹

The Wakatsuki Cabinet collapsed on 13 December and was replaced by the Inukai Cabinet. The militarist tide continued to grow within Japan. Thereafter, the Kwantung Army overran the rest of Manchuria. In 1932, it swallowed Jinzhou, Shanhaiguan, Tsitsihar, Harbin and Hailar. Rehe was taken in 1933.

Here we pause to interpret the motivations behind two highly consequential decisions: (1) The Cabinet's acquiescence to the Korea Army's unauthorized reinforcement of the Kwantung Army; and (2) Army Minister Minami and Chief of Staff Kanaya forcing the Kwantung Army to cancel the Harbin expedition. What drove these decisions?

The latter decision was influenced by the interaction of domestic political constraints, particularly those faced by Minami and Kanaya, and international strategic calculations, particularly the strategic fear of Soviet intervention. Minami and Kanaya were not doves. They had supported the expansionist line of their subordinates at the start of the crisis. But unlike their subordinates or the field army, Minami and Kanaya were directly

²³⁰ Proclamation by the Japanese Government on the Manchurian Incident, 24 September 1931, KZ, Vol. 1, 147.

²³¹ The Kwantung Army broke loose in a sudden thrust to Tsitsihar in November 1931. See Shimida, "Extension of Hostilities," 266-87.

accountable to the Emperor and the Cabinet. Apparently, Minami was restrained by the fact that Prime Minister Wakatsuki had promised the Emperor that no troops would be deployed to Harbin and Qingdao.²³² In a compromise with the Cabinet on 23 September, Minami agreed that the Kwantung Army would not move beyond Jilin in exchange for keeping Jilin.²³³

The Army HQ was also concerned about international strategic repercussions. The “General Outline of a Solution for the Manchuria-Mongolia Problem,” approved by both Minami and Kanaya in June 1937, adopted a more prudent approach than the Kwantung Army’s plans partly due to the concern over a potential Soviet move against Japan.²³⁴ The Navy and Foreign Ministries also feared that a Manchurian invasion might lead to a multilateral intervention by other great powers – analogous to the tripartite intervention in 1895 through which Japan was humiliated and forced out of Manchuria.²³⁵ In fact, on 19 September, Major-General Tatekawa Yoshitsugu from the General Staff personally told Ishiwara and his fellow conspirators that the Kwantung Army should be careful not to provoke the Soviets by invading the Russian’s sphere of influence in the north.²³⁶ But the conspirators had a different view. They believed that the Soviet Union was constrained by its domestic situation and “could not do much more than engage in propaganda work ... while there is always the danger of war with the Soviet Union if we move to Harbin, we should be able to avoid it by careful planning and instantaneous execution of our

²³² Shimida, “Extension of Hostilities,” 265.

²³³ Ibid.

²³⁴ Ogata, *Making of Japanese Foreign Policy*, 54.

²³⁵ Crowley, *Quest for Autonomy*, 128.

²³⁶ Coox, *Nomonhan*, 35.

plans, giving the Russians no opportunity to respond except with passive tactics.²³⁷ The difference in their attitudes towards military action suggests the following finding:

Observation 16: Tokyo was wary of Soviet intervention and adopted a more prudent approach towards Manchuria compared to the conspirators at the Kwantung Army, who believed that the Soviets were unlikely to intervene. The contrast in their beliefs and policy orientations supports Mechanism 2 (inadvertent enforcement).

Meanwhile, the domestic political constraints faced by the Emperor and the Cabinet also influenced the decision to acquiesce to the Korea Army's fait accompli:

Observation 17: Army radicalism made civilian leaders sensitive to the political costs of a direct clash with the Army, contributing to their acquiescence to unauthorized military actions during the Mukden crisis.

The tide of military activism was high in 1931. Wakatsuki had taken over the premiership in April 1931 from Hamaguchi Osachi, after Hamaguchi was seriously wounded in an assassination attempt by a right-wing army radical.²³⁸ Given the planned reductions in military budget under the Cabinet's administrative re-adjustment program, army officers had adopted more activist positions and saw the conflict in Manchuria in direct relation

²³⁷ Shimida, "Extension of Hostilities," 262.

²³⁸ On the domestic tensions within Japan during this period, see Takafusa Nakamura, *A History of Showa Japan, 1926-1989* (University of Tokyo Press, 1998), chapter 1.

to the army's own existence.²³⁹ With the troops already on the ground, Wakatsuki felt that he had no choice but to go along.²⁴⁰ During the Manchurian incident, imperial advisors also debated the risks and tradeoffs involved in the exercise of imperial authority: While it would be ideal if the field army fully complied with imperial orders to halt the hostilities, the Emperor's authority would be seriously damaged if the field army did not comply. Given the risks involved, as well as worries over incurring the Army's resentment towards the Emperor, the imperial advisors decided against the Emperor issuing imperial orders to stop the field army.²⁴¹ Indeed, the army's radicalism and the politicians' fears of the army's radicalism were sharply revealed in October 1931. A plot by a group of army officers to annihilate the Cabinet was exposed, but the punishment was light – the most severely punished conspirator got a twenty days' confinement.²⁴² Politicians did not want to confront the radicals in the army. Unauthorized military moves – even the October coup d'etat – were swept under the rug.²⁴³

²³⁹ Ogata, *Making of Japanese Foreign Policy*, 56-7; Jansen, *Japan and China: From War to Peace*, 379; Seki, "Manchurian Incident," 173

²⁴⁰ Crowley suggested that "[t]he decision of the premier not to challenge General Hayashi's actions was undoubtedly induced by a belief that the army and the Japanese public would accept nothing less than a reconfirmation of Japan's treaty rights in Manchuria." Crowley, *Quest for Autonomy*, 126.

²⁴¹ Stephen Large cited archival evidence to show that the imperial advisers "were deeply worried over the army's resentment of the Emperor's opposition to the fighting in Manchuria: "We agreed that the Emperor had better not say anything further unless necessary". They also agreed it would be better if Saionji, "who seems to harbor ill will toward the army", stayed away from the court for the time being." Large, *Emperor Hirohito*, 51. Jansen argued that "[i]n this kind of setting one can understand Saionji's hesitation to put forth his maximum strength in opposition. He had only one total weapon, the emperor, and this had to be saved for the ultimate crisis lest it lose its efficacy." Jansen, *Japan and China: From War to Peace*, 382.

²⁴² Jansen, *Japan and China: From War to Peace*, 382. For a summary of the "October incident," see Crowley, *Quest for Autonomy*, 131-5.

²⁴³ See Tobe, "Manchurian Incident," 97.

An activist position had crystallized within the Army HQ months before the Manchurian incident erupted. The Army General Staff and the Army Ministry had jointly agreed on the “General Outline of a Solution for the Manchuria-Mongolia Problem,” a major step approved by both Army Chief of Staff Kanaya and Army Minister Minami in June 1931. It was decided that “[t]he alleviation of anti-Japanese activities of the [Zhang Xueliang] regime shall continue to be undertaken primarily through negotiation by the Foreign Office, and the army shall maintain close contact with the Kwantung Army in order to make them act with discretion In spite of the above-mentioned efforts, should anti-Japanese activities be intensified, military action might become necessary.”²⁴⁴

But the Army HQ was also sensitive to the international and domestic implications. It emphasized that “[i]nternal as well as international understanding are absolutely necessary for the settlement of Manchurian problems”; there must be “careful preparations, such as publicizing the realities of anti-Japanese activities in Manchuria to the people of our country and of the powers, so that in the event of military action, public opinion will support the measure and the powers will not take opposing or suppressive steps.”²⁴⁵ A timeline was fixed based on this consideration: “Measures with regard to cultivation of internal and international understanding shall be undertaken with a view to achieving results in approximately one year, that is, by the spring of 1932.”²⁴⁶ The one-year time was needed for the Army to prepare for the contingency of third-power

²⁴⁴ Translated in Ogata, *Making of Japanese Foreign Policy*, 53-4. See also Seki, “Manchurian Incident,” 176; and DHR, 185-6.

²⁴⁵ Ogata, *Making of Japanese Foreign Policy*, 54.

²⁴⁶ *Ibid.*

intervention in Manchuria.²⁴⁷ The document highlighted the need for the Kwantung Army to “exercise restraint in the coming year and avoid any conflict arising from the anti-Japanese movement. In the event of a conflict, it should work toward localizing the incident and prevent it from expanding.”²⁴⁸

But the Kwantung Army’s position was much more aggressive. Tokyo limited the possibility of action within the south of Manchuria, fearing that a movement into the north might provoke Soviet intervention.²⁴⁹ The Kwantung Army wanted to seize the whole of Manchuria. And it did not want to wait for one year. Conspirators in the Kwantung Army went against Tokyo’s plans and lit the fuse on 18 September 1931. Why did they decide to do so? What were the calculations behind this momentous decision?

The calculations behind the Mukden conspiracy

Lieutenant-Colonel Ishiwara Kanji and Colonel Itagaki Seishiro led the conspirators in the Kwantung Army. The two men were responsible for the conception and implementation of the Mukden incident on 18 September 1931.²⁵⁰ In this section, I examine their motivations for triggering the Mukden crisis, based on the arguments they made in 1931 to persuade other Army officers. I focus on the material calculations, which

²⁴⁷ Peattie, *Ishiwara Kanji*, 114.

²⁴⁸ DHR, 186.

²⁴⁹ Peattie, *Ishiwara Kanji*, 97-8.

²⁵⁰ Seki described the Mukden incident “as the joint product of Ishiwara's conception and Itagaki's implementation, or perhaps of Ishiwara's ingenuity and Itagaki's influence.” Peattie recognized Ishiwara as “the driving force behind the military action.” Ogata concluded that Ishiwara and Itagaki “led the Kwantung Army ideologically, politically, and strategically, with the Commander-in-Chief and the Chief of Staff providing merely nominal leadership.” Seki, “Manchurian Incident,” 139; Peattie, *Ishiwara Kanji*, 122; Ogata, *Making of Japanese Foreign Policy*, 55.

were central in their arguments.²⁵¹ The arguments were connected by two major themes: the importance of Manchuria to Japan's security and future development; and the need to implement military action quickly.

In Ishiwara and Itagaki's view, Manchuria had critical strategic and military importance to Japan's future. Manchuria-Mongolia would be a crucial source of replenishments in a war against the United States, as well as the main battlefield in a war against the Soviet Union.²⁵² This view connects to Ishiwara's theory of a final world war between the lead nations of Asian and Western civilizations – namely, Japan and the United States. Ishiwara elaborated the theory in his earlier writings, which combined strategic analysis with his mystical belief of a predestined clash between East and West.²⁵³ To prepare for the coming conflict, Japan must secure control over Manchuria.

Ishiwara argued that “the most important prerequisite [for Japan's rise to greatness] is to secure our national defense.”²⁵⁴ If Manchuria were fully under Japanese control, it would

²⁵¹ Reading the evidence presented by Seki, Iriye highlighted that “[w]hat emerges from [Seki's account] is an almost total absence of ideology as a driving force behind military action. The Kwantung Army and the Tokyo military seem to have believed that their acts were logical expressions of a rational strategy, designed to carry out what to them was a legitimate goal of national policy.” Iriye, “Introduction: The Extension of Hostilities,” 238.

²⁵² Itagaki Seishiro, “A Military Perspective on Manchuria and Mongolia,” March 1931, in RDZDQ, 13. This was a lecture by Itagaki to mid-career army instructors. See Seki, “Manchurian Incident,” 165.

²⁵³ For a detailed analysis of Ishiwara's theory of final war, see Peattie, *Ishiwara Kanji*, chapter 3. Peattie pointed out that while Ishiwara's writings reflected some of his religious and idiosyncratic beliefs, they resonated with various assumptions held by individuals within and outside the military establishment at that time. These assumptions included an unavoidable clash with the United States and the strategic importance of dominating the Asian mainland to assure Japanese's future. Peattie, *Ishiwara Kanji*, 81-3. Unlike the radicals pushing for domestic reform in Tokyo, Ishiwara and the Kwantung Army prioritized external expansion over internal reform. See Peattie, *Ishiwara Kanji*, 73-4, 101.

²⁵⁴ Ishiwara Kanji, “Personal Views on the Manchuria-Mongolia Problem,” May 1931, KZ, Vol. 1, 82.

be hard for the Soviet Union to expand eastwards: “Our nation can relieve itself from its [defense] burden in the north, and thereafter move decisively toward China or Southeast Asia according to the needs of our national strategy.”²⁵⁵ Itagaki also emphasized Manchuria-Mongolia as Japan’s first line of defense: “From the reactive perspective, [controlling the region] would enhance the defense of Korea; and from the proactive perspective, it would check the eastern expansion of Soviet power and allow for strong speaking rights with regard to China.”²⁵⁶ Itagaki pointed out that the general terrain of the Asian mainland made it extremely difficult for foreign land armies to enter China without going through the Manchurian-Mongolian region. Hence, securing this region would place Japan on a position of strategic command with regard to China.²⁵⁷

Manchuria was also important in economic terms. Ishiwara argued that Manchuria’s agriculture produce could help to resolve the problem of food shortage for the Japanese; its coal and iron sources were sufficient to lay the foundations for heavy industries in Japan; and economic opportunities in Manchuria would help to alleviate unemployment in Japan.²⁵⁸ While the resources in Manchuria-Mongolia might not suffice to ensure Japan’s greatness, “they suffice to resolve current difficulties [in Japan] and to lay the foundations for a major leap forward.”²⁵⁹ Itagaki pointed out that the protectionist walls

²⁵⁵ Ibid. See also: Peattie, *Ishiwara Kanji*, 98. The document used the term “Southern Oceans” instead of “Southeast Asia”; the latter term is a contemporary coinage.

²⁵⁶ Itagaki, “A Military Perspective on Manchuria and Mongolia,” RDZDQ, 6.

²⁵⁷ Itagaki Seishiro, “On the Manchuria-Mongolia Problem,” 29 May 1931, in KZ, Vol. 1, 75. This was a series of lectures by Itagaki to officers in the main striking force of the Kwantung Army. See Seki, “Manchurian Incident,” 167.

²⁵⁸ Ishiwara, “Personal Views on the Manchuria-Mongolia Problem,” KZ, Vol. 1, 82-3.

²⁵⁹ Ishiwara, “Personal Views on the Manchuria-Mongolia Problem,” KZ, Vol. 1, 83. Ishiwara saw the Manchurian expansion as crucial in providing “breathing space” to Japan. See Peattie, *Ishiwara Kanji*, 97.

erected by other great powers had rendered the smaller Japanese economy to a position of dependence on the Western economies. Without a resource hinterland and an export market under Japanese influence, it would be impossible to match the other great powers.²⁶⁰ This implicated national survival. “If forced to abandon Manchuria-Mongolia due to national weakness, Japan would be entrapped between three gigantic nations – the United States, the Soviet Union and China – and relegated to the status of a minor nation like Belgium or Holland. Put in a subservient position under these great powers, national self-destruction will be the inevitable outcome.”²⁶¹

The timing of military action is another major theme in Ishiwara and Itagaki’s arguments. Military action must be executed quickly. Further delay would be detrimental as external realities were shifting in Japan’s disfavor. The Army HQ in Tokyo, however, wanted to delay military action for a year to work on domestic and international opinion.²⁶² Some in the Army General Staff felt that time was necessary to make elaborate preparations for the contingency of Soviet military movements when Japanese forces enter Manchuria.²⁶³ But Ishiwara and Itagaki were not much concerned with domestic opinion or international diplomacy.²⁶⁴ They made four arguments based on four specific windows of opportunity.

²⁶⁰ Itagaki, “On the Manchuria-Mongolia Problem,” KZ, Vol. 1, 75.

²⁶¹ Ibid.

²⁶² See the discussion on the “General Outline of a Solution for the Manchuria-Mongolia Problem” in the previous section.

²⁶³ DHR, 187.

²⁶⁴ Peattie, *Ishiwara Kanji*, 111. See also Seki, “Manchurian Incident,” 161-2. Both men were prepared to fight even in the unlikely event that foreign powers intervened in Manchuria. If it could not be avoided, they felt that Japan should fight the United States, Britain, Russia, and China simultaneously. Itagaki, “On the Manchuria-Mongolia Problem,” KZ, Vol. 1, 80; Peattie, *Ishiwara Kanji*, 71; Seki, “Manchurian Incident,” 166.

The first two were based on Japan's position vis-à-vis the Soviets and the Western powers. These windows yield Observations 17 and 18:

Observation 18: Conspirators at the Kwantung Army argued that the Soviet Union was currently weak and hence unlikely to intervene in a Sino-Japanese conflict; but in the future it would become strong and more likely to intervene. This finding supports Mechanism 2 (inadvertent enforcement) and Theory 1 (commitment theory of war).

Observation 19: Conspirators at the Kwantung Army believed that the League of Nations was unlikely to act as an external enforcer, and hence the invasion of Manchuria had little risk of external intervention. This finding supports Mechanism 1 (exogenous enforcement).

The first was the Soviet-Japanese window. Ishiwara and Itagaki believed that the Soviet Union would become stronger in the future. But at present, the Soviets were unlikely to intervene against Japan in Manchuria. Itagaki noted: "Currently, due to various factors, the Soviet Union faces major constraints in the use of military force in the Far East. However, with the five-year plan revitalizing its economy, the Soviet Union is now achieving steady success in its military development. Given its rise in national strength, there is a very high possibility that it would implement an active foreign policy in the Far East. Hence, our nation should settle the problem of Manchuria-Mongolia as quickly as

possible.”²⁶⁵ Ishiwara also warned that Soviet military capabilities had strengthened over time since the launch of the five-year plan, and that the Soviet Union would soon become Japan’s major adversary in the Far East. “But as of today, there is no need to worry too much [about the Soviets] as we move toward the construction of Manchuria [under Japan].” The present moment was opportune for creating a solid defense against the Southern penetration of Communist power.²⁶⁶

At the same time, the United States, Britain, and France were trapped in economic crisis and had competing interests in the Far East. The League of Nations was an impotent entity in the present situation. At the moment, Ishiwara was convinced that none of these actors were willing or able to intervene against Japan in Manchuria. But Japan had to move quickly and decisively, as the international situation might change and become less favorable in the future.²⁶⁷

The third and fourth windows yield Observation 20:

Observation 20: Ishiwara argued that Japan’s position in Manchuria would become vulnerable in the future, given the growth of the Chinese nationalist movement in Manchuria, and that China would become strong once it achieved national unification. Hence, the Kwantung Army should

²⁶⁵ Itagaki, “On the Manchuria-Mongolia Problem,” KZ, Vol. 1, 80.

²⁶⁶ DHR, 191. See also Peattie, *Ishiwara Kanji*, 97; Seki, “Manchurian Incident,” 149.

²⁶⁷ Seki, “Manchurian Incident,” 211; Peattie, *Ishiwara Kanji*, 114-5.

take over Manchuria quickly. This supports Theory 1 (commitment theory of war).

Ishiwara believed that while the Chiang regime was now weak and divided, it would become strong and united in the future. He argued that “China’s nationalist movement and its movement to recover its sovereign rights are inevitable, and Chiang Kaishek’s unification effort is bound to succeed. China’s national strength will gather and concentrate over time.”²⁶⁸ Japan must move into Manchuria while the window of opportunity remained open, as “there is constant infighting among the Chinese right now, and they will not be able to make a strong response over the issue of Manchuria.”²⁶⁹ China would not react strongly at this point given that it was weak and divided.

Finally, the situation in Manchuria was shifting in Japan’s disfavor. In the past, Manchuria had a close relationship with Japan to the extent that China had found it difficult to intervene. However, Manchuria was now moving closer and closer towards the Chinese side.²⁷⁰ Itagaki pointed out that Zhang Xueliang had set up a Kuomintang (KMT) branch in Manchuria and had paid official visits to Nanjing, and that KMT propaganda was infiltrating the region to the extent that even primary schools in the countryside were implementing an anti-Japanese curriculum. Anti-Japanese sentiments

²⁶⁸ DHR, 191

²⁶⁹ Ibid.

²⁷⁰ Itagaki, “On the Manchuria-Mongolia Problem,” KZ, Vol. 1, 77.

were growing strong and the situation was becoming more and more serious.²⁷¹ Japan should take control of Manchuria before the situation deteriorated further.²⁷²

These were the arguments made by Ishiwara and Itagaki to other Army officers in 1931.²⁷³ For both men, Manchuria was critical to Japan's security and future development. Their calculations of realities and trends – in Manchuria, China, the Soviet Union and the West – led them to conclude that military action must be taken quickly. They lit the fuse and sparked the Mukden crisis. The crisis led to a fifteen-year quagmire in China with momentous consequences for Japan and East Asia.

²⁷¹ Ibid. Zhang had also launched major initiatives for industrial and agricultural development, including the construction of a Chinese railway that would undermine the business of Japan's South Manchuria Railway. Ogata, *Making of Japanese Foreign Policy*, 16-17. Ishiwara argued that if Japan did not move forward and take over Manchuria, Japan would be eventually be pushed out of the region. Peattie, *Ishiwara Kanji*, 96.

²⁷² Itagaki did not think that there would be nationalist resistance from the Chinese in Manchuria. He argued that as long as their livelihoods were not threatened, the Chinese in Manchuria would be largely indifferent to the political situation just like overseas Chinese living in colonized parts of Southeast Asia. Itagaki, "On the Manchuria-Mongolia Problem," KZ, Vol. 1, 76.

²⁷³ Both men might also have personal urgency in the matter: their three-year tour of duty in Manchuria was ending and they were due for rotation. Peattie, *Ishiwara Kanji*, 114-5; Marius Jansen, "Introduction: The Manchurian Incident, 1931," in *Japan Erupts: The London Naval Conference and the Manchurian Incident, 1928-1932*, ed. James Morley (New York: Columbia University Press, 1984), 137.

6. Preventive Implications

Given the historical conditions in each crisis, what were the policy possibilities that could have prevented the outbreak of war in 1931 and 1937? The Second Sino-Japanese War in 1931 is an unusual case of a war triggered deliberately by a group of radical army officers, given particular political weakness in both Tokyo and Nanjing. With the specific circumstantial constraints – and without heroic counterfactuals on our part – it is not clear what decision-makers in the crisis could have rationally done otherwise to stop the Kwantung Army in its tracks.²⁷⁴ But what about the 1937 crisis that led to one of the most costly wars in modern history? Are there realistic policy possibilities that could have reduced the risk of war?

From the evidence, it appears that perceptions of potential vulnerability had an immediate effect on the calculations *for* war in both Tokyo and Nanjing. *Based on Theory 1, Mechanism 4 (Refined), and Mechanism 6:*

²⁷⁴ The conspirators at the Kwantung Army were determined to act. The domestic political situation in Tokyo made it hard for civilian leaders to take a strong stand against the Army. On the Chinese side, Nanjing was unlikely to use force against the Kwantung Army, as Nanjing was in such a delicate position and the force of inadvertent enforcement was so strong. It is also not clear if all-out military resistance by Nanjing would have prevented the expansion of hostilities – it could have expedited rather than deterred the Japanese deployment of troops into Manchuria. Recall, for instance, that the emergency Cabinet meeting on 19 September had focused on whether the Kwantung Army was defending itself against violence from Chinese troops. Foreign Minister Shidehara managed to (temporarily) prevent the deployment of reinforcements only after showing evidence that the Kwantung Army was the aggressor rather than the victim of violence. See DHR, 194. On the whole, historians of this period are generally pessimistic about the possibility of preventing the Sino-Japanese War in 1931. Akira Iriye, for example, pointed out that “[a]n open conflict could have been avoided only if there had been strong leadership in both Japan and China capable of imposing their determination to maintain peace on the respective populations. ... A strong government in Japan might have restrained army action in Manchuria and postponed a showdown with China on the basis of some compromise settlement on the issue of Japanese treaty rights. But the government in Tokyo was too weak and too unwilling to risk its existence by a strong stand.” Iriye, *After Imperialism*, 295.

Policy Possibilities:

Possibility (1): If feasible, send signals uncorrelated or weakly correlated with the opponent's defensive vulnerability. Be careful with signals of resolve that may deepen the opponent's vulnerability.

Possibility (2): Set a focal-point limit to the signaling, and communicate the focal point to the opponent clearly. (For instance, Japan could have amassed its troops but stopped them at a reassuring distance from Beijing.)

Possibility (3): Announce a non-offensive strategy defined in a series of conditional steps (i.e. with each increment of escalation conditional on the opponent's escalatory move.).

From the evidence, it also appears that inadvertent enforcement had an immediate effect in Tokyo on the calculations *against* war. *Based on Mechanism 2:*

Policy Possibilities: War is less likely when it will provide a strategic advantage for a third-party rival. The logic of inadvertent enforcement, however, may not be easily recognized by the opponent. For instance, there is little evidence to suggest that Chiang had recognized it in July 1937 or that Japan had tried to signal its strategic dilemma to Chiang. Signaling the logic of inadvertent enforcement to the

opponent may calm down the commitment problem and provide a more positive environment for peaceful negotiations.

These policy possibilities are historically contingent. They are based on the historical situation in which Nanjing and Tokyo found themselves in the Marco Polo Bridge crisis. These policy possibilities are also non-exhaustive. But they point us to a few potentially positive things that Tokyo or Nanjing could have done in July 1937 – but didn't.

PRELUDE TO CHAPTER 4

The archival evidence in Chapter 3 supports the theoretical predictions in Theory 1 (commitment theory of war) and Mechanism 2 (inadvertent enforcement). For Mechanism 1 (exogenous enforcement), the evidence provides direct support in the 1931 case and indirect support in the 1937 case. The 1937 case is also consistent with Mechanism 3 (endogenous enforcement), although the 1931 case is uninformative.

On the other hand, evidence for the signaling mechanisms (Mechanisms 4 to 6) is mixed and mostly indirect. Does this mean that the signaling mechanisms have weaker effects compared to the enforcement mechanisms? Such a conclusion is unfair simply because it is difficult to test signaling effects with observational data. Rarely are we able to observe the signaler's true resolve at the point of signaling or the receiver's credibility estimates at the point of signal reception. The information environment is very hard to control and analyze in a real-world setting. It is difficult to separate the effect of one signal – a sunk-

cost signal, a costly-implementation signal, or a contradictory signal – from the effects of other previous or simultaneous signals or pieces of information. Observational data are at once blessed with realistic complexities and cursed with confounders that are part of the complex reality. A confound-free test of the signaling mechanisms may be beyond the scope of observational evidence. In the next chapter, I use controlled randomized experiments to isolate the causal effects of the mechanisms as cleanly as possible.

Chapter 4

Enforcement, Signaling and War: Evidence from Ten Experiments

The previous chapter used case studies to investigate whether and how the hypothesized mechanisms operated in historical crises. This chapter uses experiments to test the causal effect of each mechanism.

Causal inference with observational data is a task riddled with difficulties. The task is especially challenging with rationalist explanations for war. Variations in the information environment or the commitment environment are hard to measure. The effects of potential confounders are impossible to rule out with observational data if the confounders are unmeasured, unobserved, or unobservable.

I use an experimental setting to control and manipulate the key parameters in the mechanisms. Random assignment of experimental treatments allows me to cleanly identify their causal effects, avoiding the pitfalls associated with potential confounders or selection effects.

I present ten experiments in this chapter. They include experiments conducted in a laboratory setting, fielded over the Internet, or embedded within surveys. The experiments involve medium- n samples recruited locally as well as large- n samples recruited nationally. Most experiments are time-shared to achieve economies of scale.

This chapter has five parts. In Chapter 2, I tested the commitment theory of war with a focus on exogenous enforcement in its simplest form. **Part 1** of this chapter tests the mechanism of inadvertent enforcement, which is a more complex form of exogenous enforcement (Experiment 1). **Part 2** focuses on endogenous enforcement based on strategic restraint (Experiment 2). **Part 3** investigates the costly signaling mechanism (Experiments 3-5). **Part 4** examines the costly implementation mechanism (Experiments 6-8). **Part 5** concludes by studying the relationship between contradictory signaling and signal credibility. Part 5 also shows how we can use experiments (Experiments 9-10) to construct stylized historical counterfactuals: While it is impossible to rerun history to answer counterfactual questions, it is possible to design experiments to test the theoretical expectations behind the counterfactuals.

This is a long chapter. The reader may wish to treat it as five mini-chapters based on its five parts. Each part is a standalone and can be read separately with cross-references. To conserve space, supplementary regression tables are relegated to Appendix B2. The experimental protocols and instructions are reproduced in Appendices A2 to A5.

Table 4.1 summarizes the main experimental results:

Table 4.1: Summary of Experimental Findings

Experiment	Type	Main Finding
#1: Inadvertent Enforcement	Game-Theoretic Lab Experiment	War incidence drops sharply under inadvertent enforcement. When enforcement is asymmetric, Player A exploits Player B with lower final offers.
#2: Endogenous Enforcement	Game-Theoretic Lab Experiment	Endogenous enforcement is successfully achieved in half the cases. In these cases, the incidence of conflict is significantly reduced.
#3: Costly Signaling I	Game-Theoretic Internet Experiment	Unexpected asymmetry in the behaviors of Signalers and Receivers: Signalers randomly assigned with high resolve are much more likely to send the threat with sunk cost. But Receiver's acquiescence rate does not respond significantly to the sunk-cost signal, contrary to model predictions.
#4: Costly Signaling II	Game-Theoretic Lab Experiment	
#5: Costly Signaling III	Contextualized Survey Experiment	Level of sunk cost (military mobilization) has no significant effect on perceived threat credibility.
#6: Costly Implementation I	Contextualized Survey Experiment	Cost of threat implementation has a significant negative effect on the credibility of the threat.
#7: Audience Costs	Contextualized Survey Experiment	Publicly announced threat to use force is more credible when it involves high audience costs.
#8: Costly Implementation II	Contextualized Survey Experiment	Threat is less credible when a past incident suggests a high cost of threat implementation.
#9: Contradictory Signaling	Experiment with Stylized Historical Counterfactuals	In treatments based on Korean-War counterfactuals: Threat is much more credible in <i>No-Acheson/No-Withdrawal</i> treatment (no salient contradiction) than <i>Acheson/Withdrawal</i> treatment (salient contradiction).
#10: Costly Implementation III	Experiment with Stylized Historical Counterfactuals	In two Korean-War counterfactuals, the cost of threat implementation reduces the credibility of the threat.

PART 1: INADVERTENT ENFORCEMENT

Under inadvertent enforcement, peace is enforced when war would create a strategic advantage for a third-party rival who might later threaten the parties in conflict. Enforcement is inadvertent as the third-party rival preserves the peace even though it has no intention to do so. The logic of inadvertent enforcement has ancient origins, but remains underdeveloped in international relations theory. In this section, I formalize the logic and show how inadvertent enforcement resolves the commitment problem and prevents war. I also implement an experiment to pin down the causal relationship.

1.1 Model

To illustrate the theory formally, I focus on how inadvertent enforcement prevents war in the shadow of a commitment problem. In a rationalist framework, war should never occur unless there is asymmetric information and/or a commitment problem. Based on Fearon's model, two states will always reach a peaceful bargain to avoid a costly war given symmetric information and the absence of a commitment problem. I begin with a commitment-problem game where war is the equilibrium outcome, and show how the introduction of inadvertent enforcement generates a peaceful equilibrium. I consider two variants of inadvertent enforcement: a symmetric variant in which the third player is a symmetric strategic rival of both the parties in conflict; and an asymmetric variant in which the third player is the rival of one but not the other.

Theoretically, a triadic bargaining model of war with shifting power can be extremely complex with many possible variations.¹ Empirically, it is generally difficult to design and implement a clean experiment when it involves complex interactions. It turns out, however, that neither the theory nor the experiment requires three-player bargaining and its many possible permutations. In fact, inadvertent enforcement requires that the third-party rival *cannot* make a bargain or alliance with a warring party, as explained in Chapter 1. This allows us to construct a three-player model that involves only bilateral bargaining between two players, excluding the third-party rival from the bargaining interactions and treating it purely as an inadvertent-enforcement device.

Hence, the goal is not to construct a general model of triadic bargaining but to locate the most basic model that allows us to isolate the treatment effect of our variable of interest – inadvertent enforcement. The theoretical motivation is to capture the logic of inadvertent enforcement under a commitment problem in its simplest form. The empirical motivation is to have an experimentally-feasible model that allows the causal relationship to be identified as cleanly as possible. The experiment evaluates the treatment effect in a baseline model of inadvertent enforcement. Once this is established, more complex models may be tested to expand our knowledge on different variants of inadvertent enforcement. Various complications – such as different configurations of the triadic relationship, different levels of power in the third-party rival, or asymmetric information on different parameters of the game – can be introduced incrementally into the baseline model in future work.

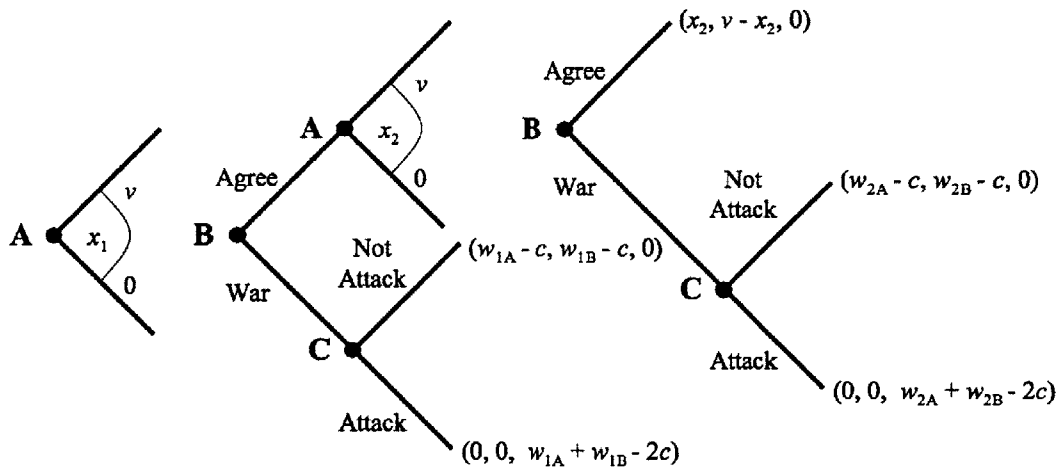
¹ For this reason, the canonical commitment-problem models of war are based on bilateral bargaining.

I begin with the two-stage commitment-problem game described in Chapter 2.² The game has two players (A and B) bargaining over a prize with the value v over two stages. In Stage 1, A makes a demand $x_1 \in [0, v]$. B observes x_1 and decides to either agree or fight. If B fights, the game ends with payoff $(w_{1A} - c, w_{1B} - c)$, with $w_{1A} = w_{1B}$, where w_{1i} is Player i 's war payoff in Stage 1 and c is the constant for the cost of war. If B agrees, the game enters Stage 2 in which A confirms its demand $x_2 \in [0, v]$. B observes x_2 and decides to either agree or fight. The game ends with payoff $(x_2, v - x_2)$ if B agrees and $(w_{2A} - c, w_{2B} - c)$ if B fights, where $v = w_{1A} + w_{1B} = w_{2A} + w_{2B}$ and $w_{2A} > w_{1A} > w_{2B} > c$. The payoffs reflect equal bargaining power between A and B in Stage 1 and a decline in relative power for B in Stage 2. There is complete information in the game.

Consider two variants of the commitment-problem game with a third-party rival (Player C). In the first variant (symmetric enforcement), C can choose to attack A and B if B chooses to fight in Stage 1 or Stage 2. Assume that C can take over the payoffs of A and B if it attacks after A and B are weakened from fighting each other. The game ends with the payoff $(0, 0, w_{1A} + w_{1B} - 2c)$ if C attacks in Stage 1 and $(0, 0, w_{2A} + w_{2B} - 2c)$ if C attacks in Stage 2. Otherwise, C gets nothing and the payoffs for A and B remain unchanged. Figure 4.1 diagrams the game.

² A two-stage model offers the simplest possible structure through which a commitment problem can be generated. To focus sharply on the commitment problem in its cleanest form, interesting variations that are non-essential to the commitment problem are black-boxed. For instance, the cost of war is assumed to be symmetric and constant; the war technology is black-boxed such that players end up with an absolute payoff without uncertainty if war is chosen; and the infinite-period bargaining framework in a more general model (e.g. Fearon's model) is reduced into two-stage bargaining to make it experimentally implementable. James Fearon, "Rationalist Explanations for War," *International Organization* 49, No. 3 (1995): 379-414.

Figure 4.1: Crisis Bargaining Game with Inadvertent Enforcement



The second variant (asymmetric enforcement) is the same as the first except that C is a third-party rival of B but not of A. Hence, C can only attack B if B chooses to fight in Stage 1 or Stage 2. The game ends with the payoff $(w_{1A} - c, 0, w_{1B} - c)$ if C attacks in Stage 1 and $(w_{2A} - c, 0, w_{2B} - c)$ if C attacks in Stage 2. Otherwise, C gets nothing and the payoffs for A and B remain unchanged.

Note that in the model, A and B operate in a strategic-interaction environment but not C, which operates in a purely decision-theoretic setting. We are interested *not* in the decision of C in itself, but in the *existence* of Player C and how that changes the behavior and outcome for Players A and B. As already mentioned, the model of inadvertent enforcement no longer applies if Player C is endogenized as a bargaining player. Hence,

Player C is excluded from the bargaining process so that it can operate purely as an inadvertent-enforcement device with minimal confounds.

In the experiment, $v = 10$, $c = 2$, $w_{1A} = w_{1B} = 5$, $w_{2A} = 7$, and $w_{2B} = 3$. The payoff parameters are constrained in the integer range within $[0, 10]$. The parameter values are selected based on the reasons discussed in Chapter 2.

Commitment-Problem Game without Inadvertent Enforcement

This variant of the model is similar to the commitment-problem game in Chapter 2 with complete information and no enforcement. The equilibrium solution is the same. Hence:

Prediction 5: War will occur with certainty in the commitment-problem game without inadvertent enforcement.

Commitment-Problem Game with Symmetric Inadvertent Enforcement

C will choose to attack in Stage 1 since $w_{1A} + w_{1B} > 2c$, just as it will also choose to attack in Stage 2 since $w_{2A} + w_{2B} > 2c$. Hence, in the subgame perfect equilibrium, A will demand $x_1 = x_2 = v - e$, where e is the smallest possible increment to B's reservation level (0) in Stages 1 and 2, and B will accept since $e > 0$. Hence:

Prediction 6: War will be avoided in the commitment-problem game with symmetric inadvertent enforcement.

Commitment-Problem Game with Asymmetric Inadvertent Enforcement

C will choose to attack in Stage 1 and Stage 2 since $w_{1B} > c$ and $w_{2B} > c$. The subgame perfect equilibrium converges to same equilibrium outcome under symmetric inadvertent enforcement: A will demand $x_1 = x_2 = v - e$ in Stages 1 and 2, and B will accept. Hence:

Prediction 7: War will be avoided in the commitment-problem game with asymmetric inadvertent enforcement.

1.2 Experimental Design and Implementation

Do the theoretical predictions hold in a controlled experiment? In experimental economics, it is common to see strategic interactions between human players in the laboratory diverge from the game-theoretic equilibrium predictions. At the same time, the experiments in Chapter 2 showed that the commitment problem has a strong positive effect on the incidence of conflict. Can inadvertent enforcement shut down the commitment problem and reduce the risk of conflict? I implement an experiment to address this question.

Setup

The experiment was programmed and implemented on z-Tree, with the participants interacting with each other anonymously through computer terminals.³ Four sessions were conducted at the MIT Behavioral Research Laboratory between 27 February and 8 March 2013. A total of 64 MIT students were recruited through the lab. Subjects were paid solely based on their performance. The computer randomly chose nine out of the fifteen rounds they played, and counted their point earnings in those rounds as payment. Each round had a total possible value of 10 points, with each point equivalent to \$0.50. Subjects earned \$19.42 on average for the one-hour session. Each subject participated in only one session.

Sequence

The experiment was embedded in the first eight rounds (Rounds 1 to 8) in each session.⁴ At the start, participants were informed that they would play a total of 15 rounds over multiple scenarios. A scenario would have one or more rounds, with each round separate and independent. Participants played Rounds 1 to 8 in this experiment before transiting into two other games (Experiments 2 and 4). They ended the experiment with a risk-aversion test (Round 15) that measured their individual risk preferences.

³ Urs Fischbacher, "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics* 10, No. 2 (2007): 171-78.

⁴ As the experiment was embedded in the first 8 rounds in each session, there is no possibility of spillover effects from the games played after Round 8. Subjects also knew that each round would be independent from previous rounds, and that they would be randomly re-matched after every round.

Instructions

Participants began by reading the instructions on their computer screens. They were placed in the role of a national decision-maker bargaining with another country for a valuable prize. The situation was framed in terms of crisis bargaining and war.⁵ The game was carefully explained, with the payoff difference between Stages 1 and 2 highlighted. In Stage 1, both countries are equally powerful: if they fight a war, each country will seize 50% of the prize (5 points) for itself. In Stage 2, Country A becomes more powerful, and if they fight a war, A will seize 70% of the prize (7 points). As war is costly, each country will lose 2 points if they fight a war. The shift in power between Stages 1 and 2 generates a commitment problem.

The instructions were written in neutral language, with a set of test questions to check understanding. After the test questions, there were three additional safeguards to ensure that participants fully understood how the game worked. First, in the screen immediately after the test questions, participants saw the answers, which showed what the participant got right or wrong. Next, participants saw a screen summarizing the game and payoffs. Finally, the information also appeared on the decision screen.

Assignment

Rounds 1-5: Subjects were randomly divided into two experimental groups (control and treatment). They remained in the same experimental group in Rounds 1 to 5. By design,

⁵ The experimental section in Chapter 2 discussed this point.

the game was similar for both groups, except that the game had a third player (C) for the treatment group but not the control group. C can choose to attack A and B whenever there is a war between A and B in Stage 1 or 2. If C attacks, it will win a decisive victory and take over the war payoffs of both A and B. The existence of C in the game constructs the inadvertent enforcement mechanism in the treatment group.

Rounds 6-8: In Round 6, all subjects in the experiment were suddenly switched into a scenario in which C can only choose to attack B (but not A) in the event of a war between A and B. If C attacks, it will take over B's payoff. This constructs an inadvertent enforcement device with a third-party rival for the declining power. Rounds 6-8 are designed to test the model predictions described in the previous section. If the model holds, we should expect two distinct patterns in the data:

- No significant difference in the incidence of conflict when subjects in the treatment group (*symmetric* inadvertent enforcement) switch into Rounds 6-8 (*asymmetric* inadvertent enforcement).
- Significant difference in the incidence of conflict when subjects in the control group (*no* inadvertent enforcement) switch into Rounds 6-8 (*asymmetric* inadvertent enforcement).

Subjects were randomly paired to each other at the start of every round. For Rounds 6-8 and for the treatment group in Rounds 1-5, two subjects in each round were randomly assigned to the role of C, with the remaining subjects randomly assigned to the roles of A

and B.⁶ This logistic strategy allows me to construct the inadvertent-enforcement device with human players without using deception or losing too many observations from the A-B pairs. Appendix A2 reproduces the experimental protocols and instructions in detail.

1.3 Experimental Results

Result 7: Inadvertent enforcement sharply reduces the incidence of war.

Figure 4.2 traces the incidence of war (war outcomes as a percentage of all outcomes) with and without inadvertent enforcement in Rounds 1 to 5. The overall incidence of war is 22% in the treatment group (inadvertent enforcement) compared to 75% in the control group (no enforcement) (two-tailed test of proportion, $p = 0.0000$, $n = 140$). In the first round, the incidence of war is 15% in the treatment group and 80% in the control group ($n = 28$: two-tailed test of proportion, $p = 0.0006$; Mann-Whitney test, $p = 0.0008$). Table 4.2 shows the incidence of war per player in each condition. The overall incidence of war per player is more than three times as high in the control group compared to the treatment group with inadvertent enforcement (two-tailed t-test, $p < 0.0001$, $n = 64$).

⁶ The decision data from C is used to generate outcomes for A and B at the end of the game. There is no variation observed in all four sessions of the experiment: everyone assigned to the role of C chose to attack in a war between A and B. Since our focus is on the decisions of A and B, it is undesirable to lose too many observations by assigning a Player C for every A-B pair. All subjects in the treatment group knew that they were playing a three-player game. However, no one (including those in the role of C) knew that the decision data from C is “shared” by other A-B pairs to generate final outcomes.

Figure 4.2: Incidence of War Across Experimental Conditions (Rounds 1-5)

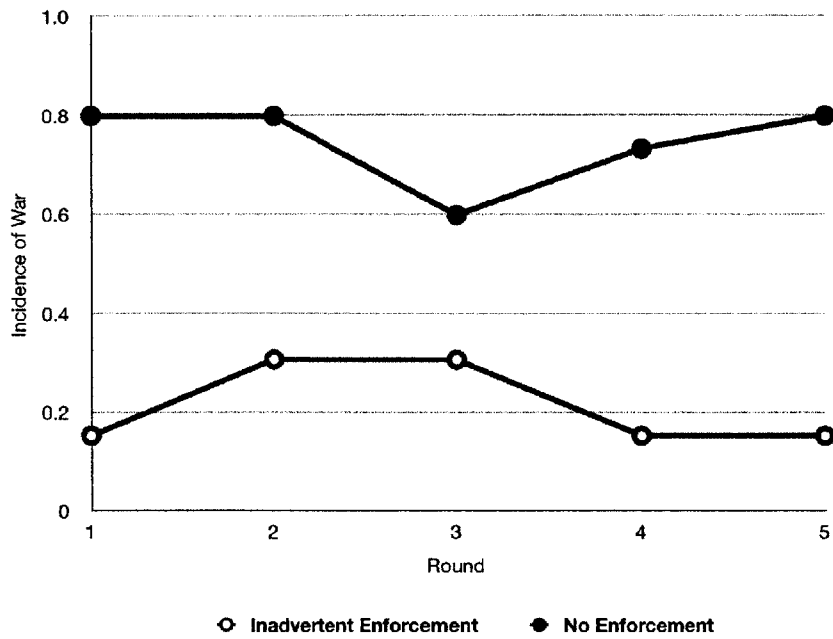


Table 4.2: Player Incidence of War

Session	No Enforcement	Inadvertent Enforcement
1	0.80	0.34
2	0.80	0.14
3	0.70	0.14
4	0.70	0.23
Overall	0.75	0.21

Notes: Maximum incidence is 1.00 (100%). Total of 64 subjects in four sessions: Session 1 ($n = 14$), Session 2 ($n = 16$), Session 3 ($n = 16$) and Session 4 ($n = 18$).

The evidence shows that war is much less likely under inadvertent enforcement. As a robustness check, I use logit models to estimate the relationship between the inadvertent-enforcement treatment and the war outcome. I use model specifications with

and without controls for individual risk preferences and the size of Player A's initial offer. Risk preference is measured on a summed score based on the risk-preference test at the end of the experiment (Round 15): the higher the score, the greater the individual willingness to take risk. Appendix A2 (last section) describes the risk-preference test. Each model uses one dyadic observation per round and session, with robust standard errors corrected for clustering at the subject level. Round and session dummies are used to control for time and session fixed effects. Table B2 in Appendix B2 shows the logit estimates. The results show that inadvertent enforcement has a negative relationship with war outcome that is strongly significant across all model specifications ($p < 0.001$). The size of the initial offer is positively and significantly related to war ($p < 0.01$). The risk-preference variable, however, is statistically insignificant ($p > 0.60$).

War can occur in either Stage 1 or Stage 2 of the game. In Rounds 1-5, the incidence of war decisions in Stage 1 is much lower at 11% with inadvertent enforcement compared to 67% without inadvertent enforcement (two-tailed test of proportion, $p < 0.0001$, $n = 140$). In the first round, the incidence of war decisions in Stage 1 is 0% with inadvertent enforcement but 60% without inadvertent enforcement ($n = 28$: two-tailed test of proportion, $p = 0.0007$; Mann-Whitney test, $p = 0.0009$). Table B3 (Appendix B2) shows logit estimates for the relationship between the inadvertent-enforcement treatment and the decision for war in Stage 1. The results are consistent with those in Table B2, with a strongly significant negative relationship between inadvertent enforcement and the Stage-1 decision for war across all model specifications ($p < 0.001$).

Result 8: When inadvertent enforcement is suddenly introduced, the incidence of war suddenly falls.

Figure 4.3: Incidence of War in Control Group (Rounds 1-8)

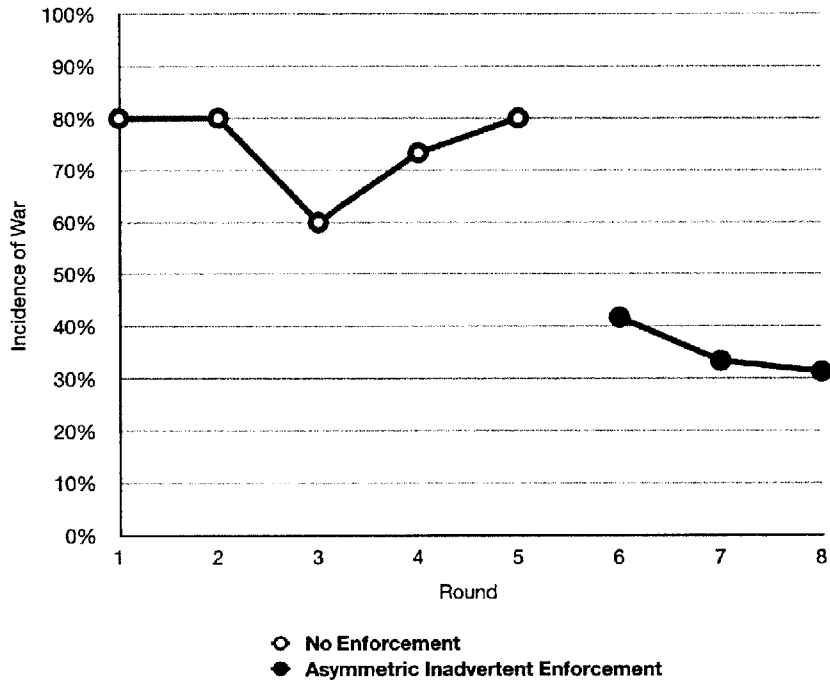


Figure 4.3 traces the change in the war incidence for the control group, which was suddenly switched to an inadvertent-enforcement treatment of the asymmetric variant in Round 6. The overall incidence of war in Rounds 1-5 is more than twice as high at 75% without inadvertent enforcement compared to 35% for Rounds 6-8 with inadvertent enforcement (two-tailed test of proportion, $p < 0.0001$, $n = 118$). The proportion of war decisions in Stage 1 is much lower at 9% with inadvertent enforcement (Rounds 6-8) compared to 67% without inadvertent enforcement (Rounds 1-5) (two-tailed test of

proportion, $p < 0.0001$, $n = 118$). At the point where the treatment was switched, the introduction of inadvertent enforcement reduced the percentage of war outcomes from 80% (Round 5) to 42% (Round 6) ($n = 27$: two-tailed test of proportion, $p = 0.0404$; Mann-Whitney test, $p = 0.0443$). War outcomes in Stage 1 fell from 80% in Round 5 to 0% in Round 6 ($n = 27$: two-tailed test of proportion, $p < 0.0001$; Mann-Whitney test, $p < 0.0001$).

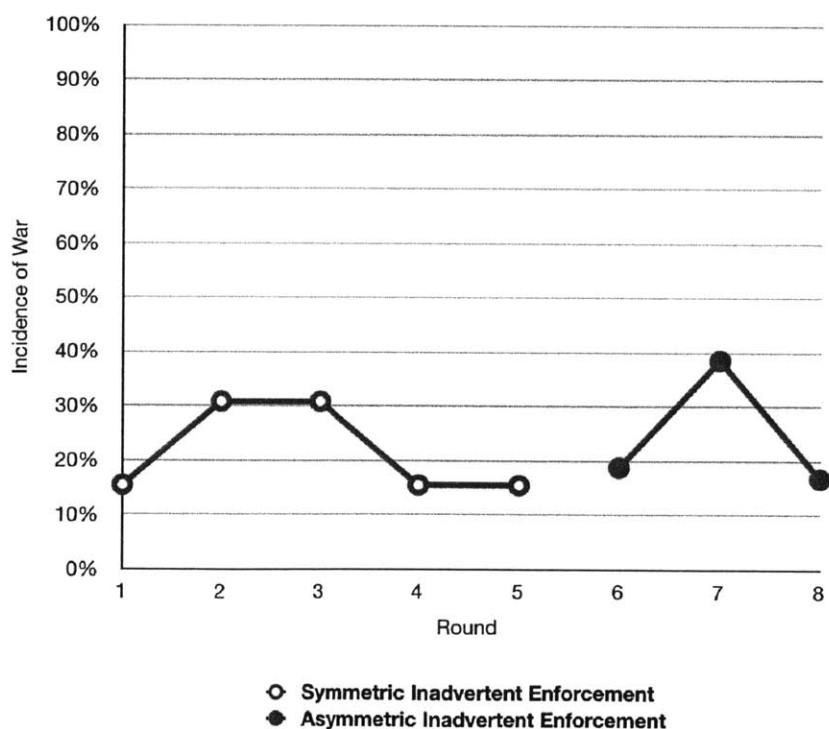
For robustness check, I use logistic regression to estimate the relationship between the asymmetric inadvertent-enforcement treatment and the war outcome. The models use one dyadic observation per round and session based on the control-group sample. Across all model specifications, there is a strong and significant negative relationship between the inadvertent-enforcement treatment and the war outcome ($p \leq 0.001$). Table B4 in Appendix B2 shows the logit estimates in detail.

Result 9: Substituting symmetric and asymmetric variants of inadvertent enforcement makes no significant difference in the incidence of war.

Figure 4.4 shows how the incidence of war evolved from Round 1 to Round 8 for the treatment group, which was introduced to the asymmetric variant of inadvertent enforcement in Round 6. There is no significant change in the incidence of war when the treatment group switched from the symmetric variant of inadvertent enforcement (Round 5) to the asymmetric variant (Round 6) ($n = 29$: two-tailed test of proportion, $p = 0.8114$; Mann-Whitney test, $p = 0.8146$). There is also no significant difference in the

treatment group's incidence of war between Rounds 1-5 (symmetric variant: 22%) and Rounds 6-8 (asymmetric variant: 24%) (two-tailed test of proportion, $p = 0.7326$; $n = 106$). Similar conclusions emerge with the treatment group's Stage-1 war decisions compared between Round 5 (8%) and Round 6 (6%) ($n = 29$: two-tailed test of proportion, $p = 0.8788$; Mann-Whitney test, $p = 0.8809$), or between Rounds 1-5 (11%) and Rounds 6-8 (7%) (two-tailed test of proportion, $p = 0.5537$; $n = 106$). Table B5 (Appendix B2) corroborates Result 9 with logit estimates showing an insignificant relationship between the enforcement-symmetry dummy and the war outcome ($p > 0.60$).

Figure 4.4: Incidence of War in Treatment Group (Rounds 1-8)



Next, I turn to an observation about Player C's behavior:

The third-party player always takes advantage of its rival(s) weakened by war.

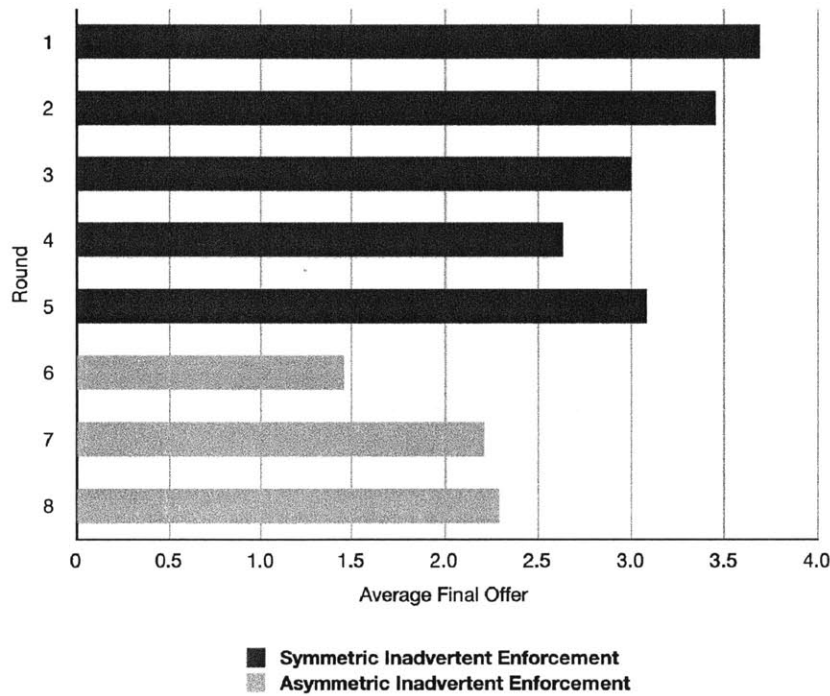
Two players in each round were randomly assigned to the role of Player C. There is no variation in Player C's decision across all rounds in the four experimental sessions: All players assigned as Player C chose to attack ($n = 64$). Note that the existence of C is designed to operate as an inadvertent-enforcement device in the experiment. There is no strategic interaction for Player C: It faces a purely decision-theoretic task in which it will always be better off if it attacks when there is a war between A and B. Hence, the outcome is not surprising, although the perfect convergence to the model prediction in a lab experiment is remarkable.

Result 10: Final offers are lower when inadvertent enforcement is asymmetric.

Figure 4.5 shows the average final offer in Stage 2 in the treatment group from Rounds 1 to 8, with a sudden switch from symmetric to asymmetric inadvertent enforcement in Round 6. The average offer made by Player A suddenly shrunk when the treatment group switched from Round 5 (symmetric enforcement) to Round 6 (asymmetric enforcement) ($n = 25$: two-tailed t-test, $p = 0.0001$; Mann-Whitney test, $p = 0.0003$). Overall, the average final offer size is lower at 2.02 under asymmetric enforcement (Rounds 6-8) compared to 3.19 under symmetric enforcement (Rounds 1-5) (two-tailed

t-test, $p < 0.0001$, $n = 102$). Table B6 in Appendix B2 corroborates Result 10 with OLS estimates showing a strong and significant relationship between the enforcement-symmetry dummy and the final offer size in Stage 2 ($p \leq 0.003$).

Figure 4.5: Average Final Offer in Treatment Group (Rounds 1-8)



Result 10 suggests that Player A is more likely to exploit Player B when inadvertent enforcement is asymmetric – when B has a third-party rival but not A. Result 9 shows, however, that there is no significant difference in the incidence of war despite A’s exploitation of B under asymmetric enforcement. The result is consistent with Prediction 7 in the model. Note, however, that both the model and the experiment are based on an environment with symmetric information and no uncertainty in the payoff parameters.

An interesting extension of both the theory and the experiment is to consider how asymmetric enforcement and potential exploitative behavior are affected by different forms of uncertainty and information asymmetry. Another useful extension is to explore how war incidence and exploitative behavior change when the third party (Player C) is a rival of Player A but not Player B.

1.4 Remarks

The results show that the introduction of inadvertent enforcement causes a sharp fall in the incidence of conflict. Consistent with the model predictions, both the symmetric and asymmetric variants of inadvertent enforcement have strong effects on conflict prevention, with no significant difference between the two. The sharpness of the results indicates the potential efficacy of inadvertent enforcement as a mechanism for shutting down the commitment problem and preventing conflict.

Both the model and experiment have focused on the logic of inadvertent enforcement in its simplest form, without analyzing or testing the effect of various realistic complications to the theory – such as different regime types, different forms of asymmetric information, or different configurations of asymmetric power between N players. The purpose is simply to establish a baseline for future explorations: To construct a baseline model that captures the theoretical essence of inadvertent enforcement, and to identify the causal implications as cleanly as possible. On this basis, various complexities may be introduced and tested in future investigations to expand our knowledge on the topic.

PART 2: ENDOGENOUS ENFORCEMENT

Under endogenous enforcement, at least one of the parties in conflict actively manages the potential shift in power that generates the commitment problem and the risk of war. Here I focus on endogenous enforcement through strategic or tactical restraint. For instance, the rising state can reduce its military spending; it can stop developing sources of potential military advantage; or it can reverse policies that lead to a rapid change in future relative power. The commitment problem and the risk of war are reduced or eliminated if the power shift is reduced or eliminated.

In the experimental literature, the closest works are studies on the arms race versus disarmament dilemma.⁷ A major contribution of this literature is the discovery that the GRIT strategy⁸ (graduated reciprocated initiatives in tension reduction) contributes significantly towards resolving the dilemma. However, while these experiments are relevant, they do not focus on the commitment problem and its resolution.

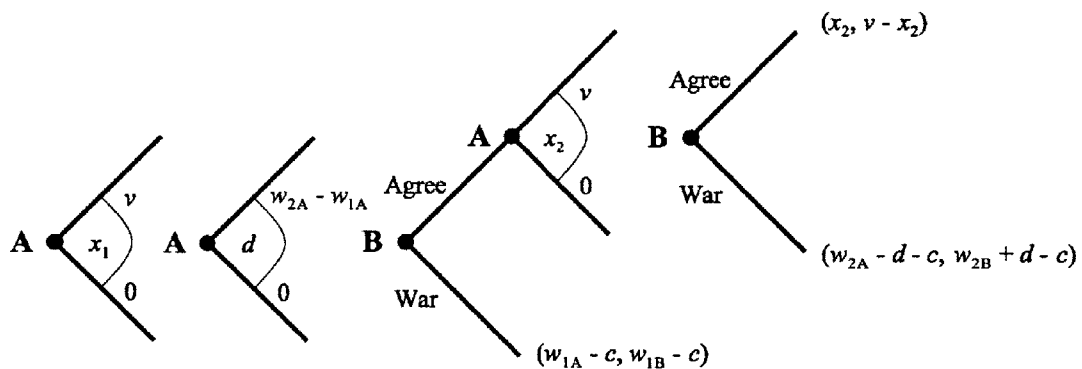
⁷ Marc Pilisuk and Paul Skolnick, "Inducing Trust: A Test of the Osgood Proposal," *Journal of Personality and Social Psychology* 8, No. 2 (1968): 121-33; Svenn Lindskold, Pamela Walters and Helen Koutsourais, "Cooperators, Competitors, and Response to GRIT," *Journal of Conflict Resolution* 27, No. 3 (1983): 521-32; Marc Pilisuk, "Experimenting with the Arms Race," *Journal of Conflict Resolution* 28, No. 2 (1984): 296-315; Svenn Lindskold and Gyuseog Han, "GRIT as a Foundation for Integrative Bargaining," *Personality and Social Psychology Bulletin* 14, No. 2 (1988): 335-45; Brian Betz, "Response to Strategy and Communication in an Arms Race-Disarmament Dilemma," *Journal of Conflict Resolution* 35, No. 4 (1991): 678-90.

⁸ Charles Osgood, *An Alternative to War or Surrender* (University of Illinois Press, 1962).

2.1 Model

I construct a model of endogenous enforcement based on a refinement of the two-stage commitment-problem game described in Chapter 2. The refinement is based on an additional decision node for Player A in Stage 1. A makes a demand $x_1 \in [0, v]$, and then makes a military reduction in Stage 1 that will reduce its war payoff in Stage 2 by $d \in [0, w_{2A} - w_{1A}]$. B observes x_1 and d , and decides whether to agree or to fight. If B fights, the game ends with payoff $(w_{1A} - c, w_{1B} - c)$, with $w_{1A} = w_{1B}$, where w_{1i} is Player i 's war payoff in Stage 1 and c is the cost of war. If B agrees, the game enters Stage 2 in which A confirms its demand $x_2 \in [0, v]$. B observes x_2 and decides whether to agree or to fight. The game ends with payoff $(x_2, v - x_2)$ if B agrees and $(w_{2A} - c - d, w_{2B} - c + d)$ if B fights, where $v = w_{1A} + w_{1B} = w_{2A} + w_{2B}$ and $w_{2A} > w_{1A} > w_{2B} > c$. Figure 4.6 shows the game tree.

Figure 4.6: Crisis Bargaining Game with Endogenous Enforcement



Note that the size of d determines the sharpness of the commitment problem for Player B. If A chooses $d = 0$ in Stage 1, the commitment problem remains unchanged, and the model reduces into the commitment-problem game in Chapter 2. In contrast, the commitment problem is shut down if A chooses $d = w_{2A} - w_{1A}$. This is because if $d = w_{2A} - w_{1A}$, the war payoffs in Stage 2 will be equalized with $w_{2A} - c - d = w_{2B} - c + d$, since $w_{1A} = w_{1B}$ and $w_{1A} + w_{1B} = w_{2A} + w_{2B}$ by definition. This translates into equal bargaining power between A and B in *both* Stage 1 and Stage 2.

In the experiment, $v = 10$, $c = 2$, $w_{1A} = w_{1B} = 5$, $w_{2A} = 7$, and $w_{2B} = 3$. The parameters are selected based on the reasons discussed in Chapter 2. For simplicity, the variable d is determined by a binary choice between 0 and $w_{2A} - w_{1A}$.

The model predictions for the experimental outcomes depend on Player A's choice of d . If A chooses $d = 0$, endogenous enforcement does not exist. There is an incentive for A to renege in Stage 2. Hence, B will disregard x_1 and only consider x_2 in Stage 1. By backward induction, A will demand $x_2 = v - w_{2B} + c$ in Stage 2, based on the amount that makes B just willing to accept. B will choose war in Stage 1 since $w_{1B} - c > v - x_2$ in Stage 2. Hence:

Prediction 8: War will occur with certainty in the commitment-problem game without endogenous enforcement.

Conversely, if A chooses $d = w_{2A} - w_{1A}$, endogenous enforcement is achieved and there is no incentive for A to renege in Stage 2. The equilibrium outcome converges to the same equilibrium described in Chapter 2 for the commitment-problem game under perfect enforcement and public information. Hence:

Prediction 9: War will be avoided in the commitment-problem game when endogenous enforcement is achieved.

2.2 Experimental Design and Implementation

Experiment 2 was time-shared with Experiment 1. Hence, it had the same experimental setting and logistic setup (see Section 1.2). Experiment 2 was conducted over three rounds after Experiment 1. Subjects knew that each round would be independent from previous rounds, and that they would be randomly re-matched after every round.

Like Experiment 1, subjects were placed in the role of a national decision-maker bargaining with another country for a valuable prize. They began with a scenario screen that highlighted two important differences compared to Experiment 1: (1) The relative power between Countries A and B in Stage 2 depends on whether A cuts its military investment in Stage 1; (2) there are only two players in the game. Thereafter, the game was carefully explained in detail. In Stage 1, both countries are equally powerful. If they fight a war, each country will seize 50% of the prize (5 points) for itself. In Stage 2, their relative power depends on whether A cuts its military investment in Stage 1. If A cuts its

investment, A and B remain equally powerful in Stage 2. Hence, each can seize 50% of the prize in a war. If A does not cut its military investment in Stage 1, A becomes more powerful in Stage 2. If they fight a war, A will seize 70% of the prize (7 points), leaving B with 30% (3 points) of the prize. The cost of war is constant at 2 points.

The instructions were set in neutral language with test questions to check the subjects' understanding of the game. After the test questions, there were three additional safeguards to ensure that participants fully understood the game.⁹ Appendix A2 reproduces the experimental protocols and instructions.

Experiment 2 investigates if the incidence of conflict is reduced when endogenous enforcement is achieved. Endogenous enforcement is *endogenously* generated by the players. Whether or not it is achieved cannot be randomly assigned exogenously.¹⁰ Note, however, that the players were randomly assigned to the role of Player A or B, and randomly matched with each other. As a consequence, all subjects had an equal chance of generating and encountering endogenous enforcement.

Does endogenous enforcement occur frequently or is it a rare phenomenon? When endogenous enforcement is achieved, does the incidence of conflict drop significantly?

How weak or strong is the effect?

⁹ In the screen after the test questions, participants saw the answers, which showed what the participant got right or wrong. Next, participants saw a screen that summarized the game and payoffs. Finally, the information appeared again on the decision screen itself.

¹⁰ It is possible to randomly assign different *opportunities* for endogenous enforcement. This is a related but different research question that can be investigated in future work.

2.3 Experimental Results

Result 11: Endogenous enforcement is achieved in half of the cases.

Endogenous enforcement requires a player to forego a future power shift that benefits him or her. Hence, one might suspect that endogenous enforcement would only occur rarely since players might resist giving up their future advantage. It turns out, however, that 51% of the players decided to make military cuts in Stage 1. This decision eliminated their military advantage in Stage 2 and shut down the commitment problem through endogenous enforcement. Table 4.3 shows the incidence of endogenous enforcement across three rounds of the experiment.

Table 4.3: Incidence of Endogenous Enforcement

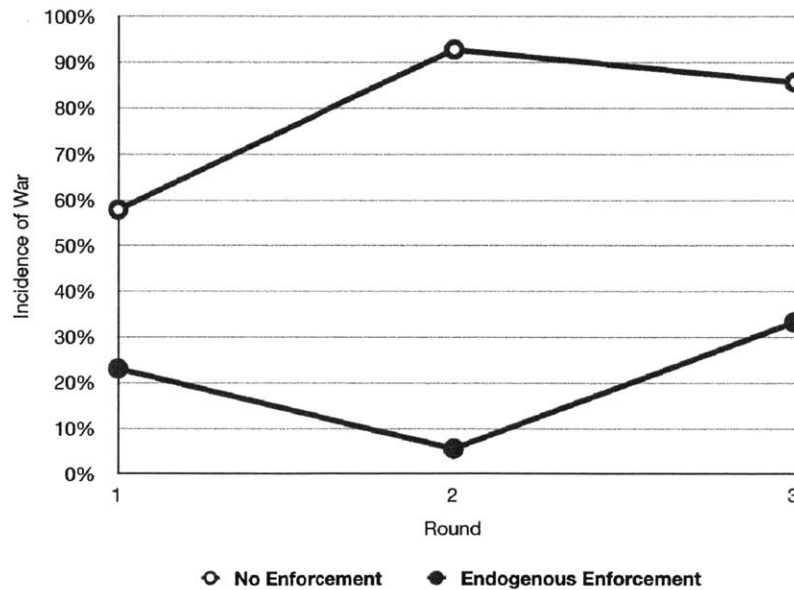
	Round 1	Round 2	Round 3	Total
Session 1	0.29	0.57	0.57	0.48
Session 2	0.50	0.75	0.75	0.67
Session 3	0.25	0.38	0.50	0.38
Session 4	0.56	0.56	0.44	0.52
Total	0.41	0.56	0.56	0.51

Notes: Maximum incidence is 1.00 (100%). Total of 64 subjects in four sessions: Session 1 ($n = 14$), Session 2 ($n = 16$), Session 3 ($n = 16$) and Session 4 ($n = 18$). Each round per session had 7 to 9 subjects randomly assigned as Player A.

Result 12: Endogenous enforcement significantly reduces the incidence of war.

Figure 4.7 shows the incidence of war with and without endogenous enforcement across three rounds of the experiment. The overall incidence of war is 77% without endogenous enforcement compared to 20% with endogenous enforcement (two-tailed test of proportion, $p < 0.0001$, $n = 96$). The incidence of war in the first round is 58% without endogenous enforcement and 23% with endogenous enforcement (two-tailed test of proportion, $p = 0.0512$, $n = 32$).

Figure 4.7: Incidence of War (With and Without Endogenous Enforcement)



**Table 4.4: Incidence of War Outcomes
(With and Without Endogenous Enforcement)**

Session	Without Endogenous Enforcement	With Endogenous Enforcement
1	0.91	0.10
2	0.88	0.38
3	0.73	0.11
4	0.62	0.14
Overall	0.77	0.20

Notes: Maximum incidence is 1.00 (100%). Total of 64 subjects in four sessions: Session 1 ($n = 14$), Session 2 ($n = 16$), Session 3 ($n = 16$), Session 4 ($n = 18$). Total of 96 dyadic observations (32 dyads x 3 rounds).

I use logit models to estimate the relationship between endogenous enforcement and the war outcome, using model specifications with and without controls for individual risk preferences and the size of Player A's initial offer. Each model uses one dyadic observation per round and session, with robust standard errors corrected for clustering at the subject level. Across all model specifications, endogenous enforcement has a strongly significant negative relationship with the war outcome ($p < 0.001$). The coefficient for the size of the initial offer is negative but does not reach statistical significance ($p > 0.05$). The risk-preference variable is insignificant at $p > 0.40$. Table B7 in Appendix B2 shows the logit estimates.

The incidence of war decisions in Stage 1 is much lower at 12% with endogenous enforcement compared to 70% without endogenous enforcement (two-tailed test of proportion, $p < 0.0001$, $n = 96$). In the first round of the experiment, the incidence of war decisions in Stage 1 is 8% with endogenous enforcement but 53% without

endogenous enforcement (two-tailed test of proportion, $p = 0.0086$, $n = 32$). Table B8 (Appendix B2) shows logit estimates for the relationship between endogenous enforcement and the decision for war in Stage 1. The results are consistent with those in Table B7, with a strongly significant negative relationship between endogenous enforcement and the decision for war in Stage 1 across all model specifications ($p < 0.001$). Neither the size of the initial offer ($p > 0.10$) nor the risk-preference variable is statistically significant ($p > 0.60$).

I use the control-group data from Experiment 1 to test the robustness of Result 12. In Experiment 1, the control group played a commitment-problem game in the absence of enforcement. If Result 12 is robust, we should expect a significant difference in war incidence between the control condition in Experiment 1 and the endogenous-enforcement condition in Experiment 2. This is indeed the case. The incidence of war is 20% with endogenous enforcement compared to 75% in the control group from Experiment 1 (two-tailed test of proportion, $p < 0.0001$, $n = 124$). The incidence of war decisions in Stage 1 is 12% with endogenous enforcement compared to 67% in the control group (two-tailed test of proportion, $p < 0.0001$, $n = 124$).

A second robustness test compares the control group in Experiment 1 with the group *without* endogenous enforcement in Experiment 2. Without endogenous enforcement, the game for Experiment 2 is theoretically equivalent to the commitment-problem game for Experiment 1 without enforcement (control condition). Hence, if Result 12 is robust, we should expect no significant difference in war incidence between the no-enforcement

condition in Experiment 1 and the no-enforcement condition in Experiment 2. This is exactly the case:

Result 13: There is no significant difference in the incidence of war between the no-enforcement condition in Experiment 1 and when endogenous enforcement is not achieved in Experiment 2.

The incidence of war is 77% in the no-enforcement condition in Experiment 2 compared to 75% in no-enforcement condition in Experiment 1 (two-tailed test of proportion, $p = 0.8097$, $n = 122$). The incidence of war decisions in Stage 1 is 70% in the former compared to 67% in the latter (two-tailed test of proportion, $p = 0.6827$, $n = 122$). This result shows that when endogenous enforcement is *not* achieved, the game for Experiment 2 generates a similar outcome as the commitment-problem game for Experiment 1 where enforcement is absent.

2.4 Remarks

The results suggest that endogenous enforcement is not a rare phenomenon. Half of subjects randomly assigned to the role of Player A were willing to give up their future military advantage to eliminate the potential power shift. By eliminating their power advantage in the future, they also shut down the commitment problem in the present. When endogenous enforcement is achieved, the incidence of conflict drops dramatically.

Experiment 2 provides some first-cut evidence on the effects of endogenous enforcement, based on the simplifying assumption of endogenous enforcement as a binary variable. Future experiments may test for variations in the treatment effect with endogenous enforcement as a continuous variable.

PART 3: COSTLY SIGNALING

For theoretical and policy-making purposes, it is crucial to understand which types of signals are credible and which are not. In international relations, this problem is understood primarily based on the theory of costly signaling, in a context where the signaler and receiver have divergent interests. To be credible, the signal should differentiate between resolved and unresolved states by carrying some costs that would discourage unresolved states from sending that signal. Fearon suggested that states have two basic ways to signal their resolve.¹¹ One of the two mechanisms is *sunk-cost signaling*, which creates direct costs that cannot be recovered, and which does not affect the relative value of escalation or compromise.¹² In this chapter, the generic term “costly signal” refers specifically to a signal with sunk cost.¹³

The theoretical logic of sunk-cost signaling is well known. But there is little empirical evidence on whether the mechanism operates in strategic interactions between states. Do resolved states use sunk costs to signal their resolve? Do sunk costs make a threat more credible? These questions are difficult to address with observational data. There are few pure cases of sunk-cost signals in international politics.¹⁴ It is hard to observe the

¹¹ James Fearon, “Signaling Foreign Policy Interests: Tying Hands Versus Sinking Costs,” *Journal of Conflict Resolution* 41, No. 1 (1997): 68-90.

¹² The second mechanism is “tied-hands signaling,” which binds the signaler to a higher cost of backing down if the opponent does not back down, but is costless to the signaler if the opponent backs down. Fearon, “Signaling Foreign Policy Interests.”

¹³ I use a strict definition of costly signaling to include only sunk-cost signaling. See Chapter 1.

¹⁴ Fearon highlighted that the effects of sunk-cost and tied-hands signals are often mixed together in real-world cases, but “it is important to see ... that two distinct mechanisms are at work, and we need to analyze them separately as ideal types to understand the strategic logic of mixed cases.” Fearon, “Signaling Foreign Policy Interests,” 70.

signaler's true resolve at the point of signaling or the receiver's credibility estimates at the point of signal reception. The most intractable problem is that the credibility effect of a sunk-cost signal is almost always confounded with the effects of other current or previous signals or pieces of information – at least some of which are completely opaque to the researcher. In short, it is extremely difficult to control and analyze the information environment in a real-world setting.

I investigate the mechanism of sunk-cost signaling in its unconfounded form with a series of experiments. In Experiments 3 and 4, subjects are randomly divided into signalers and receivers interested in a valuable prize. Both experiments are incentivized based on a signaling game with a separating equilibrium. The signaler gets a higher payoff in the game if the receiver acquiesces. The receiver who challenges the signaler gets a payoff conditional on the signaler's resolve: a high payoff if the signaler has low resolve, and a low payoff if the signaler has high resolve. The signalers are randomly assigned with high and low private valuations of the prize that generate high and low levels of resolve. I test whether signalers randomly assigned with high resolve are more likely to use a signal with sunk cost. I also test whether receivers that see a signal with sunk cost are less likely to challenge the signaler. Thereafter I measure signal credibility by eliciting the receiver's credibility estimates.

Experiments on signaling games are rare in international relations, but relevant experiments can be found in the economics literature that tests different equilibrium

selection devices in signaling games.¹⁵ The closest examples are by Miller and Plott with signaling games set in experimental markets with buyers and sellers, and by Potters and Van Winden with a basic signaling game that relates to lobbying and advertising.¹⁶ The experiments reported here, however, differ from the economics literature in various ways. First, the signaling game and experiment are structured and framed to relate directly to the international relations literature on crisis interaction. The game captures a deterrence crisis with two parties contending for a valuable prize; the signaler issuing a threat to the receiver; and the receiver deciding whether to confront the signaler. Unlike signaling experiments in which the receiver either receives a message or doesn't, the receiver in this experiment *always* receive a threat from the signaler with the same content: the only difference is whether the threat carries a sunk cost.

¹⁵ See Jordi Brandts and Charles Holt, "An Experimental Test of Equilibrium Dominance in Signaling Games," *American Economic Review* 82, No. 5 (1992): 1350-65; Jeffrey Banks, Colin Camerer, and David Porter, "An Experimental Analysis of Nash Refinements in Signaling Games," *Games and Economic Behavior* 6, No. 1 (1994): 1-31; David Cooper, Susan Garvin, and John Kagel, "Signaling and Adaptive Learning in an Entry Limit Pricing Game," *RAND Journal of Economics* 28, No. 4 (1997): 662-83; David Cooper and John Kagel, "Are Two Heads Better than One? Team versus Individual Play in Signaling Games," *American Economic Review* 95, No. 3 (2005): 477-509. In international relations, Tingley and Walter's cheap-talk experiment showed that when signalers and receivers are given a private channel of costless communication, it deters the receivers in early periods of play and makes the signalers more eager to fight, compared to the control group without any communication between signaler and receiver. Dustin Tingley and Barbara Walter, "Can Cheap Talk Deter? An Experimental Analysis," *Journal of Conflict Resolution* 55, No. 6 (2011): 996-1020.

¹⁶ Miller and Plott constructed experimental markets where sellers could choose to add product enhancements that differed in cost based on the quality of their product. The study found that in general, high-quality sellers are more likely to opt for more costly signals. Specifically, markets with relatively low marginal costs of signaling a high-quality product separated at least in the final periods, but not markets with relatively high signaling costs. The general finding is replicated in Potters and Van Winden. In the experiment, the signaler chooses between a costly message and no message based on five different treatments representing different parameter configurations, and twenty periods of play with the signalers and receivers reversing roles in the last ten periods. The study found that receivers respond more positively to a costly message compared to no message at all, with the statistical results reported based on the last ten periods of play. The game has a pooling equilibrium and a semi-pooling equilibrium, with no separating equilibrium. See Ross Miller and Charles Plott, "Product Quality Signaling in Experimental Markets," *Econometrica* 53, No. 4 (1985): 837-72; Jan Potters and Frans Van Winden, "Comparative Statics of a Signaling Game: An Experimental Study," *International Journal of Game Theory* 25, No. 3 (1996): 329-53.

Second, the experiments are designed to provide a sharp test of the mechanism of sunk-cost signaling. As will be described later, the relative payoffs of the two signaler types are configured to make type-separation possible and the type-separation logic as transparent as possible. I also black-box the signaler's reaction to the receiver's choice: By definition – and as reflected in the payoffs – the resolved (or unresolved) type will always (or never) fulfill its threat to fight the receiver if the receiver does not acquiesce. This design strategy ensures that the receiver's focus in the game falls sharply on the signaler type and the signal sent, with absolute uncertainty on how a given type of signaler would react.

Finally, I measure signal credibility not only based on the receiver's action, but also based on a measurement of signal credibility as perceived and reported by the receiver. I also use a survey experiment to elicit signal credibility estimations from a diverse national sample.

I use a variety of design strategies to test the robustness of my results. First, I replicate the experiment in two different environments: over the Internet (Experiment 3) and in the laboratory (Experiment 4). Second, I construct in Experiment 3 an additional experimental group in which receivers have no information about the payoffs for the two types of signalers. This makes type-separation reasoning impossible for receivers, and it allows me to test whether the credibility of sunk-cost signaling is conditional on type-separation reasoning. Finally, I implement a survey experiment (Experiment 5), which contextualizes sunk-cost signaling in an international crisis scenario. This facilitates a

straightforward elicitation of credibility estimates without any strategic interaction. While there is variation in the point estimates, the main conclusions are generally consistent across the different designs.

Experimental replication in different environments and with different subject samples allows us to see if the findings are generalizable beyond a particular subject pool. Experiment 5 also provides contextualization based on an international crisis scenario to enhance the realism of the experimental context. Of course, I cannot eliminate the possibility that national leaders in the same setting may respond to sunk-cost signaling in a different way. It is also possible that leaders with different predispositions may respond differently to sunk-cost signaling in different issue contexts. However, the experiments help to shed light on whether and how the mechanism of sunk-cost signaling operates under ideal *ceteris-paribus* conditions. The results provide an empirical baseline for our theoretical understanding, as well as for future investigations of the mechanism with other subject samples or in different contexts.

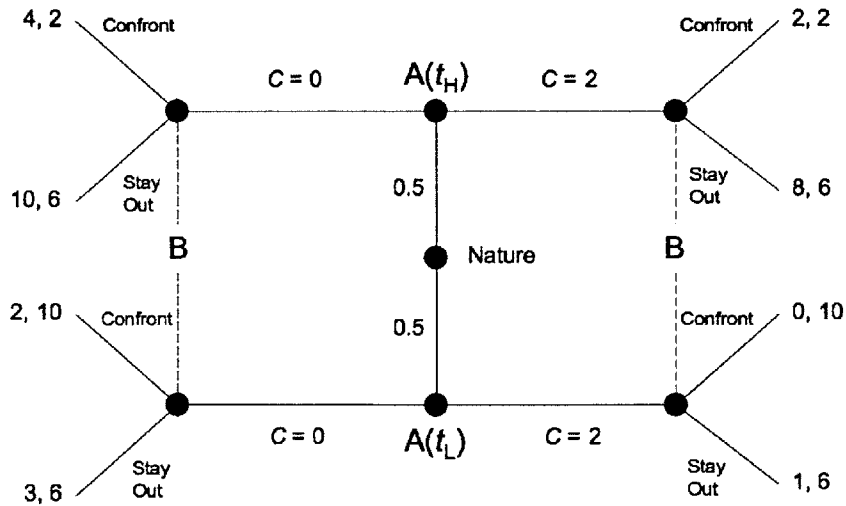
3.1 Signaling Game

I construct a signaling game that captures the general mechanism of sunk-cost signaling in a pure and simple form. The game sets up a deterrence crisis between two players, A and B, who are interested in a valuable prize. A sends a threat to B that it will fight B if B does not stay out. There are two types of Player A: true type and fake type. By definition,

a true type will *always* fulfill its threat while the fake type will *never* fulfill the threat. Hence, the true type is high-resolve and the fake type is low-resolve.

The game is straightforward. Nature randomly assigns A as either a high-resolve type or low-resolve type unknown to B. A chooses either a threat with a sunk cost ($c = 2$) or a threat without a sunk cost ($c = 0$). B observes the threat and decides whether to confront A or to stay out. If B stays out, the payoff is $(10 - c, 6)$ if A is a high-resolve type, and $(3 - c, 6)$ if A is a low-resolve type. If B confronts, the payoff is $(4 - c, 2)$ if A is a high-resolve type (i.e. A will fulfill its threat and fight B), and $(2 - c, 10)$ if A is a low-resolve type (i.e. A will not fulfill its threat and will not fight B). Figure 4.8 diagrams the game.

Figure 4.8: Sunk-Cost Signaling Game



The payoff parameters are chosen with the following considerations. First, the payoffs are constrained within the integer range in $[0, 10]$. The experiment operates within eleven whole numbers to keep the calculations as straightforward as possible for the subjects. Second, B's payoffs are constrained such that an equal expected value (6) between the two choices (confront or stay out) is assured, given a 0.5 ex-ante probability that A is a high-resolve type. B's maximum payoff is the same as the maximum payoff for a high-resolve-type A; B's minimum payoff is the same as the minimum payoff for a low-resolve-type A, disregarding sunk cost which is a voluntary choice made by A. Finally, A's payoffs are constructed to ensure that a threat with sunk cost is type separating, subject to the first and second constraints. Note that the threat with sunk cost is strictly dominated by the threat without sunk cost for Player A of the low-resolve type: the low-resolve type gets *at most* 1 if it sends the costly threat, but *at least* 2 if it sends the costless threat. Hence, a low-resolve type should never send the costly threat. This implies that a high-resolve type can effectively separate itself from a low-resolve type by sending the costly threat.

The game has a separating perfect Bayesian equilibrium as follows: A will send the threat with sunk cost if it is a high-resolve type and the threat without sunk cost if it is a low-resolve type. B will stay out if it receives the costly threat and will confront if it receives the costless threat. In B's belief, given that the costless threat is observed, the probability that A is a high-resolve type is 0 while the probability that A is a low-resolve type is 1.

Prediction 10: The signaler will send the threat with sunk cost if it is a high-resolve type and the threat without sunk cost if it is a low-resolve type.

Prediction 11: The receiver will acquiesce if it receives the threat with sunk cost and will confront if it receives the threat without sunk cost.

Proof. Since the costly threat ($c = 2$) is strictly dominated by the costless threat ($c = 0$) for the low-resolve type (t_L), t_L will never play $c = 2$. If a separating equilibrium exists, it will require the high-resolve type (t_H) to play $c = 2$ and t_L to play $c = 0$. B's beliefs $\mu(\cdot)$ on A's type are derived from Bayes' rule. Given $c = 0$, B's best response is *Confront*, as the expected utility $\mu(t_L | c = 0)(10) + \mu(t_H | c = 0)(2)$ exceeds $\mu(t_L | c = 0)(6) + \mu(t_H | c = 0)(6)$, the expected utility for *Stay Out*. Given $c = 2$, $\mu(t_L | c = 2)(10) + \mu(t_H | c = 2)(2) < \mu(t_L | c = 2)(6) + \mu(t_H | c = 2)(6)$, and B's best response is *Stay Out*. In equilibrium, t_L will not deviate from $c = 0$, since for her $c = 0$ strictly dominates $c = 2$. t_H 's payoff along the equilibrium path is $U_A(c = 2; \textit{Stay Out}; t_H)$. If t_H deviates from $c = 2$ and plays $c = 0$, B's belief remains as before and $U_A(c = 0; \textit{Confront}; t_H) < U_A(c = 2; \textit{Stay Out}; t_H)$. Hence, t_H has no incentive to deviate. A separating equilibrium exists as described. ■

3.2 Experimental Design and Implementation

Experiment 3 implements the signaling game over the Internet. Experiment 4 replicates the signaling game in a laboratory setting. Experiment 5 is embedded in an online survey. Table 4.5 summarizes the differences between the three experiments. The next three subsections describe the design and implementation procedures for each experiment.

Table 4.5: Summary of Experiments

	Experiment 3	Experiment 4	Experiment 5
Type	Internet-based game	Laboratory game	Internet-based survey
Activity	Signaling game	Signaling game	Credibility rating
Incentivized	Yes	Yes	No
Sample	U.S. adult residents recruited through AMT	MIT students recruited through lab	U.S. adult residents recruited through AMT
Rounds	One-shot	Three rounds	One-shot
Subjects	339	64	635

Experiment 3 (Internet-Based)

375 U.S. adult residents were recruited on 21 and 22 February 2013 through Amazon.com's Mechanical Turk (AMT). The validity of AMT as an experimental tool has been carefully tested across different fields in social science, including economics

and political science.¹⁷ Participants in Experiment 3 were linked from AMT to the online platform where the experiment was hosted. The experiment was programmed and conducted on the Qualtrics platform. Each participant received \$0.51 as participation fee and a bonus payment between \$0.00 and \$1.00 that varied based on the game outcome. Participants played the sunk-cost signaling game at the start of the session. Thereafter, they transitioned into a bargaining game followed by a risk-aversion game that measured their risk preferences. The session ended with a set of demographic questions.¹⁸

Participants were told that they would play three different games. The computer would randomly choose one out of the three games they played, and count their point earnings in that game as bonus payment. Each game had a total possible value of 10 points, with each point equivalent to \$0.10 in bonus. The average total earning was \$1.02.

¹⁷ See John Horton, David Rand, and Richard Zeckhauser, "The Online Laboratory: Conducting Experiments in a Real Labor Market," *Experimental Economics* 14, No. 3 (2011): 399-425; and Adam Berinsky, Gregory Huber, and Gabriel Lenz, "Evaluating Online Labor Markets for Experimental Research: Amazon.Com's Mechanical Turk," *Political Analysis* 20, No. 3 (2012): 351-68. See also Gregory Huber, Seth Hill, and Gabriel Lenz, "Sources of Bias in Retrospective Decision Making: Experimental Evidence on Voters' Limitations in Controlling Incumbents," *American Political Science Review* 106, No. 4 (2012): 723-4, for a succinct defense of AMT as a recruitment tool for experimental studies; and Adam Berinsky, Kai Quek, and Michael Sances, "Conducting Online Experiments on Mechanical Turk," *The Experimental Political Scientist* 3, No. 1 (2012): 2-6, for an overview of AMT. AMT is a service by Amazon.com linking employers to online workers. Berinsky, Huber and Lenz showed that the AMT sample is demographically similar to the sample used in the unweighted American National Election 2008-09 Panel Study (ANESP). The AMT sample is slightly more Democratic, more liberal and more knowledgeable about politics. The potential problems of habitual participations and repeated participation are negligible. The study concluded that "if we treat the [AMT] as a means for conducting internally valid experiments, instead of a representative sample, the [AMT] respondent pool is very attractive." Berinsky, Huber and Lenz, "Evaluating Online Labor Markets," 361.

¹⁸ The average time to completion is 17 minutes. This is likely to be an overestimate, as the Qualtrics timer began once the respondent was linked from AMT to Qualtrics, but a small number of respondents might not have immediately started and finished the session. There were ten observations in the dataset with a timed duration of 40 minutes or more, of which five were timed at 50 minutes or more.

In the sunk-cost signaling game, participants were randomly divided into three groups: one group of signalers (Player A) and two groups of receivers (Player B). Those assigned as Player A were randomly divided into true types and fake types. All groups received the same instructions, except that the second group of Player B was given no information about the payoffs for the two types of signalers. Hence, type-separation reasoning was deliberately made impossible¹⁹ for the second group of Player B. This allows us to test if the credibility of sunk-cost signaling is conditional on type-separation reasoning.

The instructions highlighted that Player A knows whether it is a true type or a fake type, but A's type is unknown to Player B. B gets a high payoff if it confronts a fake type but a low payoff if it confronts a true type. The signaling game was explained to participants in detail, with questions at the end to test their understanding. Participants also saw a summary of the game and payoffs twice: on the screen just before the decision screen, and on the decision screen itself. Thereafter, the participants made their decisions. Player A (the signaler) would decide to send either a threat with sunk cost ("Threat X with cost = 0 points") or a threat without sunk cost ("Threat Y with cost = 2 points"). Player B (the receiver) would observe the signal and decide whether to confront or to stay out. Player B would also be asked to assess the likelihood that it had encountered a true-type Player A. Appendix A3 reproduces the full experimental instructions.

To generate the game outcome, each participant's decision was randomly matched with the decision of another participant in the opponent role. Opponents were randomly drawn

¹⁹ Zeinab Partow and Andrew Schotter, "Does Game Theory Predict Well for the Wrong Reasons: An Experimental Investigation," Working Paper (Department of Economics, New York University, 1993).

from participants who had already played the game in the opponent role. I collected the pool of opponent decisions by implementing the same game in a pre-experiment three weeks earlier on AMT with 73 participants. These decisions were programmed into the actual experiment to generate the game outcomes for payment purposes, and the different frequencies of costly signals conditional on the opponent's signaler type. Participants in the pre-experiment were excluded from the actual experiment.²⁰ Section 3.3 reports the relevant details.

To ensure that respondents paid attention and understood the signaling game, respondents were tested with four questions on how the game worked.²¹ Two questions were straightforward. The other two questions were designed to be hard tests: they were more complicated and could not be easily answered without careful thought. Aside from the test questions, which also served a training purpose, there were three additional safeguards to ensure that respondents understood how the game worked. First, in the screen immediately after the test questions, respondents saw the answers to the questions, which showed what the respondent got right or wrong. Next, respondents saw a screen that summarized the game and the payoffs.²² Finally, the information also appeared on the decision screen.

²⁰ The recruitment notice for the actual experiment prohibited repeat participants. Each AMT subject has a unique "Worker ID," which allows me to trace repeat participants and exclude them from the dataset.

²¹ The fourth question was omitted for the second experimental group of Player B, who had received no information about the payoffs for the two signaler types. See Appendix A3.

²² The summary screen was excluded in the first session of Experiment 4 ($n = 14$). The conclusions from Experiment 4 remain unchanged when Session 1 data is excluded from the statistical analysis.

There is a potential concern that Internet respondents may not pay sufficient attention to the experimental setting. I dealt with the concern in five ways. First, I recruited subjects from AMT, who are known to be more attentive than the average Internet respondent.²³ Second, the monetary bonuses were performance-contingent, providing real incentive for respondents to play and “win” the game. Third, recruitment was restricted to those with a minimum 95% approval rate for prior AMT tasks. The 95% threshold is higher than the 90% standard usually used in AMT experiments, and it prevents respondents without an excellent work record from participating in the game. Fourth, the data analysis excludes all subjects who answered more than one test question incorrectly. The final sample has a total of 339 subjects after excluding 36 subjects. Finally, there were three additional safeguards for respondents to revise and confirm their understanding of the game, as described earlier.

Experiment 4 (Laboratory-Based)

Experiment 4 was time-shared with Experiments 1 and 2. It was implemented in the same sessions with the same experimental setting and logistic setup (see Section 1.2). Participants played eleven rounds under Experiments 1-2 before transiting into the sunk-cost signaling game. After playing three rounds of the signaling game, participants played a risk-aversion game in the final round that measured their risk preferences.²⁴

²³ Berinsky, Huber and Lenz, “Evaluating Online Labor Markets.”

²⁴ The potential for spillover effects should be limited as the signaling game is distinctly different from Experiments 1-2, and subjects know that each round is independent from previous rounds. Subjects know that they will be randomly matched with different opponents in the signaling game.

The rules and instructions in the signaling game were largely similar in Experiments 3 and 4 (see Appendix A3), with three basic differences. First, Experiment 4 did not construct an additional experimental group of Player B for which type-separation is impossible. I did not replicate the type-separation test in the laboratory, given the smaller sample size in Experiment 4 compared to Experiment 3. Second, Experiment 4 used a one-stage elicitation of credibility estimates, as two-stage elicitation would be more cumbersome and less effective with repeated rounds.²⁵ Finally, subjects in Experiment 4 played three rounds of the signaling game instead of a single round. This allows us to collect more observations without deviating too far from a one-shot game setting. Subjects played the first round without knowing if the second or third round would be similar to the first. In each round, subjects were randomly assigned as either Player A or Player B, and randomly matched with one another. To minimize learning effects, subjects were *not* told the game outcomes – which would leak information about the receiver’s choice and the signaler’s type – until the end of the entire experiment.

²⁵ One-stage elicitation presents Player B (the receiver) with the full menu of options when they were asked if they thought that their opponent (the signaler) was likely to be a true type: “Very likely”, “Somewhat likely”, “More likely than unlikely”, “Neither likely nor unlikely”, “More unlikely than likely”, “Somewhat unlikely”, and “Very unlikely”. Appendix 4.2 describes the two-stage elicitation process in Experiment 3. Note that the language used in the two elicitation procedures is comparable but not exactly the same. It seems unlikely that the difference in the retrospective credibility ratings between the two experiments is a mere artifact of the elicitation procedures; the difference is more likely due to differences in the subject sample and experimental setting. It would have been cleaner to use a one-stage elicitation procedure in both experiments. This was not done, as Experiment 4 was a replication test that occurred only after Experiment 3 was implemented. Future experiments on credibility with multiple replications may aim for exact similarity in the elicitation procedures.

Experiment 5 (Survey-Based)

Experiment 5 is a survey experiment that contextualizes sunk-cost signaling in an international crisis scenario. The experiment elicits credibility estimates from respondents without any strategic interaction. Unlike Experiments 3 and 4, there is no signaling game and respondents are not assigned into a specific role. Instead, they are presented with a scenario and asked to assess the credibility of a threat made by a country in the scenario. Experiment 5 was embedded in a time-shared survey conducted on the Qualtrics platform. 1,057 U.S. adult residents were recruited from 6 to 11 April 2012 through AMT, of which 635 subjects were randomly allocated to Experiment 5.²⁶ Each participant received \$0.51 for completing the survey.

To construct the sunk-cost signal in its unconfounded form, the signal is deliberately decontextualized and disassociated from specific real-world examples in Experiments 3 and 4. It is possible, however, that credibility estimates change when we move from a pure and abstract “sunk-cost signal” to an impure but concrete real-world example. Experiment 5 is designed as a robustness test with a concrete example of a sunk-cost signal (military mobilization) located in a specific international crisis context (territorial dispute). While there are few pure cases of sunk-cost signals in international politics, military mobilization is frequently cited as a classic example of a sunk-cost signal,²⁷

²⁶ The remaining subjects were allocated to Experiments 6 and 7.

²⁷ James Fearon, “Threats to Use Force: Costly Signals and Bargaining in International Crises,” PhD dissertation (University of California, Berkeley, 1992); James Fearon, “Domestic Political Audiences and the Escalation of International Disputes,” *American Political Science Review* 88, No. 3 (1994): 577-92; and Fearon, “Signaling Foreign Policy Interests.”

though it is admittedly not a pure case.²⁸

Respondents began the survey experiment by reading about a foreign crisis scenario, in which two states had staked their claims on an important piece of territory. One of the states (“Country X”) threatened to fight a war if the other state moved into the territory. Respondents were told that Country X had mobilized its military. 210 respondents were randomly divided into the control group and the treatment group. By design, the two groups differed only in one dimension: Respondents in the treatment group were told that military mobilization was “very costly”, while those in the control group were told that military mobilization was “not very costly.” “Very costly” and “not very costly” were used instead of “costly” and “costless”, since the physical deployment of troops cannot plausibly be costless. The dependent variable is a credibility rating on a seven-point scale that measures the respondent’s perceived likelihood that Country X would fulfill its threat.

Military mobilization as a sunk-cost signal is confounded if respondents infer a correlation between the cost of military mobilization and the cost of fighting. To control for this potential confounder, Experiment 5 has a separate component with a 2x2 factorial design using the remainder of the respondent pool ($n = 425$). By design, this component is exactly the same as the earlier component (with $n = 210$) except for an additional sentence: “At the same time, it is clear that fighting a war at this time [will be / will not be] very costly to X.” Respondents were randomly assigned into one of four

²⁸ Fearon, “Signaling Foreign Policy Interests,” 70; Branislav Slantchev, “Military Coercion in Interstate Crises,” *American Political Science Review* 99, No. 4 (2005): 533-47.

experimental conditions that differed along two dimensions: whether military mobilization was “very costly” or “not very costly” to Country X, and whether fighting a war would be “very costly” or “not very costly” to Country X. The two sentences were presented in random order. Interacting the two possibilities across the two dimensions creates the four experimental conditions. This allows me to isolate the effect of costly mobilization on credibility estimates, while explicitly controlling for the effect driven by the cost of fighting. Appendix A4 reproduces the experimental instructions.

3.3 Experimental Results

The results show that signalers randomly assigned with high resolve are much more likely to use a signal with sunk cost. But signals with sunk cost do not have a significant effect on the receiver’s acquiescence rate. There is an unexpected asymmetry between the behaviors of signalers and receivers: Signalers believe that sunk costs make their threats more credible, and hence they choose to suffer sunk costs willingly. But receivers do not necessarily respond in line with the signalers’ belief, despite the sunk costs suffered.

Sending Costly Signals

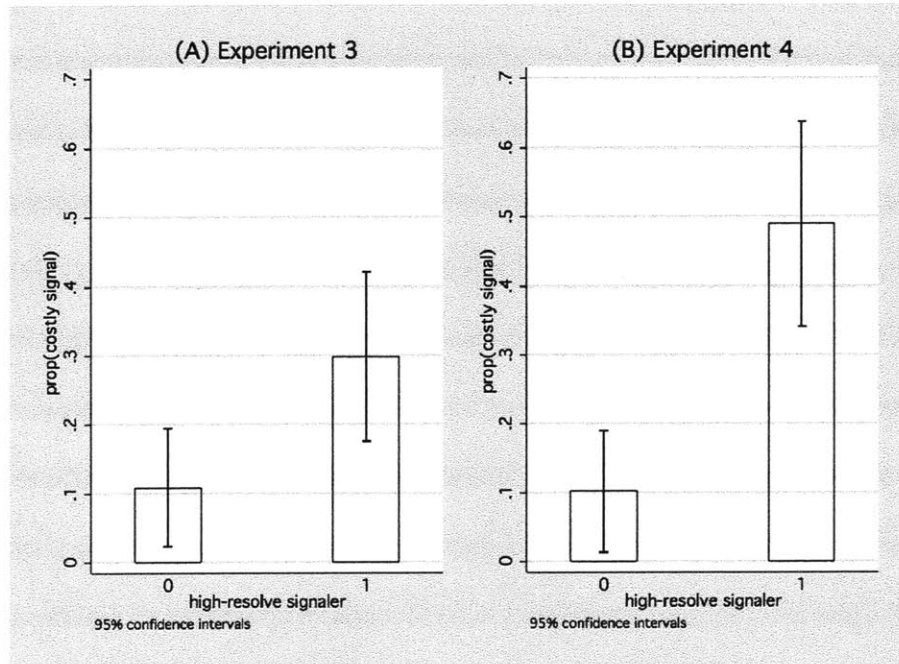
Result 14: Signalers randomly assigned with high resolve are much more likely to use a signal with sunk cost.

In Experiments 3 and 4, signalers are randomly assigned as either a true type or a fake type. By definition, a true type will always fulfill the threat (high resolve) while a fake type will never fulfill the threat (low resolve). Their payoffs reflect their difference in resolve: If Player B (the receiver) stays out, the true type gets $10 - c$ but the fake type gets $3 - c$. If B confronts, the true type gets $4 - c$ but the fake type gets $2 - c$. While A (regardless of type) will always get more if B stays out, a true-type A gets more than a fake-type A. The rest of the chapter will refer to a true type as a high-resolve player and a fake type as a low-resolve player.

Do high-resolve signalers use sunk costs to signal their resolve? Figure 4.9 compares the percentages of high-resolve signalers and low-resolve signalers who sent a costly threat. In Experiment 3, 30% of signalers randomly assigned as high-resolve signalers chose the threat with sunk cost compared to 11% of low-resolve signalers (two-tailed test of proportion, $p = 0.0132$, $n = 112$).²⁹ In total, costly threats made up 21% of all threats sent regardless of type. 74% of costly threats were sent by high-resolve signalers.

²⁹ Unless stated otherwise, the p-values in parentheses in the rest of this chapter are based on a two-tailed test of proportion.

Figure 4.9:
Proportion of Signalers Sending the Costly Threat (Experiments 3 and 4)



In Experiment 4, 49% of high-resolve signalers chose the costly threat compared to 10% of low-resolve signalers ($p < 0.0001$, $n = 96$). 29% of all threats sent (regardless of type) were costly threats, and high-resolve signalers sent 82% of all costly threats. In the first round of the signaling game, 43% of high-resolve signalers sent a costly threat compared to 0% of low-resolve signalers ($p = 0.0021$, $n = 32$). Costly threats accounted for 19% of all threats sent in the first round (regardless of type), all of which were sent by high-resolve signalers.

The evidence shows that high-resolve signalers are much more likely to choose the costly threat than low-resolve signalers. The logit estimates presented in Table 4.6 confirm the result. The logit models analyze the relationship between the level of

resolve and the choice of costly threat. The binary variable, *High-Resolve*, is coded 1 if the signaler was randomly assigned as a high-resolve signaler, and 0 if otherwise. The model specifications include a baseline model without controls and an alternative model that includes a control for individual risk preference (for Experiments 3 and 4) and round and session fixed-effects (for Experiment 4). Risk preference is measured on a summed score based on the risk-aversion game at the end of the experiment (see Appendix A3): the higher the score, the greater the individual willingness to take risk. Table 4.6 shows the logit estimates. The *High-Resolve* binary variable is positive and significant across the different model specifications in both Experiment 3 ($p < 0.020$) and Experiment 4 ($p \leq 0.001$). The risk-preference variable is statistically insignificant ($p > 0.20$) in both experiments.

Table 4.6:
Logit Estimates – Determinants of Costly Threat

	Experiment 3		Experiment 4	
	(1)	(2)	(3)	(4)
High-resolve	1.244 (0.523)*	1.242 (0.529)*	2.138 (0.613)**	2.142 (0.616)**
Risk-preference		0.183 (0.152)		0.012 (0.353)
Constant	-2.100 (0.434)**	-2.630 (0.671)**	-2.676 (0.661)**	-2.668 (1.198)*
Round & session fixed effects	No	No	Yes	Yes
Log-likelihood	-53.688	-52.974	-46.316	-45.928
Prob>Chi ²	0.017	0.044	0.002	0.005
Pseudo-R ²	0.056	0.069	0.201	0.198
N	112	112	96	94

Notes: ** $p \leq 0.01$; * $p \leq 0.05$. In parentheses are robust standard errors, which are corrected for clustering at the subject level for Experiment 4. Round and session dummies are used to control for round and session fixed effects.

Responding to Costly Signals

Result 15: Sunk-cost signals do not have a significant effect on the receiver's acquiescence rate.

Result 16: The sunk-cost signaling mechanism works well at the signaler's end but not at the receiver's end. In line with what the model predicts, high-resolved signalers are much more likely to send sunk-cost signals. But the receiver's acquiescence rate does not respond to the sunk-cost signal, contrary to what the model predicts.

Since receivers were randomly matched with signalers, receivers in the same game had an equal probability of observing a sunk-cost signal. However, there is a difference in how signalers were matched to receivers in Experiments 3 and 4. In Experiment 4, subjects in each session-round were randomly assigned to their roles as signaler or receiver, and randomly paired with one another. In Experiment 3, however, real-time matching was not feasible as subjects entered the Internet-based game on a rolling basis at different timings. Hence, I collected a pool of signaler decisions by implementing the same signaling game three weeks earlier on AMT in a pre-experiment. From the signaler's point of view, there is no substantive difference between the signaling game in the actual experiment and in the pre-experiment.³⁰ The pay rates were the same in both

³⁰ This is not the case for players in the receiver role. In the pre-experiment, receivers made two conditional decisions: whether to confront or to stay out *conditional* on receiving a costly threat, and *conditional* on receiving a costless threat. I collected this data simply to generate outcomes and determine payment (for signalers) in the pre-experiment and in the actual experiment. Note that because

experiments.³¹ 30 signaler decisions (15 from high-resolve signalers and 15 from low-resolve signalers) were collected from the pre-experiment.³² I programmed the decision data into Experiment 3 to generate the game outcomes for receivers, and the different frequencies of costly signals conditional on their opponent's signaler type (high-resolve or low-resolve). Based on the signaler decisions in the pre-experiment, there was a 9/15 chance of receiving a costly threat for receivers randomly matched with a high-resolve signaler, and a 3/15 chance for receivers randomly matched with a low-resolve signaler. Receivers did not know the actual probabilities, except that they had a 50% ex-ante probability of meeting either signaler type.

Do sunk costs make a threat more credible to receivers? First, we look at receiver decisions: whether receivers choose to stay out or to confront, conditional on whether they receive a threat with or without sunk cost. Given how the payoffs are structured, we should expect receivers to stay out if they believe they have encountered a high-resolve opponent. By definition, a high-resolve opponent will always fulfill its threat to fight – and depress the receiver's payoff to the minimum – if the receiver does not stay out. Hence, the receiver's decision to stay out provides a behavioral measure of the receiver's perceived credibility of the received threat.

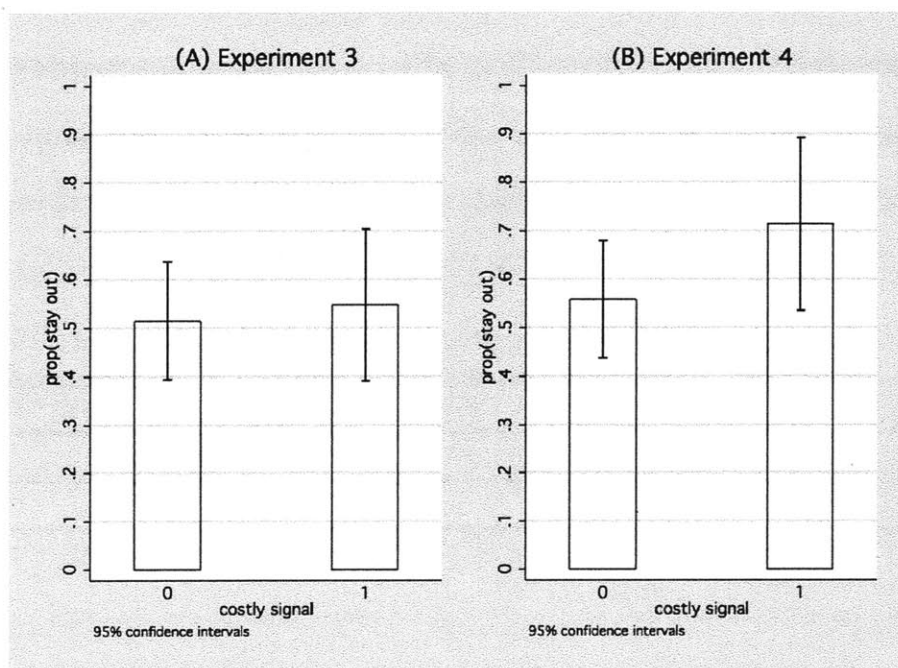
the signaler has the first move in the signaling game, the receiver decision is *not known to the signaler* at the point of decision. Hence, the content of the receiver decision does not impact the choices made by the signaler.

³¹ Unlike in the actual experiment, game outcomes in the pre-experiment could only be generated through random matching of receivers and signalers after all participants had completed the game on AMT. Participants in the pre-experiment were informed of this fact.

³² The data includes the first 15 respondents who completed the experiment in each signaler group.

Figure 4.10 compares the percentage of receivers who chose to stay out based on whether they received a threat with sunk cost. In Experiment 3, the percentages are quite similar across the two groups: 55% of receivers who saw a costly threat decided to stay out compared to 51% of receivers who saw a costless threat ($p = 0.7369$, $n = 110$). On the whole, 53% of all receivers decided to stay out in Experiment 3.

Figure 4.10:
Proportion of Receivers Who Stayed Out (Experiments 3 and 4)



60% of receivers stayed out in Experiment 4. On the whole, 71% of receivers who received a costly threat stayed out in Experiment 4, compared to 56% of receivers who received a costless threat ($p = 0.1568$, $n = 96$). This is a wider difference compared to that in Experiment 3, but it is statistically insignificant. In the first round of the signaling game, 50% of receivers who received a costly threat stayed out and exactly 50% of

receivers who received a costless threat also stayed out ($p = 1.00$, $n = 32$).

It is possible that the decision to stay out is influenced by how risk-averse the receiver is. To control for this possibility, I use logit models with a control for individual risk preference (for Experiments 3 and 4), and round and session fixed-effects (for Experiment 4), to estimate the relationship between receiving a costly threat and staying out. The costly-threat variable is a dummy variable coded as 1 if the receiver received a threat with sunk cost, and 0 if otherwise. Table 4.7 shows the logit estimates compared to a baseline model without controls. The models detect no significant relationship between the costly-threat dummy and the decision to stay out in Experiment 3 ($p > 0.50$) and Experiment 4 ($p > 0.10$). The risk-preference variable is significant in Experiment 3 ($p = 0.040$) but insignificant in Experiment 4 ($p > 0.20$).

Table 4.7:
Logit Estimates – Determinants of Staying Out

	Experiment 3		Experiment 4	
	(1)	(2)	(3)	(4)
Costly threat	0.132 (0.395)	0.229 (0.414)	0.694 (0.476)	0.648 (0.481)
Risk-preference		-0.323 (0.157)*		0.269 (0.245)
Constant	0.059 (0.244)	1.050 (0.519)*	0.932 (0.455)*	0.303 (0.648)
Round & session fixed effects	No	No	Yes	Yes
Log-likelihood	-76.026	-73.266	-58.786	-57.750
Prob>Chi ²	0.738	0.119	0.084	0.156
Pseudo-R ²	0.001	0.037	0.088	0.097
N	110	110	96	95

Notes: ** $p \leq 0.01$; * $p \leq 0.05$. In parentheses are robust standard errors, which are corrected for clustering at the subject level for Experiment 4.

The evidence suggests that a threat with sunk cost does not make receivers more likely to stay out compared to a threat without sunk cost. On this measure, there is no statistically significant difference between a threat with sunk cost and a threat without.

Measuring Perceived Credibility

I also collected data from a non-behavioral measure of signal credibility, which was self-reported by the receiver on a seven-point scale. After receivers had made their decisions in Experiments 3 and 4, they were asked: “Do you think that Player A is a TRUE type?” The receiver responded on a seven-point scale that ranged from “Very unlikely” (0) to

“Very likely” (6). At the midpoint of the credibility scale was “Neither likely nor unlikely” (3).

Figure 4.11:
Average Credibility Scores (Experiments 3 and 4)

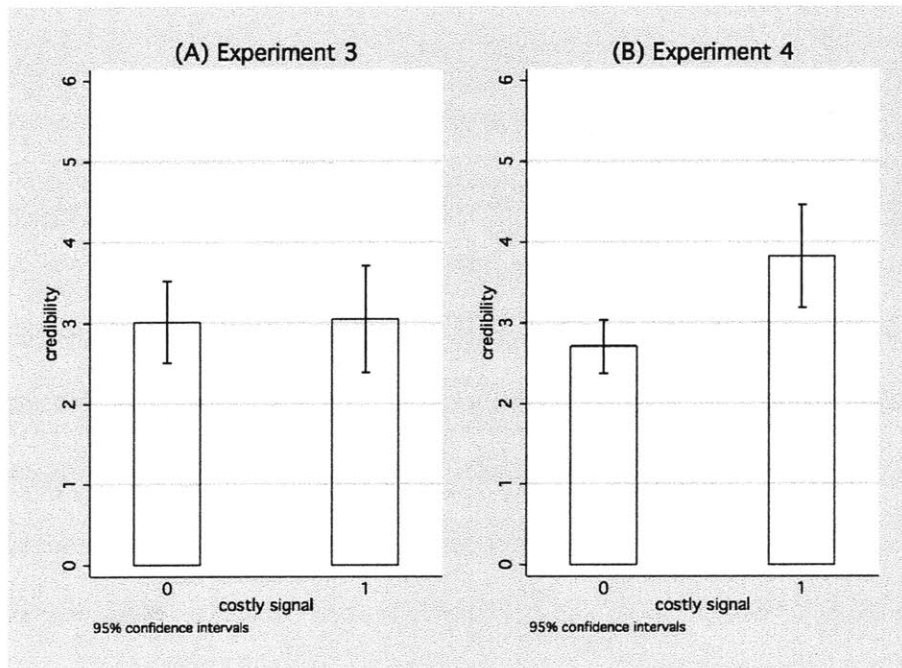


Figure 4.11 compares the credibility estimates by receivers based on whether they received a threat with sunk cost. In Experiment 3, receivers gave the costly threat an average credibility score of 3.05 compared to 3.01 for the costless threat, with no significant difference between the two scores (two-tailed t-test, $p = 0.9369$, $n = 110$). However, Experiment 4 yields a different result: the costly threat received an average credibility score of 3.82, while the costless threat had an average score of 2.71 (two-tailed t-test, $p = 0.0010$, $n = 96$). Table 4.8 shows the ordered logit estimates. Restricting the comparison to the first round of Experiment 4, the costly threat received an average score

of 3.83, while the costless threat had an average score of 2.81 ($n = 32$: two-tailed t-test, $p = 0.1164$; Mann-Whitney test, $p = 0.1204$).

Hence, when they were asked to reflect on the situation after making their decision, the MIT students in the laboratory experiment retrospectively assessed the sunk-cost signal as more credible on average than the signal without sunk cost. However, a diverse sample of U.S. adults recruited from the Internet responded with similar credibility assessments to the sunk-cost signal and the signal without sunk cost.

Table 4.8:
Ordered Logit Estimates – Determinants of Credibility Score

	Experiment 3		Experiment 4	
	(1)	(2)	(3)	(4)
Costly threat	0.036 (0.347)	0.058 (0.350)	1.597 (0.571)**	1.566 (0.612)**
Risk-preference		-0.140 (0.159)		0.062 (0.256)
Cutpoint 1	-1.912 (0.319)	-2.359 (0.611)	-4.339 (0.662)	-4.172 (0.855)
Cutpoint 2	-0.585 (0.239)	-1.021 (0.549)	-2.337 (0.505)	-2.168 (0.856)
Cutpoint 3	-0.317 (0.236)	-0.748 (0.541)	-1.263 (0.410)	-1.088 (0.780)
Cutpoint 4	0.381 (0.236)	-0.049 (0.540)	0.068 (0.430)	0.203 (0.733)
Cutpoint 5	0.573 (0.240)	0.142 (0.548)	1.015 (0.433)	1.152 (0.715)
Cutpoint 6	1.713 (0.292)	1.290 (0.560)	2.761 (0.489)	2.899 (0.673)
Round & session fixed effects	No	No	Yes	Yes
Prob>Chi ²	0.918	0.669	0.001	0.000
Pseudo-R ²	0.000	0.003	0.068	0.068
N	110	110	96	95

Notes: ** $p \leq 0.01$; * $p \leq 0.05$. In parentheses are robust standard errors, which are corrected for clustering at the subject level for Experiment 4.

Experiment 5 provides a separate test based on a concrete example of a sunk-cost signal (military mobilization) in a specific international crisis context (territorial dispute). In this survey experiment, respondents were presented with a foreign-crisis scenario that was exactly the same, except that military mobilization was “very costly” to Country X (the signaler state in the scenario) in the treatment group, and “not very costly” in the control group. The dependent variable is measured on a seven-point credibility scale based on the respondent’s perceived likelihood that Country X would fulfill its threat to fight Country Y. The scale ranges from “Very unlikely” (0) to “Very likely” (6), with “Neither likely nor unlikely” (3) at the midpoint.

The treatment group in Experiment 5 gave an average credibility score of 5.00 while the control group gave 5.23. The difference is statistically insignificant (two-tailed t-test, $p = 0.2908$, $n = 210$). Table 4.9 shows the average credibility scores in the separate 2x2 factorial experiment that controls for the effect driven by the cost of fighting. Again, there is no significant difference in credibility scores between high-cost and low-cost military mobilizations when the cost of fighting is high for Country X (two-tailed t-test, $p = 0.4409$, $n = 211$), or when the cost of fighting is low (two-tailed t-test, $p = 0.5435$, $n = 214$). By contrast, there are highly significant differences in credibility scores between respondents who saw that the cost of fighting is high compared to those who saw that the cost of fighting is low, given high-cost (two-tailed t-test, $p = 0.0032$, $n = 213$) and low-cost (two-tailed t-test, $p < 0.0001$, $n = 212$) military mobilizations.

**Table 4.9:
Average Credibility Scores in Experiment 5**

		Sunk Cost	
		High	Low
Cost of Threat Implementation	High	4.72 (83%)	4.51 (81%)
	Low	5.37 (95%)	5.48 (96%)

Note: In parentheses are the percentages of respondents who believed the threat in each condition.

Type-Separation Reasoning and Perceived Credibility

Does the credibility of sunk-cost signaling depend on type-separation reasoning? Experiment 3 investigates by constructing an additional experimental group of receivers for whom type-separation reasoning is impossible. This group of receivers had no information about the payoffs for the two types of signalers, except that all signalers regardless of type would receive a higher payoff if the receiver stayed out. By this design, I make the logical reasoning behind the sunk-cost signaling mechanism impossible for the receivers, forcing them to assess the costly or costless threat at an intuitive level.

53% of receivers in this group chose to stay out. 62% of receivers who received a costly threat stayed out compared to 45% who received a costless threat ($p = 0.0716$, $n = 117$). The costly threat received an average credibility score of 3.13, while the costless threat had an average score of 2.60 (two-tailed t-test, $p = 0.1436$, $n = 117$). Table 4.10 shows the logit and ordered logit estimates. The costly-threat dummy has a positive relationship with the decision to stay out that is marginally significant ($p = 0.045$) in the logit model that controls for individual risk-preference. In this model, the risk-preference variable has

a significant negative relationship with staying out ($p = 0.006$). However, in the ordered logit regressions with the credibility score as the dependent variable, neither the costly-threat dummy nor the risk-preference variable is significant ($p > 0.10$ and $p > 0.30$ respectively).

Table 4.10:
Logit and Ordered Logit Estimates – Effect of Costly Threat

	<i>Dependent variable:</i> <i>Staying out</i>		<i>Dependent variable:</i> <i>Credibility score</i>	
	Logit (1)	Logit (2)	Ordered logit (1)	Ordered logit (2)
Costly threat	0.676 (0.379)	0.790 (0.394)*	0.506 (0.331)	0.502 (0.331)
Risk-preference		-0.515 (0.188)**		-0.126 (0.130)
Constant	-0.194 (0.256)	1.304 (0.635)*		
Cutpoint 1			-2.161 (0.353)	-2.539 (0.503)
Cutpoint 2			-0.352 (0.244)	-0.717 (0.436)
Cutpoint 3			0.082 (0.244)	-0.279 (0.438)
Cutpoint 4			0.939 (0.262)	0.580 (0.445)
Cutpoint 5			1.139 (0.265)	0.781 (0.451)
Cutpoint 6			2.172 (0.308)	1.814 (0.461)
Prob>Chi ²	0.074	0.005	0.127	0.148
Pseudo-R ²	0.020	0.083	0.006	0.008
N	117	117	117	117

Notes: ** $p \leq 0.01$; * $p \leq 0.05$. Robust standard errors in parentheses.

Compared to the original receiver group (for whom type-separation was possible), there is no statistically significant difference in either the stay-out decision or the credibility score among those who saw the costly threat ($n = 97$: $p = 0.4842$ and $p = 0.8494$

respectively). Among those who saw the costless threat, there is also no significant difference in the stay-out decision or the credibility score ($n = 130$: $p = 0.4722$ and $p = 0.2407$ respectively). On the whole, it does not appear that type-separation reasoning was a crucial driver of the receiver's credibility assessment in Experiment 3.

3.4 Remarks

The concept of costly signaling is central to our understanding of interstate communication under asymmetric information. But the sunk-cost signaling mechanism is difficult to test with observational data: Aside from measurement problems, the effect of a sunk-cost signal is almost always confounded with the effects of other pieces of information, at least some of which are unobserved by the researcher. Yet it is important to know if and how the mechanism operates, in order to understand better the logic of signaling in international relations.

I test the mechanism experimentally. The purpose is to provide an empirical baseline for our theoretical understanding of sunk-cost signaling, and for future investigations of the mechanism in different contexts, by testing whether and how sunk-cost signaling operates under ideal conditions. A novel feature is to use three separate experiments to test the same mechanism: a signaling-game experiment over the Internet; a signaling-game experiment in the laboratory; and a survey experiment that contextualizes sunk-cost signaling in an international crisis scenario. The experiments involved two different samples: a targeted sample of MIT students and a diverse sample of subjects recruited

across the U.S. over the Internet. Experimental replication in two different environments and with two very different subject samples allows us to test if the findings are generalizable and robust. Controlling the information environment is challenging even in an experimental setting. Since each experiment has its imperfections and limitations, it is important to test if the main conclusions are consistent across the different experiments.

Using three different tests complicates the conclusions, but it gives a clearer sense of which results are strong and which are not. First, signalers randomly assigned with high resolve are much more likely to use a signal with sunk cost. This is a clear result. It is replicated both over the Internet with a diverse national sample as well as in a laboratory setting with a narrow sample of MIT students. The result suggests that the sunk-cost signaling mechanism operates well from the angle of the signaler.

But it is unclear whether sunk costs make a threat more credible from the angle of the receiver. Here the results are ambiguous. Statistically significant effects are hard to detect, but there is some non-behavioral evidence based on retrospective assessment by the MIT sample in the laboratory experiment. However, a diverse sample of U.S. adults in the Internet-based experiment gave similar credibility assessments to threats with or without sunk costs.

It is useful to highlight that the experimental results should not be taken as a conclusion that a sunk-cost signal has no credibility effect. What the results suggest is that the sunk-cost signaling mechanism has a strong and clear effect at the signaler's end, but a

relatively weak and less clear effect at the receiver's end. The signaler component of the mechanism is validated and replicated in the experiments, but not the receiver component of the same mechanism.

This suggests an asymmetry between the beliefs and behaviors of signalers and receivers in the experiments. It appears that the logic of sunk-cost signaling is more straightforward and apparent to the signaler than to the receiver. The task of the receiver assessing credibility may be more complicated than the task of the signaler choosing which signal to send. Why this is the case is in itself an interesting puzzle. One possible reason is the existence of a biased asymmetric-information environment where the receiver knows less than the signaler. To the receiver, the signaler's type is a mystery and deception is always a possibility. There may thus be greater uncertainty at the receiver's end than at the signaler's end, resulting in a cognitive load that is heavier for the receiver than it is for the signaler.

While it is beyond the scope of this project, it would be very interesting to replicate the experiment with extremely high stakes or monetary rewards to see if they can motivate receivers to overcome the cognitive load. The replication is especially interesting because the experimental literature suggests the *reverse* of this intuitive hypothesis: experiments on payment-based performance have shown that very large stakes *increase* mistakes rather than decrease them.³³ It would be fascinating to test whether the intuitive expectation or its reverse holds in the sunk-cost signaling context. The result of the

³³ Dan Ariely, Uri Gneezy, George Loewenstein, and Nina Mazar, "Large Stakes and Big Mistakes," *Review of Economic Studies* 76, No. 2 (2009): 451-69.

replication would also connect to the general and very practical question of whether extrinsic incentivization can help to improve judgment in an environment cursed by an adverse signal-to-noise ratio.

In the experiments here, many sources of noise that are likely in real life have been shut off by design: Type-separation is made feasible and transparent; reputational concerns and the shadow of the future are removed; and uncertainty over the signaler's reaction to the receiver's choice is minimized by design. The experiments suggest, however, that the receiver may not be able to respond optimally to the sunk-cost signal even in an idealized environment. This points to a potential tragedy in sunk-cost signaling: Signalers may choose to suffer the sunk cost because they believe that it makes their signal more credible. But receivers do not necessarily respond in line with the signalers' belief, despite the sunk cost suffered. The consequences are wasted resources and a suboptimal outcome for both parties.

PART 4: COSTLY IMPLEMENTATION

Under M5, the perceived credibility of a signal is determined by the perceived costs of implementing that signal. The threat to use force is not credible when its implementation is costly. Observers discount the threat because they know that it will be costly for the threatener to do what it threatened to do.

I test M5 with survey experiments based on diverse national samples recruited online across the United States. In these experiments, I randomly assign subjects to experimental treatments that differ based on the cost of threat implementation. Experiments 6 and 7 are conducted through AMT (see Section 3.2). Experiment 8 is conducted with a different national sample constructed based on U.S. census targets on key demographic variables. My objective, of course, is not so much to make inferences about population parameters, as to test whether M5 holds across multiple experiments with large and diverse samples of respondents.

Experiment 6: Costly Implementation

In Experiment 6, respondents assess the credibility of a threat made by a country in an international crisis scenario. I implemented Experiment 6 with Experiments 5 and 7 in the same survey wave in April 2012, which recruited 1,057 U.S. adult residents through AMT. 212 subjects in the survey wave were randomly allocated to Experiment 6.³⁴

³⁴ The remaining subjects were randomly allocated to Experiments 5 and 7.

Respondents began the experiment by reading about a foreign crisis scenario, in which two states have staked their claims on an important piece of territory. One of the states (“Country X”) threatens to fight a war if the other state moves into the territory. 212 respondents were randomly divided into one of two experimental groups. By design, the two groups saw a scenario that was exactly the same except for a difference in one sentence: One group was informed that “it is clear that fighting a war at this time will be very costly to X”, while the other group was told that fighting a war “will not be very costly”. Again, “very costly” and “not very costly” were used instead of “costly” and “costless”, since fighting a war cannot plausibly be costless. The estimated treatment effect should therefore be a conservative estimate that tends to understate rather than overstate the effect size. Appendix A4 provides the full text of the survey experiment.

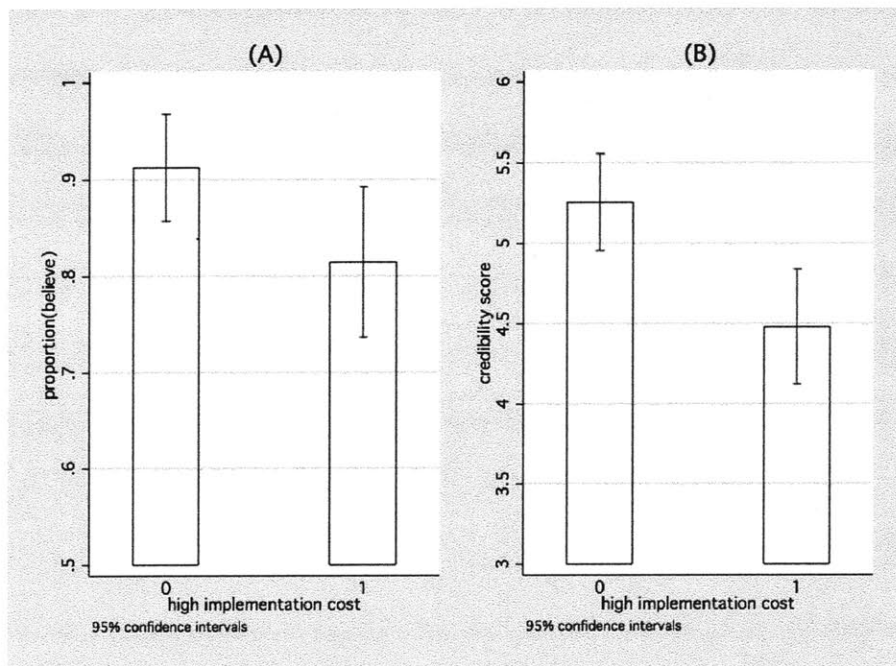
I analyze two dependent variables based on two branched survey questions. The first question asks respondents whether they think the threatener is likely to fulfill its threat of war if the other state ignores the threat. Respondents choose one of three options: “likely”, “unlikely” or “neither likely nor unlikely”. The proportion of those who choose “likely” (or “unlikely”) provides a measure of the proportion of respondents who believe (or do not believe) the threat. Those who choose “neither likely nor unlikely” are excluded in this measure. A second branched question asks respondents their perceived likelihood that the threatener will fulfill its threat. Those who chose “likely” (or “unlikely”) in the first question will choose between “very likely” and “somewhat likely” (or “very unlikely” and “somewhat unlikely”) in the second question. Those who

chose “neither likely nor unlikely” in the first question will be asked if they “lean toward believing”, “lean toward disbelieving” or “lean neither way”. Thus, the second question constructs a seven-point scale that gives a more precise measure of perceived credibility.

Result 17: The cost of threat implementation has a negative effect on the credibility of the threat.

Experiment 6 provides evidence for M5. The cost of implementing the threat has a direct negative effect on its perceived credibility. Figure 4.12 summarizes the findings.

Figure 4.12:
(A) Proportion of Believers and (B) Average Credibility Scores in Experiment 6



91% of respondents believed the threat when it was known that implementing the threat “will not be very costly”. On the other hand, the percentage is 81% in the experimental condition where it was known that fighting a war “will be very costly” – a difference of 10 percentage points ($p = 0.0423$, $n = 200$).³⁵ More precisely, the average credibility score for the threat to use force is 5.25 when the cost of threat implementation is low and 4.48 when the implementation cost is high. The 0.8-point difference on the seven-point scale is significant at $p = 0.0012$ (two-tailed t-test, $n = 212$).

Experiment 5 provides further evidence to support the results in Experiment 6. Experiment 5 (see Section 3.3) is similar to Experiment 6 except that it uses a 2x2 factorial design to capture the interaction between the cost of threat implementation and the sunk-cost signal.

Table 4.9 (in Section 3.3) shows the average credibility scores and the relative proportions of believers across the four experimental conditions. The threat of war is less credible when the threat is costly to implement, in both the low sunk-cost condition (two-tailed t-test, $p < 0.0001$, $n = 212$) and the high sunk-cost condition (two-tailed t-test, $p = 0.0032$, $n = 213$). Table 4.9 also shows that fewer respondents believed the threat when they learned that the threat is costly to implement ($p = 0.0005$, $n = 198$ in low sunk-cost condition; $p = 0.0058$, $n = 204$ in high sunk-cost condition).

³⁵ Recall that the n excludes those who chose “neither likely nor unlikely” in the first branched question.

Experiment 7: Audience Costs

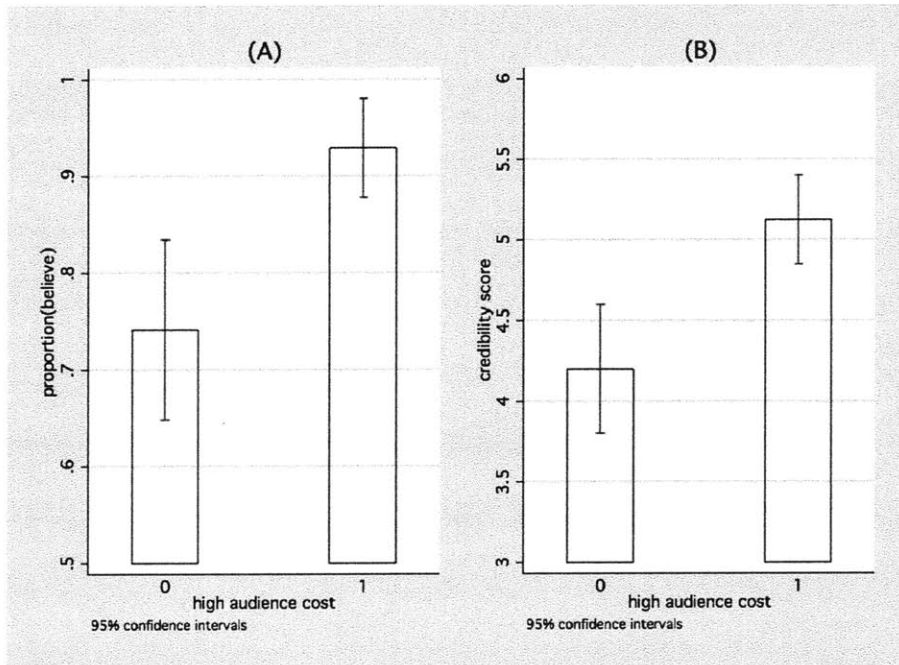
Experiment 7 investigates the credibility effect of audience costs. As discussed in Chapter 1, audience costs should increase perceived credibility by increasing the cost of non-implementation. The threat of war should be more credible if it is costly *not* to implement the threat. While there are a number of survey experiments on the existence and determinants of domestic audience costs, there is no experiment that tests whether and how audience costs affect perceived credibility.

Experiment 7 had a similar setup as Experiment 6 and was implemented in the same survey wave. 210 subjects in the survey were randomly allocated to Experiment 7. The foreign crisis scenario was largely similar to that in Experiment 6. Respondents were randomly assigned to one of two experimental scenarios. In both scenarios, the Country X “has publicly announced its threat of war.” Respondents in the first experimental group were told that “it will be very costly” to the threatener if it reneges on the threat (high audience costs). Respondents in the second experimental group were told that reneging on the threat “will not be very costly” (low audience costs). The focus is on the costs of backing down after publicly announcing the threat to use force. The experiment did not specify whether the audience costs were domestic or international. Appendix A4 provides the text of the survey experiment.

Result 18: A publicly announced threat to use force is more credible when it involves high audience costs.

Figure 4.13 shows how variations in audience costs shape perceived threat credibility.

Figure 4.13:
(A) Proportion of Believers and (B) Average Credibility Scores in Experiment 7



In the scenario where audience costs are high, 93% of the respondents believed the threat to use force. In the experimental condition where audience costs are low, the percentage is 74%, a drop of 19 percentage points. This difference is substantially large and statistically significant ($p = 0.0004$, $n = 188$). Meanwhile, the average credibility score is also higher at 5.12 when audience costs are high compared to 4.20 when

audience costs are low (two-tailed t-test, $p = 0.0002$, $n = 210$). Overall, the results show that respondents are more willing to believe a publicly announced threat to use force when they think that audience costs are high.

Experiment 8: Costly Implementation Signals with Realistic Complications

Experiments 8, 9 and 10 were implemented in the same survey wave based on a sample of 1,241 U.S. adults. The experiments were fielded from 24 to 27 January 2012 through Survey Sampling International (SSI), which constructed a national sample based on U.S. census targets on age, geography, income and education.³⁶ 536 subjects in the survey wave were randomly allocated to Experiment 8.³⁷

Experiment 8 constructed three different robustness tests for M5. Subjects were randomly divided into one of the three tests. Each had a vignette that started with the same setup, in which two countries (Country X and Country Y) have staked their claims on an important piece of territory. An international crisis is triggered when X sent its troops into the territory. The dependent variables are similar to those used in Experiments 5-7. Appendix A5 provides the full text of the survey experiment.

³⁶ SSI is a survey sampling firm that has been used by researchers across multiple fields ranging from medicine and environmental science to management and political science. For further details on SSI, see: <http://www.surveysampling.com/>.

³⁷ The remaining subjects were randomly allocated to Experiments 9 and 10.

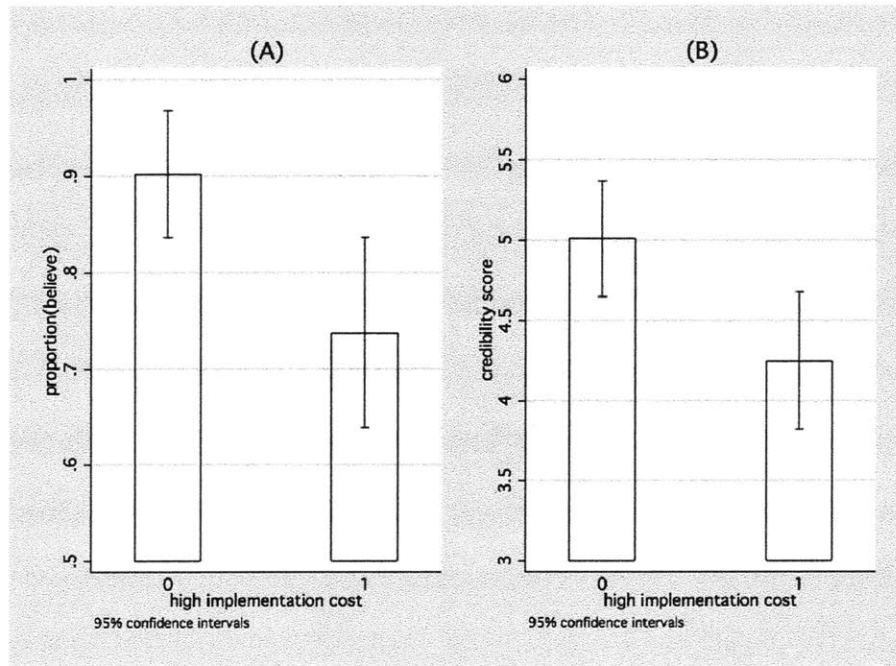
Results

The first test investigates if M5 holds when the cost of threat implementation is inferred based on a past incident. In this test, subjects were not told whether it is costly to implement the threat to use force. Instead, they made their own inference based on a past experience. Respondents were told about a previous incident where the two countries fought after Country X sent its troops into the territory. The respondents were randomly divided into one of two experimental groups. One group learned that “Y had a strong military, so it did not take a major effort for Y to push X out from the territory previously.” The other group learned that “X had a strong military, so it took a major effort for Y to push X out from the territory previously.” Hence, respondents in the two groups saw a scenario that was the same, except for the implied difference in the cost of threat implementation based on a previous crisis.

Result 19: A threat is more credible when a past incident suggests that the threatener can fulfill its threat easily.

The credibility effect of M5 remains significant when the cost of threat implementation is inferred based on a past incident. Figure 4.14 describes the result.

Figure 4.14:
(A) Proportion of Believers and (B) Average
Credibility Scores in Experiment 8 (First Test)



90% of the respondents believed the threat in the scenario where the *threatener* had a strong military and had easily expelled the challenger in a previous incident. In contrast, the percentage is 74% in the condition where the *challenger* had a strong military and it took the threatener a major effort to expel the challenger. The difference of 16 percentage points is significant at $p = 0.0062$ ($n = 162$). The average credibility score is also higher at 5.01 in the former condition compared to 4.25 in the latter condition (two-tailed t-test, $p = 0.0077$, $n = 182$). The results show that respondents find a threat more credible when a previous incident suggests that the threatener can fulfill its threat easily.

Second and third tests

The second and third tests investigate if M5 holds when the crisis context is thickened with other types of signals used in actual historical crises. The second test introduces actual verbal threats used in the correspondences between China and India on the eve of the 1962 Sino-Indian War. The threats were issued by China and subsequently translated by the Indian Government in a parliamentary white paper. The verbal threats are inserted ad-verbatim into the experimental scenarios in quotes. As before, the country identities are neutralized. In the experiment, Country Y made a diplomatic protest and later warned that “he who plays with fire will eventually be consumed by fire.”³⁸ Subsequently, Country Y lodged “the strongest and most serious protest” against the intrusion and warned that if Country X continued “to spread the flames of war, it must bear full responsibility for the resulting casualties on both sides and all other consequences that may ensue.”³⁹

The third test is similar to the second test, except that it *further* thickens the crisis context with both verbal threats as well as military moves on the ground. In addition to one verbal threat (the “strongest and most serious protest” paragraph), the scenarios in the third test have Country Y increasing its troops at the border and building military installations around the disputed territory. These operational moves on the ground were similar to those taken by China on the eve of the 1962 Sino-Indian War.

³⁸ Note by Chinese Ministry of Foreign Affairs to Embassy of India in China, 13 September 1962, in Ministry of External Affairs, *Notes, Memoranda and Letters Exchanged Between the Governments of India and China: White Paper No. 7* (New Delhi: Government of India, 1962), 68.

³⁹ Note by Chinese Ministry of Foreign Affairs to Embassy of India in China, 11 October 1962, in Ministry of External Affairs, *White Paper No. 7*, 106.

In each test, the respondents were randomly divided into one of two experimental groups. One group was told: “It is known that fighting a war at this time will be very costly to Y.” The other group learned that fighting a war “will not be very costly”. Appendix A5 provides the full text.

The credibility effect of M5 remains significant in the second test, in which the experimental context is thickened with examples of verbal threats taken ad-verbatim from the 1962 Sino-Indian crisis.

Figure 4.15:
(A) Proportion of Believers and (B) Average Credibility Scores in Experiment 8 (Second Test)

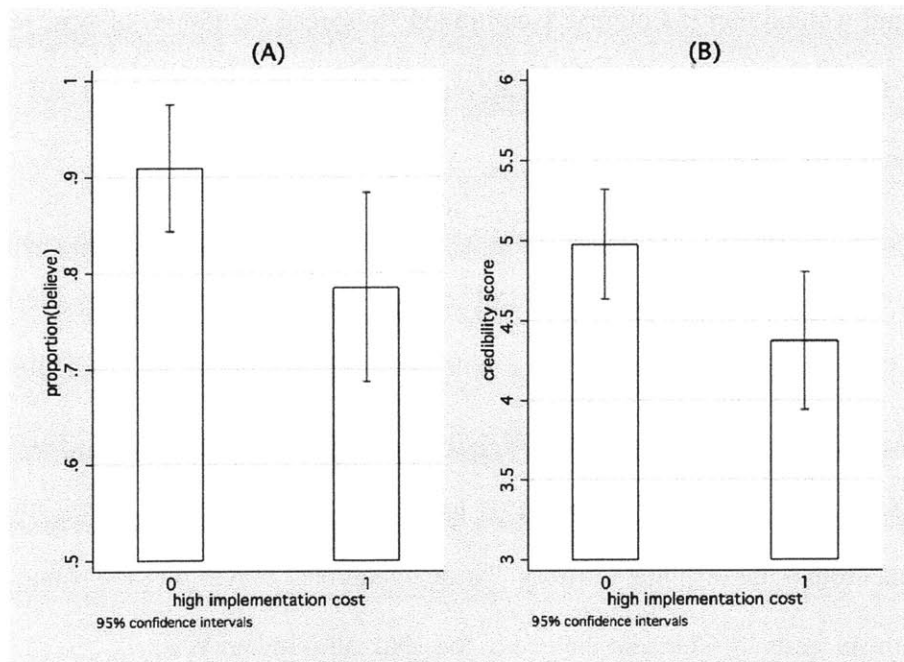


Figure 4.15 shows that 91% of the respondents believed the threat in the experimental condition with a low cost of threat implementation. The percentage falls to 79% in the

condition with a high implementation cost, a drop of 12 percentage points ($p = 0.0362$, $n = 147$). The average credibility score is also lower at 4.37 in the condition with a high implementation cost compared to 4.98 in the condition with a low implementation cost (two-tailed t-test, $p = 0.0301$, $n = 173$). The evidence suggests that the credibility effect of M5 remains robust when the signaling environment is thickened with verbal threats.

However, the statistical significance is diminished in the third test – when the experimental context is *further* thickened with both verbal threats as well as military moves on the ground.

Figure 4.16:
(A) Proportion of Believers and (B) Average Credibility Scores in Experiment 8 (Third Test)

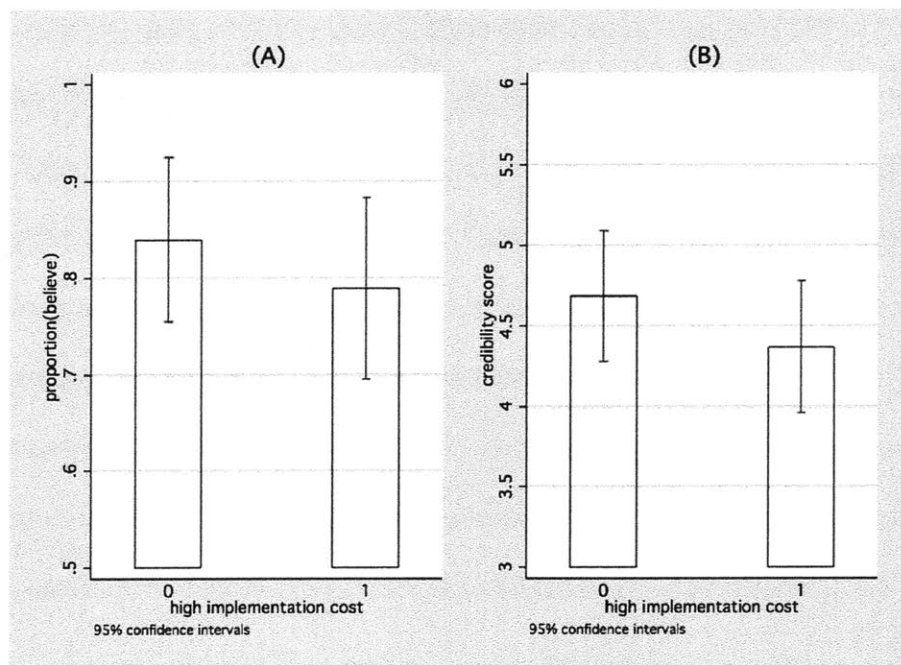


Figure 4.16 shows that 84% of the respondents believed the threat in the condition with a low implementation cost. The percentage is 79% in the condition with a high implementation cost ($p = 0.4244$, $n = 151$). Likewise, the credibility score is higher at 4.69 in the condition with a low implementation cost compared to 4.37 in the condition with a high implementation cost. While the direction of change agrees with M5, the difference does not reach statistical significance (two-tailed t-test, $p = 0.2775$, $n = 181$).

PART 5: CONTRADICTION SIGNALING

Based on M6, signals are less credible when they involve a salient contradiction. A contradiction occurs when signals of strong resolve are mixed with signals of weak resolve. The contradiction is salient if it is registered and considered by the receiver. A series of signals with a salient contradiction is noisier and less credible.

Experiment 9: Contradictory Signaling based on Korean-War Counterfactuals

Experiment 9 was implemented in the same survey wave described in the last section, based on a national sample constructed with U.S. census targets on key demographic variables. 352 subjects in the survey wave were randomly allocated to Experiment 9.

Experiment 9 tests M6 using historical counterfactuals based on the Korean War: Did contradictory signals from Washington compromise the credibility of the U.S. commitment to defend South Korea? In the summer of 1949, the U.S. withdrew its troops from South Korea. In January 1950, Secretary of State Dean Acheson publicly omitted South Korea from the U.S. defense perimeter in his National Press Club speech. It is known that Stalin and Mao had discussed Acheson's speech, and the speech is believed to have influenced them as they assessed the credibility of the U.S. commitment to defend South Korea.⁴⁰ Stalin subsequently concluded that "the

⁴⁰ Thomas Christensen, *Worse Than a Monolith: Alliance Politics and Problems of Coercive Diplomacy in Asia* (Princeton, NJ: Princeton University Press, 2011), 53. For a contrary view, see James Matray, "Dean Acheson's National Press Club Speech Reexamined," *Journal of Conflict Studies* 22, No. 1 (2002): 28-55.

prevailing mood [in the U.S.] is not to interfere.”⁴¹

Would the U.S. commitment be more credible if Acheson had *included* South Korea within the U.S. defense perimeter in his National Press Club speech? Would credibility be enhanced if Washington did *not* withdraw its troops from South Korea? It is impossible to know for sure since we can never rerun history. But we can devise a theoretical test by using experiments to construct stylized historical counterfactuals.

Experiment 9 randomly assigned respondents to one of four counterfactual scenarios. Each scenario began with an international crisis setup similar to the one in Experiment 8, with neutralized country identities. The four counterfactuals differed along two dimensions: (1) whether the threatener withdrew or continued to maintain its troops on the territory (*Withdrawal* or *No-Withdrawal*); and (2) whether the threatener included or did not include “the defense of the territory as one of its “core security interests” in a public announcement of its national defense strategy” (*No-Acheson* or *Acheson*). The *Acheson/Withdrawal* condition is Counterfactual 0 – it is not strictly a counterfactual since it stylizes what actually happened on the eve of the Korean War. The remaining experimental conditions are true counterfactuals: The *No-Acheson/No-Withdrawal* condition is Counterfactual 1, the *No-Acheson/Withdrawal* condition is Counterfactual 2, and the *Acheson/No-Withdrawal* condition is Counterfactual 3. Appendix A5 provides the full text used in each counterfactual scenario.

⁴¹ William Stueck, *Rethinking the Korean War: A New Diplomatic and Strategic History* (Princeton, NJ: Princeton University Press, 2002), 73.

Results

If M6 holds, our data should show three patterns. Each pattern suggests a test for M6:

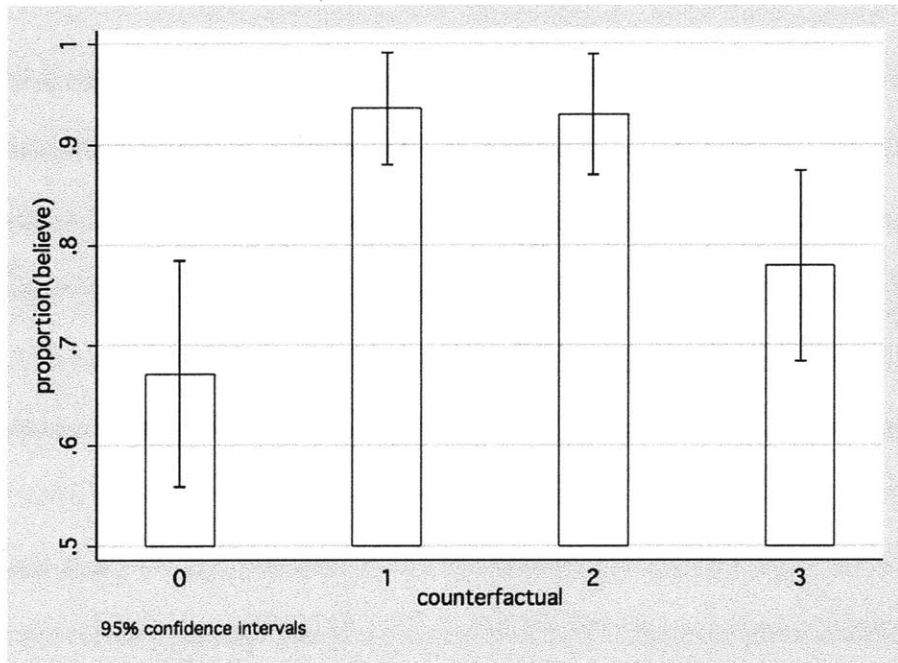
Test 1: The threat of war is *more credible* in the *No-Acheson/No-Withdrawal* condition (Counterfactual 1) than in the *Acheson/Withdrawal* condition (Counterfactual 0).

Test 2: The threat of war is *less credible* in the *Acheson/Withdrawal* condition (Counterfactual 0) than in either the *No-Acheson/Withdrawal* condition (Counterfactual 2) or the *Acheson/No-Withdrawal* condition (Counterfactual 3).

Test 3: The threat of war is *more credible* in the *No-Acheson/No-Withdrawal* condition (Counterfactual 1) than in either the *No-Acheson/Withdrawal* condition (Counterfactual 2) or the *Acheson/No-Withdrawal* condition (Counterfactual 3).

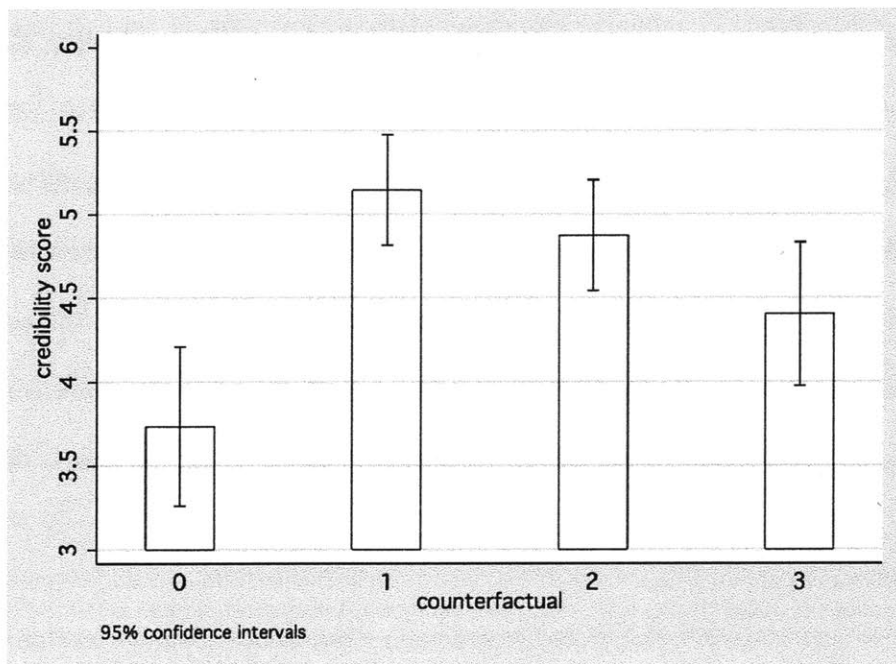
Test 1 is the most important. This is because if M6 is correct, the *No-Acheson/No-Withdrawal* condition should be the *most* credible and the *Acheson/Withdrawal* condition the *least* credible. By contrast, M6 makes no prediction about the relative credibility of the *No-Acheson/Withdrawal* condition and the *Acheson/No-Withdrawal* condition. Which of the two is the more credible remains an open question. But the experiment can suggest a plausible answer. Figures 4.17 and 4.18 show how perceived threat credibility changes across the four counterfactuals.

Figure 4.17:
Distribution of Believers Across Four Korean-War Counterfactuals



Notes: Counterfactuals 0 (Acheson/Withdrawal), 1 (No-Acheson/No-Withdrawal), 2 (No-Acheson/Withdrawal), and 3 (Acheson/No-Withdrawal).

Figure 4.18:
Average Credibility Scores Across Four Korean-War Counterfactuals



Result 20: Signals of resolve are less credible when they involve a salient contradiction in the series of signals sent.

M6 passes Test 1: The threat of war is indeed more credible in the *No-Acheson/No-Withdrawal* condition (Counterfactual 1) than in the *Acheson/Withdrawal* condition (Counterfactual 0). 94% of respondents believed the threat in the *No-Acheson/No-Withdrawal* condition compared to 67% in the *Acheson/Withdrawal* condition. This is a large and highly significant difference of 26 percentage points ($p < 0.0001$, $n = 148$). The average credibility score in the *No-Acheson/No-Withdrawal* condition is 5.15 but only 3.73 in the *Acheson/Withdrawal* condition. The difference of almost one and a half points on the seven-point scale is significant at $p < 0.0001$ (two-tailed t-test, $n = 173$).

M6 largely passes Test 2: The threat of war is less credible in the *Acheson/Withdrawal* condition (Counterfactual 0) than in either the *No-Acheson/Withdrawal* condition (Counterfactual 2) or the *Acheson/No-Withdrawal* condition (Counterfactual 3). 67% of respondents believed the threat in the *Acheson/Withdrawal* condition compared to 93% in the *No-Acheson/Withdrawal* condition ($p = 0.0001$, $n = 142$) and 78% in the *Acheson/No-Withdrawal* condition. The latter difference, however, is not significant ($p = 0.1425$, $n = 147$). Using a more precise measure, the average credibility score in the *Acheson/Withdrawal* condition is 3.73 compared to 4.88 in the *No-Acheson/Withdrawal* condition (two-tailed t-test, $p = 0.0001$, $n = 176$) and 4.40 in the *Acheson/No-Withdrawal* condition (two-tailed t-test, $p = 0.0383$, $n = 175$), with both differences significant at the 5% level.

The results of Test 3 are mixed: The threat of war is more credible in the *No-Acheson/No-Withdrawal* condition (Counterfactual 1) than the *Acheson/No-Withdrawal* condition (Counterfactual 3) but not the *No-Acheson/Withdrawal* condition (Counterfactual 2). 94% of respondents believed the threat in the *No-Acheson/No-Withdrawal* condition compared to 78% in the *Acheson/No-Withdrawal* condition ($p = 0.0052$, $n = 155$). The average credibility score in the former condition is 5.15 compared to 4.40 in the latter condition (two-tailed t-test, $p = 0.0072$, $n = 176$). On the other hand, there is no significant difference in the percentage of respondents who believed the threat in the *No-Acheson/No-Withdrawal* condition (94%) and the *No-Acheson/Withdrawal* condition (93%) ($p = 0.8957$, $n = 150$). The average credibility score is 5.15 in the former condition and 4.88 in the latter condition (two-tailed t-test, $p = 0.2499$, $n = 177$). This result suggests the relatively weaker influence of the *Withdrawal* condition vis-à-vis the *Acheson* condition.

In line with this suggestion, the *No-Acheson/Withdrawal* condition (Counterfactual 2) turns out more credible than the *Acheson/No-Withdrawal* condition (Counterfactual 3). 93% of respondents believed the threat in the former condition compared to 78% in the latter condition ($p = 0.0093$, $n = 149$). However, the statistical significance is diminished with a more precise measure based on credibility scores. The score is higher at 4.88 in the *No-Acheson/Withdrawal* condition than 4.40 in the *Acheson/No-Withdrawal* condition but does not reach statistical significance at the 5% level (two-tailed t-test, $p = 0.0844$, $n = 179$).

Experiment 10: Costly Implementation Theory in the Korean-War Counterfactuals

Experiment 10 was implemented in the same survey wave as Experiments 8 and 9. A total of 353 subjects in the survey wave were randomly allocated to Experiment 10.⁴²

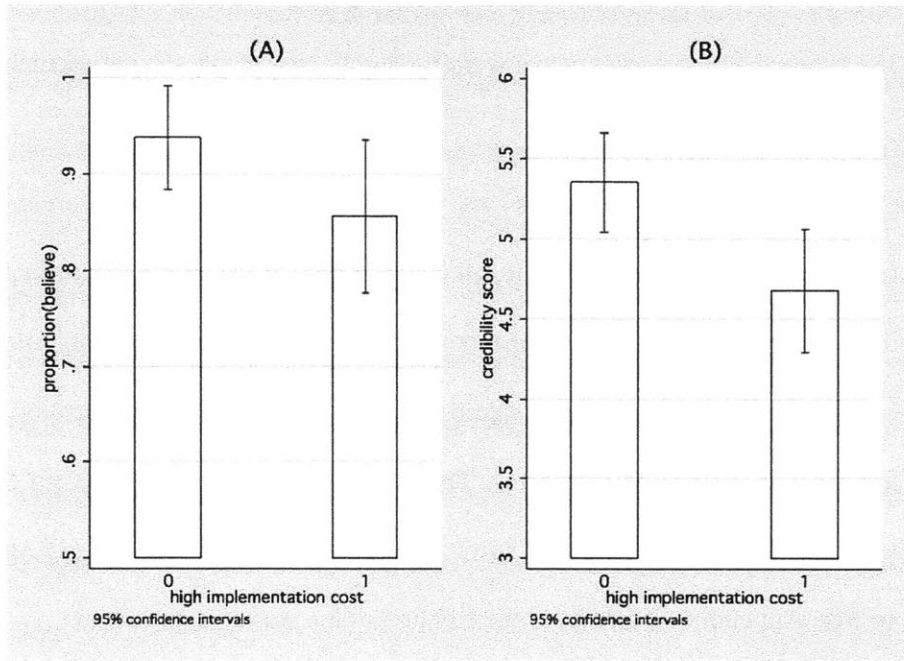
Experiment 10 provides a robustness test for M5 based on the Korean War counterfactuals. It uses Counterfactual 0 (*Acheson/Withdrawal*) and Counterfactual 1 (*No-Acheson/No-Withdrawal*) from Experiment 9. Respondents were randomly allocated to either Counterfactual 0 or 1. Respondents in each counterfactual condition were randomly assigned to one of two scenarios. The scenarios were the same except for the cost of threat implementation (high or low), as described in the earlier experiments pertaining to M5. Appendix A5 provides the full text of the survey experiment.

Results

Does the cost of threat implementation affect credibility in the Korean War counterfactuals? Figures 4.19 and 4.20 describe the results.

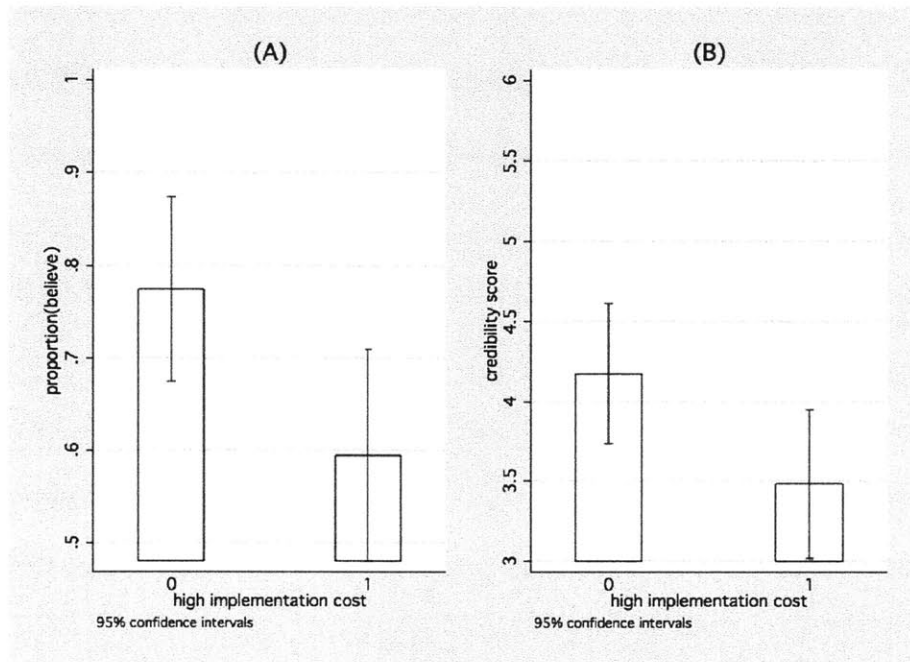
⁴² These are different subjects from those allocated to Experiments 8 and 9.

Figure 4.19:
(A) Proportion of Believers and (B) Average Credibility Scores
Under Korean-War Counterfactual 1 (“No Acheson/No Withdrawal”)



Under Counterfactual 1 (*No-Acheson/No-Withdrawal*), 86% of the respondents believed the threat when they knew that threat implementation was very costly. In contrast, the percentage is 94% in the scenario where threat implementation was not very costly, a difference of 8 percentage points that is significant at the 10% level but not at the 5% level ($p = 0.0911$, $n = 158$). The statistical significance, however, is sharpened with a more precise measure based on credibility scores. The average credibility score is lower at 4.68 with high implementation cost compared to 5.36 with low implementation cost (two-tailed t-test, $p = 0.0072$, $n = 177$).

Figure 4.20:
(A) Proportion of Believers and (B) Average Credibility Scores
Under Korean-War Counterfactual 0 (“Acheson/Withdrawal”)



In the context of Counterfactual 0 (*Acheson/Withdrawal*), 77% of the respondents believed the threat in the scenario with a low cost of threat implementation. In contrast, the percentage is 59% in the condition with a high implementation cost, a significant drop of 18 percentage points ($p = 0.0199$, $n = 145$). The average credibility score is also lower at 3.48 with high implementation cost compared to 4.17 with low implementation cost (two-tailed t-test, $p = 0.0331$, $n = 176$).

Results from the different experiments point in a similar direction: the credibility of a threat is negatively influenced by the perceived costs of implementing that threat. The costly implementation theory largely holds despite the contextual variations across

several survey experiments. Respondents significantly discounted the threat to use force when they knew that it would be costly for the threatener to do what it threatened to do.

Chapter 5

Wars in East Asia: Assessing Old Hypotheses, Inferring New Hypotheses

Sixteen international wars were fought in East Asia in the last century (see Table 1.2). What caused these wars? What were the motivations and decision processes leading to war? Does the historical evidence support the theories in Chapter 1? What new theories or scope conditions can be inferred from the historical cases?

Chapter 4 conducted experimental tests of the mechanisms described in Chapter 1. Chapter 3 examined real-world decision processes. Based on two archival case studies, it provided a fine-grained assessment of whether the theoretical expectations fit the historical facts. In this chapter, I examine the origins of sixteen East Asian wars fought in the 20th century. The chapter has two theoretical goals. The first is theory refinement: To use the full case universe to check the external validity of the mechanisms and to suggest theoretical refinements. The second is theory generation: To infer alternative hypotheses on war that fall outside the rationalist mechanisms described in Chapter 1.

I divide this chapter into two parts. Part 1 deals with theory assessment and refinement. I divide this part into seven sections corresponding to Theory 1 and Mechanisms M1-M6. Part 2 infers new hypotheses from the East Asian case universe. I divide this part into four sections corresponding to four sets of hypotheses. The first two retain the unitary-

state assumption. The last two remove the assumption and focus on domestic causes of war. The chapter ends with five general observations on the East Asian cases.

Let us first begin with a broad comparative view. Four figures (Figures 5.1 to 5.4) allow us to visualize the case universe in four different dimensions.

Figure 5.1: War in East Asia, 1900-2000¹

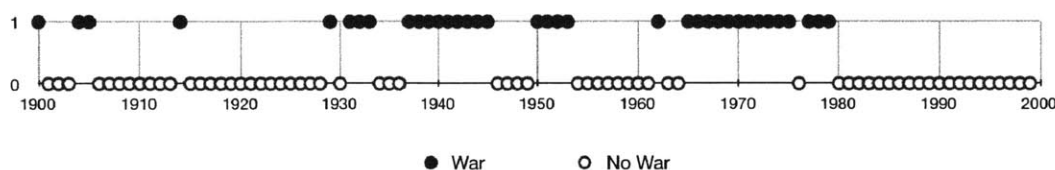


Figure 5.1 plots the war variable over a hundred years. The war variable is 1 (red dot) if there was an international war ongoing within East Asia in any part of that year, and 0 (blue circle) if otherwise. There was about a quarter-century of international peace after 1905 and before 1929, if we exclude the opportunistic small-scale war launched by Japan against Germany in World War I.² The longest period of international peace is from 1980 to the present. In an uncanny manifestation of the 20-80 rule, the two decades between 1935 and 1955 accounted for more than 80% of all battle fatalities in the whole case universe.

¹ See the international wars listed in Table 1.1. Note that a value of 0 on the war variable does not necessarily mean that peace prevailed. There were, for instance, the Taiwan Strait crises in 1954 and 1958; the battles between North and South Vietnam in 1963-64; the Sino-Soviet border clashes in 1969; the Sino-Vietnamese border clashes in 1987; and the Spratly Islands crisis in 1988.

² This case can be excluded under the numerical definition of a war, as it falls short of a thousand battle fatalities. However, Japan had formally declared war on Germany, and it was clear to both countries that they were at war.

Figure 5.2: Comparative Intensity of War (Battle Fatalities)³

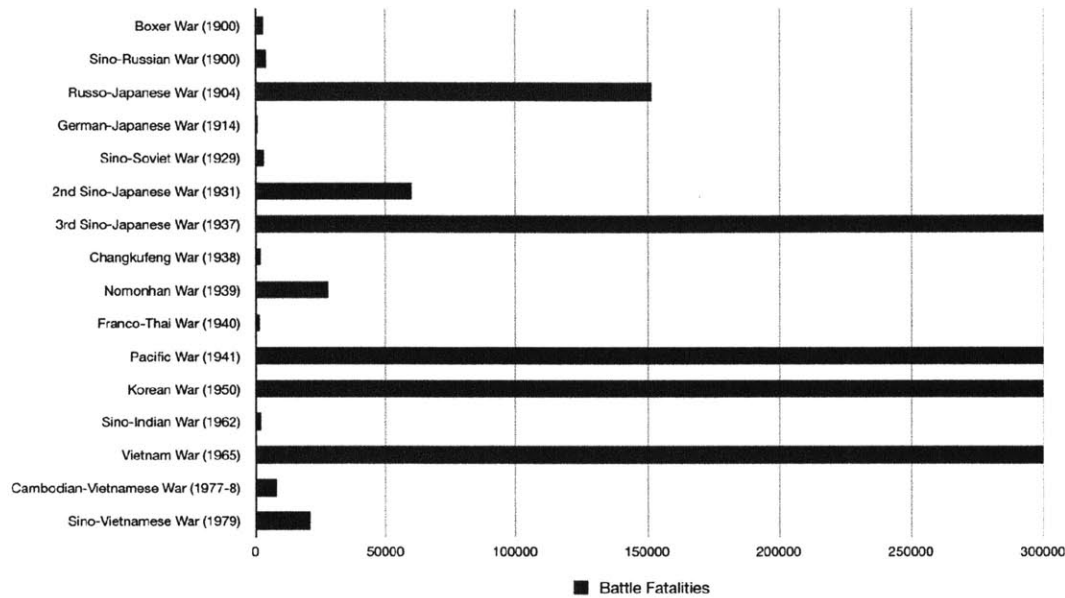


Figure 5.2 compares the intensity of war across sixteen cases, using battle fatalities as the measure. In order to visualize the data for the low-intensity wars, the maximum of the scale is set at 300,000 battle fatalities. Three features stand out in the figure. First, four wars exceed the maximum scale: the Third Sino-Japanese War, the Pacific War, the Korean War, and the Vietnam War. Battle fatalities in these wars ranged from more than three times the maximum scale (Korean War) to more than eleven times the maximum scale (Pacific War). Second, there is a wide variance in war intensity across the cases, with a cluster of low-intensity wars (< 10,000 fatalities) at one end and a cluster of high-intensity wars (> 100,000 fatalities) at the other end. Only three wars lie between the

³ Source: Correlates of War 4.0. In parenthesis is the year of war initiation. Statistics on the German-Japanese War are based on Micheal Clodfelter, *Warfare and Armed Conflicts: A Statistical Reference to Casualty and Other Figures, 1618-1991* (Jefferson, NC: McFarland, 1992), 746-7.

extremes: the Second Sino-Japanese War, the Nomonhan War, and the Sino-Vietnamese War. Third, the majority of the wars were low-intensity wars. Excluding the Cambodian-Vietnamese War, fatalities in low-intensity wars ranged between 1,000 to 5,000. These wars made up almost half the case universe.

Figure 5.3: Comparative Intensity of War (War Duration)⁴

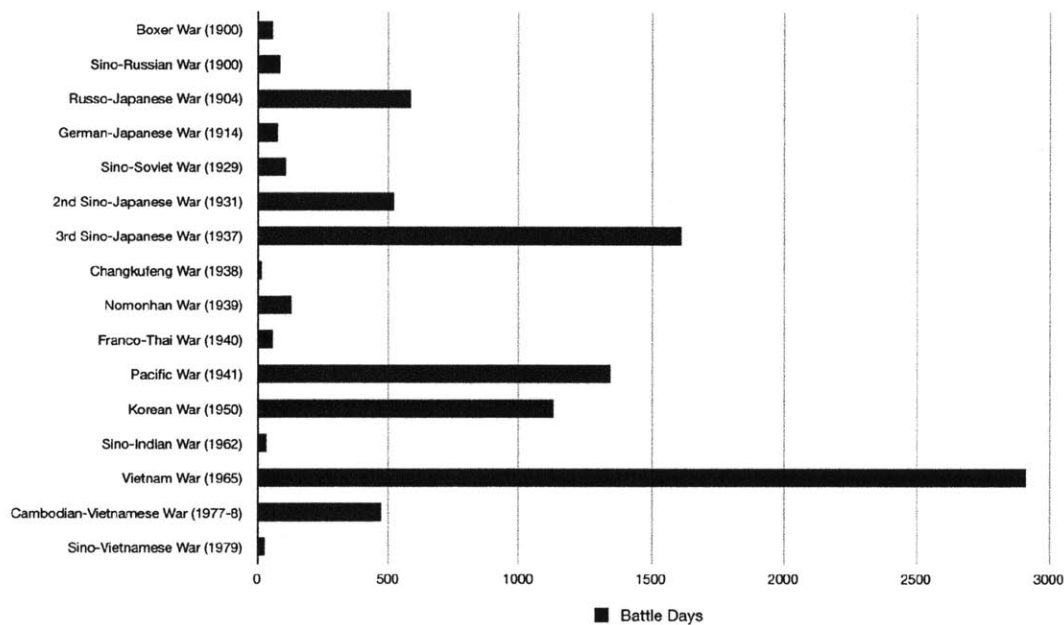


Figure 5.3 measures the intensity of war based on war duration. Based on Figures 5.2 and 5.3, the relative war intensity is largely consistent between the two measures, with the obvious exceptions of the Cambodian-Vietnamese and Sino-Vietnamese wars, and the Vietnam War vis-à-vis the Pacific War and Third Sino-Japanese War. The majority of the

⁴ Source: Correlates of War 4.0. In parenthesis is the year of war initiation. The start date for the Korean War is defined as 24 June 1950. The duration of the Vietnam War is calculated based on the period from 7 February 1965 (start date of regular U.S. bombing of North Vietnam) to 27 January 1973 (start date of ceasefire based on the Paris Peace Accords). The Third Sino-Japanese War merged into the Pacific War in 1941.

wars were relatively short. Seven wars lasted less than a hundred days, of which five ended within two months. But there is also a wide variance in the duration of war. Six wars lasted more than five hundred days. Four of the five high-intensity wars lasted more than a thousand days.

Figure 5.4: Relative Power Between War Parties (Based on CINC Scores) ⁵

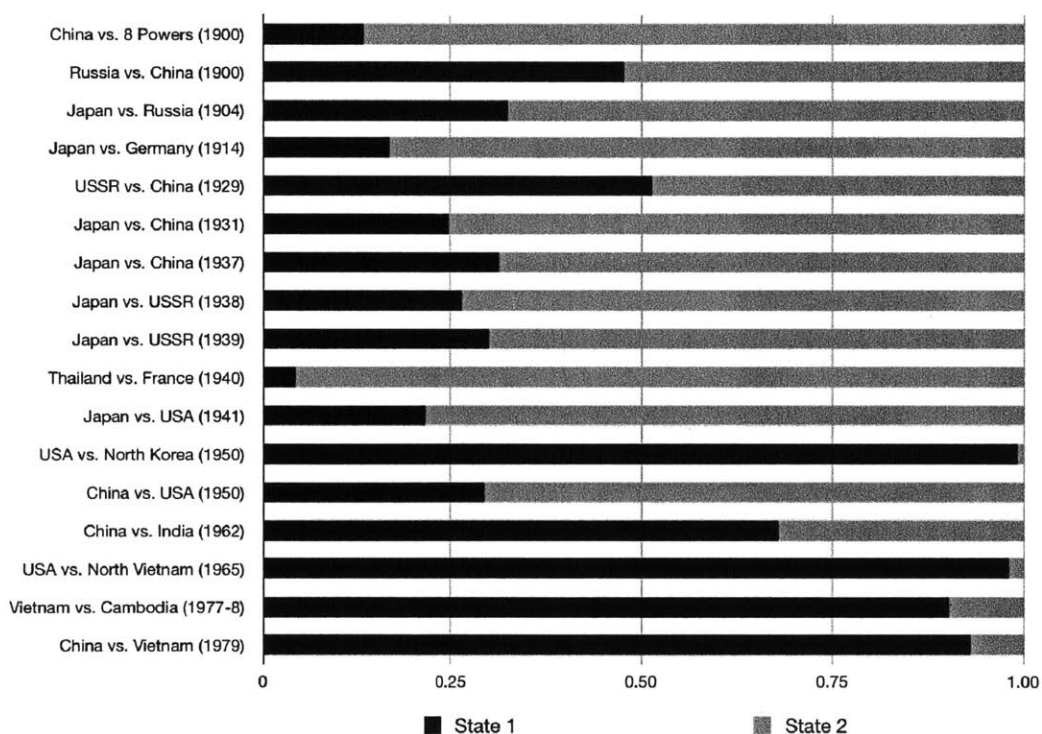


Figure 5.4 measures the relative power between the warring parties based on the Composite Index of National Capability (CINC). “State 1” ⁶ is the state that made the formal decision for war. Relative power is measured by the proportion of the CINC score

⁵ Source: National Material Capabilities Dataset 4.0. Year of war initiation in parenthesis. “State 1” refers to the states listed before “vs.” on the y-axis.

⁶ “State 1” refers to the states listed before “vs.” on the y-axis.

of each side over the total CINC scores of both sides. The CINC score for each state essentially sums its six capability components (total national population, urban population, iron and steel production, energy consumption, military personnel, and military expenditure), with each component standardized based on its world share and averaged across the six components.⁷

Figure 5.4 reveals a puzzle: As many as nine wars were fought between states with at least a 3:1 disparity in relative national capabilities. In five cases, it was the disadvantaged state that made the decision for war. Why would a state choose to fight when there is a 3:1 power disparity? One way to resolve the puzzle is to point to the common problems associated with interstate war statistics. The CINC score is an aggregate measure that does not always accurately reflect the relative military balance. For instance, the CINC score for Japan does not manage to capture its military superiority over China in the 1930s. Furthermore, the data quality varies widely across the years and cases. Thus, the measurement error may vary widely from country to country, making bilateral comparisons imprecise. It is also well known that quantitative aggregations often suppress important cross-case and cross-time differences that become obvious with the conduct of case studies. For instance, the wars initiated in 1914 (by Japan) and 1940 (by Thailand) were opportunistic wars waged at a time when a previously stronger opponent was weakened and distracted by other wars (i.e. Germany in 1914 and France in 1940). The CINC scores do not reflect these historical facts.

⁷ J. David Singer, "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985" *International Interactions* 14 (1987): 115-32; J. David Singer, Stuart Bremer, and John Stuckey, "Capability Distribution, Uncertainty, and Major Power War, 1820-1965," in *Peace, War, and Numbers*, ed. Bruce Russett (Sage, 1972), 19-48.

But the puzzle remains even with the exclusion of the miscoded cases. Excluding the cases mentioned above, there remain six wars with at least a 3:1 disparity in national capabilities between the warring parties. Two of them are especially striking: Japan attacked the U.S. in 1941; China in 1900 went to war with the world's eight most powerful nations all at once. What were the motivations behind these fateful choices? To what extent do the theories in Chapter 1 explain these cases?

Let us now assess the power and limitations of the rationalist mechanisms based on the East Asian case universe.

PART 1: THEORY ASSESSMENT

1.1 Exogenous Enforcement (M1)

Exogenous enforcement reduces the risk of war by increasing the cost of breaking a peaceful agreement or status-quo arrangement. When leaders see an external actor that can act as an enforcer, they are less likely to risk war.

Kim Il-sung proposed to attack South Korea in 1949 but Stalin rejected the plan, fearing that the U.S. would intervene to enforce the agreement between Moscow and Washington on the 38th parallel.⁸ Mao Zedong also dissuaded Kim with a similar argument: the U.S. and Japan might intervene if North Korea crossed the 38th parallel.⁹ Later, Stalin approved Kim's war proposal only after he became more convinced that the U.S. would not intervene.¹⁰ Even then, Stalin made it clear that his approval was conditional on

⁸ William Stueck, *Rethinking the Korean War: A New Diplomatic and Strategic History* (Princeton, NJ: Princeton University Press, 2002), 70.

⁹ Shen Zhihua and Li Danhui, *After Learning to One Side: China and Its Allies in the Cold War* (Washington, DC and Stanford, CA: Woodrow Wilson Center Press and Stanford University Press, 2011), 26.

¹⁰ Kathryn Weathersby concluded that "Stalin's assessment of whether an attack on the South would prompt the Americans to intervene" was "[t]he key factor in [Stalin's] decision for war." Kathryn Weathersby, "The Soviet Role in the Korean War: The State of Historical Knowledge," in *The Korean War in World History*, ed. William Stueck (University Press of Kentucky), 68. Kim argued to Stalin that the U.S. would not intervene because "(1) it would be a decisive surprise attack and the war would be won in three days; (2) there would be an uprising of 200,000 Party members in South Korea; (3) there were guerrillas in the southern provinces of South Korea; and (4) the United States would not have time to participate." Sergei Goncharov, John Lewis, and Xue Litai, *Uncertain Partners: Stalin, Mao, and the Korean War* (Stanford, CA: Stanford University Press, 1993), 144. See also Shen and Li, *After Learning to One Side*, 29; Stueck, *Rethinking the Korean War*, 74. Different theories for Stalin's change of mind are summarized in Shen Zhihua, *Mao, Stalin and the Korean War: Trilateral Communist Relations in the 1950s* (Routledge, 2012), 106-7. The factors influencing Stalin's assessment of a "changed international situation" are analyzed in Weathersby, "Soviet Role," 68-9.

Mao's approval, and that Kim should depend on Mao, not Stalin, to bail him out if the U.S. intervened.¹¹

Conversely, leaders are more likely to risk war when they believe that exogenous enforcement is unlikely. In 1929, reassured that the Great Powers – especially Japan – would not intervene in a Sino-Soviet war, the Soviet Union expanded its military operations to recover its extraterritorial claims in Manchuria.¹² In 1931, conspirators at the Kwantung Army made plans to invade China with the calculation that the League of Nations would be ineffective.¹³ In 1940, Premier Phibun moved his troops into French Indochina partly because he expected no military interference from Britain and the United States, despite their verbal opposition. He was right: neither the British nor the Americans were keen to intervene militarily.¹⁴

Exogenous enforcement can also take a “soft” form through non-military measures, such as economic sanctions, diplomatic isolation and adverse international opinion. It appears that leaders are sensitive to “soft” enforcement even if it does not change their minds about war. Japan was anxious about international opinion on the eve of the 1904 Russo-Japanese War.¹⁵ In the period leading to the 1979 Sino-Vietnamese War, there was an intense competition between Beijing and Hanoi to enlist diplomatic support from other

¹¹ Stueck, *Rethinking the Korean War*, 73.

¹² Cheng Tianfang, *A History of Sino-Russian Relations* (Washington DC: Public Affairs Press, 1957), 154; Zeng Yeying, Huang Daoxuan, and Jin Yilin, *Zhonghua Minguo Shi* [History of Republican China], Vol. 7 (Beijing: Zhonghua Shuju, 2011), 221.

¹³ See Chapter 3.

¹⁴ See Nicholas Tarling, *Britain, Southeast Asia and the Onset of the Pacific War* (Cambridge, UK: Cambridge University Press, 1996), 254.

¹⁵ Ian Nish, *The Origins of the Russo-Japanese War* (London: Longman, 1985), 222-3.

Southeast Asian states: both sides wanted to gain the moral high ground and avoid diplomatic repercussions.¹⁶ That the Sino-Vietnamese War would not turn international opinion against China was one of the arguments for war Deng Xiaoping offered to the Politburo in December 1978.¹⁷

But divergent beliefs about the existence of exogenous enforcement can be dangerous. This divergence led to the Sino-Soviet War in 1929. Nanjing thought that exogenous enforcement would restrain Moscow, but it didn't. Believing that Moscow would not wage war, Nanjing took a hard-line towards the negotiations with Moscow in July-August 1929. Marshal Zhang Xueliang feared that the Soviets would attack Manchuria if negotiations did not bear fruit. But Foreign Minister Wang Zhengting assured Zhang that "while [the Soviet] side has deployed itself in preparation for war, it is bound by the Kellogg-Briand Pact and will definitely not dare to make itself the enemy of the world by declaring war against us."¹⁸ Even after negotiations failed, Chiang Kaishek telegraphed Zhang: "Judging from the current situation, the Soviet Union is only strong on the outside

¹⁶ Nayan Chanda, *Brother Enemy: A History of Indochina since the Fall of Saigon* (San Diego, CA: Harcourt Brace Jovanovich, 1986), 317-20; Robert Ross, *The Indochina Tangle: China's Vietnam Policy, 1975-1979* (New York: Columbia University Press, 1988), 218-23; Ezra Vogel, *Deng Xiaoping and the Transformation of China* (Cambridge, MA: Harvard University Press, 2011), 280-93. Ross suggested that China and Vietnam competed for the goodwill of other Southeast Asian states because "[w]ith the dry season approaching and the growing possibility of war, Hanoi and Beijing sought every additional increment of association in order to deter its adversary from taking hostile action." Ross, *Indochina Tangle*, 219.

¹⁷ King Chen, *China's War with Vietnam, 1979* (Stanford, CA: Hoover Institution, 1987), 87-8.

¹⁸ Zeng, Huang and Jin, *Zhonghua Minguo Shi*, 214-5. U.S. President Herbert Hoover had set the celebration of the Kellogg-Briand Pact of 1928 outlawing war on 24 July 1929. Cheng, *Sino-Russian Relations*, 153.

but weak inside; beside using bribes and threats, it has no other means.”¹⁹ The Soviet Union marched into Manchuria. The Sino-Soviet War offers a useful lesson:

Observation 21: Divergent beliefs about the existence of exogenous enforcement can increase the risk of war.

1.2 Inadvertent Enforcement (M2)

Inadvertent enforcement is a special type of exogenous enforcement. Under inadvertent enforcement, a third-party rival *unintentionally* enforces a peaceful agreement or status-quo arrangement. Leaders try to avoid war when they expect war to create a strategic advantage for their third-party rival. The strategic fear of third-party opportunism dampens the appetite for war. Even hawkish leaders may become war-averse when war makes them vulnerable to a third-party rival. The logic of inadvertent enforcement casts a long shadow in the East Asian case universe.

In July 1929, Marshal Zhang Xueliang pushed for a negotiated settlement with the Soviets, fearing Japanese intervention if a Sino-Soviet war broke out.²⁰ On 26 July, Zhang urged the Nanjing government to adopt his proposal for a peaceful settlement with the Soviets “in order to prevent a third party [Japan] from fishing an advantage [from the

¹⁹ Zeng, Huang and Jin, *Zhonghua Minguo Shi*, 215.

²⁰ *Ibid*, 212.

situation].”²¹ Chiang Kaishek wrote in his diary on 27 July: “the two Zhangs [Marshal Zhang Xueliang and Governor Zhang Zuoxiang] fear that the Soviets will commence hostilities and that Japan will exploit the opportunity [to intervene]; hence they rush for a negotiated settlement without any consideration of the [destabilizing] consequences [on the central government’s policies].”²²

In 1931, Tokyo's strategic fear of Soviet intervention motivated a more cautious approach towards Manchuria vis-à-vis the Kwantung Army, who moved much more aggressively in the belief that the Soviets were unlikely to cause trouble. Likewise in 1937, Tokyo’s initial efforts to contain the July crisis were made in the belief that a war in China would open a strategic advantage for the Soviets in the north.²³

In July 1938, the Changkufeng crisis erupted when Soviet soldiers occupied Changkufeng Hill and started building fortifications. Yet the Japanese Korea Army recommended to Tokyo that military conflict with the Soviet Union should be avoided as Japan was at war with China.²⁴ The Army General Staff at Tokyo concurred.²⁵

²¹ Xue Xiantian, *Minguo Shiqi Zhongsu Guanxi Shi: 1917-1949* [History of Sino-Soviet Relations in the Republican Era: 1917-1949], Vol. 1 (Beijing: Zhonggong Dangshi Chubanshe, 2009), 228.

²² Zeng, Huang and Jin, *Zhonghua Minguo Shi*, 213.

²³ See Chapter 3.

²⁴ See Telegraph 913 from Korea Army HQ to Vice Chief of AGS Hayao Tada, Vice War Minister Hideki Tojo, and Chief of the General Staff of the Kwantung Army Rensuke Isogai, in Michael Kikuoka, *The Changkufeng Incident: A Study in Soviet-Japanese Conflict, 1938* (Lanham, MD: University Press of America, 1988), 72-4. See also Stuart Goldman, *Nomonhan 1939: The Red Army Victory That Shaped World War II* (Annapolis, MD: US Naval Institute Press, 2012), 59; Hata Ikuhiko, “The Japanese-Soviet Confrontation, 1935-1939,” in *Deterrent Diplomacy: Japan, Germany, and the USSR, 1935-1940: Selected Translations from Taiheiyō Sensō E No Michi, Kaisen Gaikō Shi*, ed. James Morley (New York: Columbia University Press, 1976), 142.

²⁵ The AGS replied to Telegraph 913 on 14 July: “The central authorities agreed with the opinion of the local command.” Kikuoka, *Changkufeng Incident*, 79.

Lieutenant-General Tojo Hideki (who later led Japan into the Pacific War) telegraphed the Korea Army HQ on 15 July: “With regard to the Soviet violation of the border in Changkufeng Hill, the central authorities decided to handle it through diplomatic negotiations. Even if the Soviet Union does not withdraw its troops in response to our demand, the local command should still take cautious considerations regarding immediate military action to drive out the Soviets.”²⁶

Conversely, the risk of war increases when inadvertent enforcement is removed. In fact, leaders can sometimes remove inadvertent enforcement endogenously through pacts or alliances.²⁷

Observation 22: War is more likely when the risk of third-party opportunism is reduced through pacts or alliances.

While the Russo-Japanese War was not directly caused by the Anglo-Japanese alliance forged in 1902, the alliance shaped Japan’s strategic circumstances in 1904. Lord Lansdowne, the British Foreign Secretary, reported to King Edward VII in April 1904 that “although [the alliance was] not intended to encourage the Japanese Government to

²⁶ Kikuoka, *Changkufeng Incident*, 80. Major-General Hashimoto Gun, who was at the AGS during the crisis, recalled the incident as “a kind of agitation by the Russians who were needling the Japanese in connection with the warfare in China ... The Supreme Command [AGS] at that time absolutely did not wish any kind of aggravation with the Soviet Union especially before the Wuchang-Hankow operation in China” Kikuoka, *Changkufeng Incident*, 79.

²⁷ Inadvertent enforcement depends on a commitment problem with the third-party rival. Inadvertent enforcement is weakened if the commitment problem is reduced. Depending on the strategic situation, a neutrality pact can be subject to a commitment problem. For example, after the German-Soviet war began in June 1941, Japan had considered renegeing on the Soviet-Japanese Neutrality Pact signed in April 1941. See Records of Imperial Conference, 2 July 1941, in Nobutaka Ike, *Japan’s Decision for War: Records of the 1941 Policy Conferences* (Stanford, CA: Stanford University Press, 1967), 87.

resort to extremities, [it] had, and was sure to have, the effect of making Japan feel that she might try conclusions with her great rival in the Far East, free from all risk of a European coalition such as that which had on a previous occasion deprived her of the fruits of victory.”²⁸ Hence, as Ian Nish pointed out, the Japanese decision for war in 1904 was made “with a fair prospect of a straightforward two-party conflict.”²⁹

And Japanese leaders had tried hard to limit the war to a two-party conflict. The Cabinet resolution on 28 December 1903 emphasized the need to keep China out of the conflict for fear that China’s participation would tempt other powers to join in as well: “[If China starts fighting against Russia], it will tempt the Powers who will try to cash in for themselves by intervening straightaway, while Japan will be so fully involved fighting to the north that she will have no occasion to worry about what is going on elsewhere and will even lose her corner in south China in the end. Needless to say, this is equivalent to our pulling chestnuts out of the fire for someone else ... We must do our utmost to prevent this.”³⁰

Thirty years later, the strategic concern over third-party opportunists continued to haunt Japanese leaders. In April 1941, Japan signed the Soviet-Japanese non-aggression pact with the hope of securing her northern flank to free her hands in the south.³¹ In June 1941, the German-Soviet War suddenly broke out. Japan wanted to take advantage of the

²⁸ Nish, *Origins of the Russo-Japanese War*, 257.

²⁹ *Ibid.*

³⁰ *Ibid.*, 200.

³¹ Walter LaFeber, *The Clash: A History of U.S.-Japan Relations* (New York: W. W. Norton, 1997), 195-6.

fact that the Soviets were tied down by the Nazi invasion to invade Indochina.³² Even then, Tokyo still feared that the Soviets would attack Japan while she was stuck in the south. To further prevent that possibility, Army Chief of Staff Sugiyama Gen argued that Japan must quickly invade the south in the winter season when a Soviet offensive would be ineffective. He warned the imperial conference in September 1941 that “if we should miss this seasonal opportunity, we will not be able to achieve security in the North during our operations in the South.”³³ One of the reservations Japanese leaders had on the eve of the Pacific War was the possibility of Soviet interference, especially if the war dragged over time.³⁴

In early 1962, Chinese leaders were concerned that Chiang Kaishek would launch an attack with American support in the event of a Sino-Indian war. With the ongoing crisis at Laos, Laos could become a corridor for Chiang’s invasion. By late summer 1962, however, the Laos crisis ended with the Geneva Conference and the Taiwan Strait crisis ended with assurances from Washington that it would not support an attack on Mainland China. The new developments convinced China that a Sino-Indian war would be limited to a bilateral war.³⁵

³² Akira Iriye, *The Origins of the Second World War in Asia and the Pacific* (London: Longman, 1987), 140.

³³ Records of Imperial Conference, 6 September 1941, in Ike, *Japan’s Decision*, 142. See also Records of Imperial Conference, 5 November 1941, in Ike, *Japan’s Decision*, 227.

³⁴ Iriye, *Origins of the Second World War*, 177.

³⁵ Allen Whiting, *The Chinese Calculus of Deterrence: India and Indochina* (Ann Arbor, MI: University of Michigan Press, 1975), 85-6; John Garver, “China’s Decision for War with India in 1962,” in *New Directions in the Study of China’s Foreign Policy*, ed. Alastair Iain Johnston and Robert Ross (Stanford, CA: Stanford University Press, 2006), 110-1.

These cases suggest that the strategic fear of third-party opportunism dampens the incentive for war. They highlight the logic of inadvertent enforcement by which a third-party rival becomes an *unintentional* enforcer. For example, in the last two cases, Chiang reduced China's incentive for fighting India and Stalin reduced Japan's incentive for invading Indochina – even though Chiang and Stalin had probably never thought of protecting India or Indochina. Unintentionally, they generated positive security externalities.

But there are analogous cases of exogenous enforcement that fall outside the strict definition of inadvertent enforcement. Here, the third party acts as an *intentional* enforcer because of its obligation arising from an ex-ante alliance or alignment with one of the two parties in conflict. Potential aggressors are restrained through the fear of third-party *obligation*, rather than through the strategic fear of third-party *opportunism* that defines inadvertent enforcement.

In 1950, Washington had reservations that crossing the 38th parallel to annex North Korea would trigger Soviet and Chinese intervention. In 1964-65, Lyndon Johnson and his advisors were frequently haunted by the fear of large-scale Chinese intervention analogous to that in the Korean War.³⁶ In both cases, China was aware that it played a role in the U.S. strategic calculations.

³⁶ Khong Yuen Foong, *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965* (Princeton, NJ: Princeton University Press, 1992), chapter 5.

In 1978, Vietnamese leaders decided to eliminate the Pol Pot regime but were concerned that China would intervene. To deter a Chinese attack from the north while it was tied down in Cambodia, Vietnam concluded a twenty-five year treaty of friendship and mutual cooperation with the Soviet Union in November 1978 before it invaded Cambodia in December 1978.³⁷

The Soviet-Vietnamese treaty increased China's fear of Soviet intervention as Chinese leaders deliberated for war against Vietnam. At the Central Military Commission meeting on 7 December 1978, which decided to launch a war in Vietnam, concerns were raised over a Soviet reprisal attack from the north while China was tied down in Vietnam. The decision for war was eventually made after Chinese leaders were convinced that there were insufficient Soviet forces to mount a large-scale attack on China; and a quick and limited offensive against Vietnam was unlikely to provoke Soviet intervention or adverse international opinion.³⁸ Deng Xiaoping was assured that the risk of Soviet intervention would be very low if the war were brief.³⁹ According to estimates, the Soviet military would require 14 days or more to prepare for an attack, but by then China would have wrapped up its operations in Vietnam.⁴⁰

³⁷ D. R. SarDesai, *Vietnam: Past and Present* (Boulder, CO: Westview Press, 2005), 161. Article Six of the treaty stated: "If either side is attacked or exposed to the threat of attack, the two signatory powers will immediately confer with each other in order to remove this threat and take appropriate and effective steps to safeguard the peace and security of both countries." Chanda, *Brother Enemy*, 321-2. See also Nguyen Van Canh, *Vietnam under Communism, 1975-1982* (Stanford, CA: Hoover Institution Press, 1983), chapter 11.

³⁸ Zhang Xiaoming, "China's 1979 War with Vietnam: A Reassessment," *China Quarterly* 184 (2005): 859; Zhang Xiaoming, "Deng Xiaoping and China's Decision to Go to War with Vietnam," *Journal of Cold War Studies* 12, No. 3 (2010): 14, 25.

³⁹ Vogel, *Deng Xiaoping*, 528.

⁴⁰ Zhang, "China's 1979 War," 859. Ezra Vogel also suggested that "Deng personally believed that because the Soviets were then in the final stages of negotiating the Strategic Arms Limitation Treaty

The 1979 case offers a general lesson:

Observation 23: War is less likely when leaders fear a third-party attack from the rear. But the risk of war increases when leaders believe that war can be finished rapidly before the third party can organize itself for an attack.

China made several moves to reduce the risk of a Soviet attack. China courted the U.S. and quickly restored diplomatic relations, despite U.S. insistence on arms sales to Taiwan.⁴¹ China also announced that its invasion would be limited in its goals and would not threaten the survival of the Vietnamese regime.⁴² The Politburo insisted that regardless of battlefield success, Chinese forces must halt and withdraw after taking two provincial capitals at the Vietnamese border.⁴³ Furthermore, the combat operations would not involve the use of aircrafts, as an air war would increase the risk of Soviet intervention.⁴⁴ All Soviet troop movements would also be kept under close surveillance by Chinese intelligence.⁴⁵

(SALT) II with the United States, they would be reluctant to disrupt the negotiations by engaging in a land war in Asia." Vogel, *Deng Xiaoping*, 527.

⁴¹ Ross, *Indochina Tangle*, 228, 254.

⁴² *Ibid.*, 229.

⁴³ Zhang, "Deng Xiaoping and China's Decision," 2.

⁴⁴ Vogel, *Deng Xiaoping*, 528.

⁴⁵ *Ibid.*

Inadvertent enforcement at the domestic level

As the 1931 case (Chapter 3) shows, inadvertent enforcement can also operate at the domestic level. The logic is similar except that the third-party rival is a *domestic* rival: Leaders are less likely to choose war if it creates a potential political vulnerability that their domestic rivals cannot commit *not* to exploit.

Until November 1964, Lyndon Johnson tried to avoid a major escalation of U.S. involvement in Vietnam. A key consideration behind Johnson's restraint was the presidential election in November.⁴⁶ He recognized that fighting a war in Vietnam would yield an electoral advantage to his Republican opponent. Hence, his presidential campaign emphasized a desire to limit U.S. involvement in Vietnam, casting Johnson as a prudent contrast vis-à-vis the hawkish Barry Goldwater. In a campaign speech, Johnson declared: "Sometimes our folks get a little impatient ... Sometimes they rattle their rockets some, and they bluff about their bombs. But we are not about to send American boys 9 or 10,000 miles away from home to do what Asian boys ought to be doing for themselves."⁴⁷ Robert Schulzinger argued that the end of the election "removed one brake on a more assertive U.S. policy in Vietnam."⁴⁸

⁴⁶ Fredrik Logevall, *The Origins of the Vietnam War* (Longman, 2001), 59, 65; Robert Schulzinger, *A Time for War: The United States and Vietnam, 1941-1975* (New York: Oxford University Press, 1997), 134, 138, 144; Gerard DeGroot, *A Noble Cause? America and the Vietnam War* (Longman, 2000), 127.

⁴⁷ George Herring, *America's Longest War: The United States and Vietnam, 1950-1975* (New York: McGraw-Hill, 2002), 146; Schulzinger, *Time for War*, 155.

⁴⁸ Schulzinger, *Time for War*, 163.

Inadvertent enforcement reversed

But the same mechanism can also work in the other direction: Leaders are *more* likely to risk war if their war-avoidant behavior would create political vulnerabilities that domestic rivals cannot commit *not* to exploit.

In the Boxer War, Beijing's decision to fight eight Great Powers simultaneously is puzzling in the light of China's national interests. But that seemingly irrational decision might have made sense for Empress Dowager Cixi, who usurped imperial power in 1898. She had misperceived the naval attacks on the Dagou Forts as part of a Western plot to depose her and reinstate Emperor Guangxu.⁴⁹ If the powers marched into Beijing, her survival would be threatened as power shifted to the emperor's supporters. She announced to the Imperial Court on June 17: "Everyone here has learnt about today's event. And I have to declare war for the sake of saving the country. ... If the war did not go well and the country could not be saved, [you] should not blame me alone and say that the Empress Dowager has lost the three-hundred-year-old dynasty founded by the ancestors."⁵⁰

⁴⁹ Xiang Lanxin, *The Origins of the Boxer War: A Multinational Study* (London: RoutledgeCurzon, 2003), 256, 355. In Chester Tan's narrative of the imperial audience on 17 June, "the Empress Dowager announced that she had just received from the Powers a demand covering four points: (1) a special place to be assigned to the Emperor for residence; (2) all revenues to be collected by the foreign ministers; (3) all military affairs to be committed to their hands ... The fourth point, which the Empress Dowager omitted to mention in the Imperial Council, was to restore the rule to the Emperor." Chester Tan, *The Boxer Catastrophe* (New York: Octagon Books, 1967), 73. See also Joseph Esherick, *The Origins of the Boxer Uprising* (University of California Press, 1987), 302. Some historians suggested that the hawkish Prince Duan had forged the ultimatum, which was not issued by the western powers. For a detailed discussion of the ultimatum, see Xiang, *Boxer War*, 294-6.

⁵⁰ Xiang, *Boxer War*, 296. The "event" might have referred to the Dagou attacks or the alleged ultimatum, or both.

Political vulnerabilities may also be created through domestic audience costs. This creates a direct connection between domestic audience costs and the risk of war.

In 1940, France approached Thailand for a non-aggression pact with the diplomatic understanding that Thailand could recover some of its lost territories in French Indochina as a concession for signing the pact. The high hopes were dashed when Germany invaded France and the new Vichy government reversed course. Premier Phibun was trapped by the uncompromising position adopted by Vichy France: Recovering the lost territories would give a huge boost to his political prestige, but “losing” them again would cause an enormous public backlash.⁵¹ Historians studying this period reported that “there were domestic political rivals waiting to take advantage,” and Phibun had expressed fears “that he might be eclipsed by another military figure ... if he did not please the ultra-nationalists in the officer corps.”⁵² The British ambassador reported from Bangkok: “Military party and 'lingos' are in full control and the Prime Minister has joined them ... the [French] bombing of Nakhon Panom has roused them to a frenzy of indignation. Mass demonstrations demanding war continue to be held, and are even more insistent than before.”⁵³ The political costs of backing down were high with nationalistic sentiments at a feverish pitch. When the French Indochinese authorities refused to be moved after months of negotiations, Phibun launched the Franco-Thai War.

⁵¹ David Wyatt, *Thailand: A Short History* (New Haven, CT: Yale University Press, 2003), 245; Joseph Wright, *The Balancing Act: A History of Modern Thailand* (Oakland, CA: Pacific Rim Press, 1991), 106; E. Thadeus Flood, “The 1940 Franco-Thai Border Dispute and Phibuun Sonkhraam's Commitment to Japan,” *Journal of Southeast Asian History* 10, No. 2 (1969): 310, 323.

⁵² Flood, “Franco-Thai Border Dispute,” 325; Richard Aldrich, *The Key to the South: Britain, the United States, and Thailand During the Approach of the Pacific War, 1929-1942* (Kuala Lumpur: Oxford University Press, 1993), 268.

⁵³ Aldrich, *Key to the South*, 280-1.

Nationalistic sentiments were also riding high in India on the eve of the 1962 Sino-Indian War. A classified internal history by India's Ministry of Defense concluded that the Nehru government was pressed to push forward "under tremendous pressure from the Parliament, the Press and the public." "Sadly unfamiliar with military matters, these vociferous and strident opinions accused the Government of lack of will, and insisted that the Indian territory already occupied by China must be liberated at the earliest, if necessary by armed might. The debates in Parliament and the editorials in the national dailies from 1960 to 1962 make shocking reading today."⁵⁴

In the Cambodian case during 1977-78, mass nationalism was a product of government policy. Khieu Samphan, the head of state under the Pol Pot regime, explained the regime's anti-Vietnamese stance: "To unite our compatriots through the Party, bring our workers up to their highest level of productivity, and to make the youths' [troops] ardor and valor in combat even greater, the best thing we can do was incite them to hate the youths [pejorative for Vietnamese] more and more everyday. Our bang-phaaun [older and younger brothers and sisters] are willing to make any sacrifice the minute we wave the "Hate Vietnam" flag in front of them."⁵⁵ Mounting a nationalistic flag, the Pol Pot regime escalated its border attacks on Vietnam in 1977-78. Cambodia refused to compromise

⁵⁴ P. B. Sinha and A. A. Athale. *History of the Conflict with China, 1962* [Restricted], ed. S. N. Prasad (New Delhi: History Division, Ministry of Defence, 1992), xxiii. Steven Hoffmann reported that "[f]or political reasons (i.e., so as not to forfeit public confidence and to retain diplomatic flexibility), [Premier] Nehru felt that his government should have the army do its best to push the Chinese out ... [Defense Minister] Krishna Menon also believed that Indian public opinion would not forgive any surrendering of territory." Steven Hoffmann, *India and the China Crisis* (University of California Press, 1990), 128.

⁵⁵ Nguyen-Vo Thu-Huong, *Khmer-Viet Relations and the Third Indochina Conflict* (Jefferson, NC: McFarland & Company, 1992), 96.

even after a major putative attack by Vietnam in December 1977.⁵⁶ On 22 January 1978, Phnom Penh publicly announced that its conflict with Vietnam was “not a dispute which can be resolved through compromise and negotiation.”⁵⁷

The last three cases yield a common lesson:

Observation 24: Nationalistic fervor increases the domestic audience costs for leaders and reduces their incentive to back down in an international confrontation.

The road to the Vietnam War was also taken in the shadow of audience costs. As Larry Berman and Stephen Routh pointed out, “[o]nce policies were implemented under Eisenhower that supported the Diem regime, it became increasingly difficult for subsequent presidents to roll back those obligations because the costs of losing increased with each presidential promise to maintain the government in the South.”⁵⁸ The Johnson

⁵⁶ David Chandler argued that “[t]he Vietnamese had invaded Cambodia intending to spur negotiations; instead, it was as if they had poked a beehive with a stick. Cambodia's leaders were now convinced that Vietnam wanted to swallow their country. They began preparing for a holy war.” David Chandler, *Brother Number One: A Political Biography of Pol Pot* (Boulder, CO: Westview Press, 1999), 143.

⁵⁷ Ben Kiernan, *The Pol Pot Regime: Race, Power, and Genocide in Cambodia under the Khmer Rouge, 1975-79* (New Haven, CT: Yale University Press, 2008), 388. In his biography of Pol Pot, David Chandler speculated that “[p]erhaps [Pol Pot] saw fighting with Vietnam as a way of producing solidarity and lessening unrest.” It is also probable that Pol Pot “saw Vietnamese “aggression” as connected with “treasonous elements” inside the Cambodian Communist party.” Chandler, *Brother Number One*, 142.

⁵⁸ Larry Berman and Stephen Routh, “Why the United States Fought in Vietnam,” *Annual Review of Political Science* 6 (2003): 193. Logevall argued that until late February 1965, the domestic political costs of reversing the course in Vietnam were potentially manageable: “At home, a negotiated withdrawal would have won the support of powerful voices in elite public opinion, including Johnson's own Democratic leadership in Congress. But the president and his men nevertheless chose war. ... Time and again over the years they had publicly vowed American steadfastness in the struggle, had declared the outcome in Vietnam to be crucial to US security, and they were loathe to change their tune now.” Logevall, *Origins of the Vietnam War*, 92. See also Fredrik Logevall,

administration was sensitive to domestic audience costs as it progressively deepened the U.S. commitment in Vietnam.⁵⁹ National Security Adviser McGeorge Bundy used the Korean War analogy to highlight the enormous political backlash on the Johnson administration “if we should be the first to quit in Saigon”.⁶⁰ Johnson feared that showing weakness in Vietnam would shatter his domestic standing and strengthen those who opposed his domestic policies. He predicted that “[i]f I don’t go in now and they show later that I should have ... they’ll . . . push Vietnam up my ass every time.”⁶¹

Johnson’s concern was to avoid domestic political losses, rather than to seek domestic political benefits. But escalation decisions can be taken under a “two-sided” temptation – when there are simultaneous expectations of domestic political benefits from escalation and domestic political losses from backing down. The U.S. decision to cross the 38th parallel in the Korean War was made under a two-sided temptation. The decision was expected to yield a large domestic political dividend for the Truman administration; on the other hand, a halt at the 38th parallel would provide political ammunition for the Republican Party in the impending mid-term elections.⁶² The Truman administration was particularly tempted by the concept of “rollback” after suffering intense attacks from domestic opponents for its alleged weakness in the face of the communist menace.⁶³

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

⁵⁹ According to Robert Schulzinger, “Johnson told the Chairman of the Joint Chiefs of Staff that he did not want to lose South Vietnam before the election, but he did not “want to get the country into war” before the vote either.” Schulzinger, *Time for War*, 138.

⁶⁰ Schulzinger, *Time for War*, 132.

⁶¹ Herring, *America’s Longest War*, 136.

⁶² Rosemary Foot, *The Wrong War: American Policy and the Dimensions of the Korean Conflict, 1950-1953* (Ithaca, NY: Cornell University Press, 1985), 67-74.

⁶³ Peter Lowe, *The Korean War* (New York: St. Martin’s Press, 2000), 38.

The U.S. crossing of the 38th parallel was a disaster. It led to the Chinese overt entry into the Korean War. Like Truman, Mao faced a two-sided temptation in the decision to intervene. Allowing the U.S. to eliminate the communist regime in Korea would embolden the reactionaries within China. By contrast, the decision to fight the U.S. would yield the opportunity to purge reactionaries and mobilize the Chinese masses to realize his revolutionary domestic goals.⁶⁴ Together with the decision to enter the Korean War, Mao launched intensive mass mobilization campaigns under the banner of “Resist America, Assist North Korea” that fundamentally transformed the political and social landscape in China.⁶⁵

When inadvertent enforcement did not work

In three cases, the decision-making seems to contradict the logic of inadvertent enforcement. One of them is China’s entry into the Korean War in 1950. Mao was not deterred by the possibility of a coastal attack from Taiwan while the People’s Liberation Army was tied down in Korea. Here, inadvertent enforcement might have been

⁶⁴ Chen Jian, *China's Road to the Korean War: The Making of the Sino-American Confrontation* (New York: Columbia University Press, 1994), 83, 160; Stueck, *Rethinking the Korean War*, 108; Thomas Christensen, *Useful Adversaries: Grand Strategy, Domestic Mobilization, and Sino-American Conflict, 1947-1958* (Princeton, NJ: Princeton University Press, 1996).

⁶⁵ Chen Jian, “Reorienting the Cold War: The Implications of China's Early Cold War Experience, Taking Korea as a Central Test Case,” in *The Cold War in East Asia, 1945-1991*, ed. Tsuyoshi Hasegawa (Washington, DC and Stanford, CA: Woodrow Wilson Center Press and Stanford University Press, 2011), 88-9; Chen, *China's Road to the Korean War*, 137-41, 190-4. Chen Jian reported that at the end of the Korean War, “organized resistance to the new regime had been destroyed, land in the countryside had been redistributed and the landlord class had been eliminated, the national bourgeoisie was under the tight control of the Communist state, and the petit bourgeois intellectuals had experienced the first round of Communist reeducation.” Chen, “Reorienting the Cold War,” 89.

suspended for two reasons. First, Mao did not seem to anticipate a long war. Chen Jian reported that Mao and his colleagues were expecting “a regional war, a conventional war, a short war, and a limited war.”⁶⁶ If so, this provides further evidence for Observation 23 discussed earlier. Second – and perhaps more importantly – Mao believed that China would become even more vulnerable in the future if North Korea collapsed and U.S. forces lined the Chinese border.⁶⁷ At that time, a U.S. attack in North China from Korea coordinated with an attack in East China from Taiwan would be much more perilous. Hence, it is possible that the fear of a two-front war in the present was overridden by the fear of a *worse* two-front war in the future.

The logic of inadvertent enforcement also did not seem to work on the French Indochinese side in late 1940 before the Franco-Thai War. Fearing a Japanese attack, the Indochinese authorities should have secured themselves by persuading Thailand to ratify the anti-aggression pact, through making concessions on the border readjustments badly wanted by Premier Phibun. Instead, the Indochinese authorities balked. Their intransigent stance might be due to a principal-agent problem, which will be discussed at the end of this chapter.

During the Changkufeng crisis in 1938, Japanese leaders were split into two camps of thought. One camp believed that it was imprudent to enter into a conflict with the Soviet Union, given that the Japanese army was tied down in China. The other camp, however, reasoned that Japan must act firmly to deter the Soviets from interfering at the border

⁶⁶ Chen, *China's Road to the Korean War*, 178.

⁶⁷ See the subsequent section (Section 1.4) on the Chinese decision to intervene in the Korean War.

while the Japanese were fighting in China. An offensive plan based on the latter view was eventually approved at the five ministers' conference on 19 July, with the prime minister, the army minister, and the finance minister overriding the objections of the foreign minister and the navy minister.⁶⁸ The draft proposal for imperial sanction explained: "In view of the critical combat situation in progress in China, aggravation with the Russians at the border should be avoided by all means. However, if [we yield to the Soviets at Changkufeng], it is feared that similar incidents will occur one after another. Moreover, the position of Changkufeng Hill is considered to be least dangerous in escalating the dispute with large forces. Accordingly, it is rather a right judgment to make a one-shot attack in order to intimidate the fighting spirit of the Soviet Union. Without question, the offensive action should be carried out by surprise for limited objectives. ..."⁶⁹ The proposal, however, was rejected by the Emperor who allegedly chastised the army minister for the reckless behavior of the army since the 1931 Mukden crisis.⁷⁰ Nonetheless, the 1938 and 1950 cases suggest a common lesson:

⁶⁸ Goldman, *Red Army Victory*, 61.

⁶⁹ Kikuoka, *Changkufeng Incident*, 83-5. In 1939, Colonel Inada Masazumi (Chief of the AGS Operations Section and the original advocate of the offensive plan) claimed that his intent was to conduct an *iryoku teisatsu* (tactical probe or reconnaissance-in-force) to test Soviet intentions. The fear that the Soviet Union would exploit Japan's involvement in China to attack Manchuria had led to Tokyo's hesitation on launching a massive military operation in Wuhan. Inada's *iryoku teisatsu* was designed to test if the Soviets would respond at Changkufeng with a major attack in Manchuria, before Japan committed herself heavily in the Wuhan operation. Goldman, *Red Army Victory*, 60; Kikuoka, *Changkufeng Incident*, 87; Alvin Coox, *Nomonhan: Japan against Russia, 1939*. Vol. 1, (Stanford, CA: Stanford University Press, 1990), 124; Alvin Coox, *The Anatomy of a Small War: The Soviet-Japanese Struggle for Changkufeng-Khasan, 1938* (Westport, CT: Greenwood Press, 1977), chapter 4.

⁷⁰ Coox, *Nomonhan*, 130-1; Kikuoka, *Changkufeng Incident*, 86.

Observation 25: The risk of war can increase when leaders expect that military inaction will make them even more vulnerable to their third-party rival in the future.

1.3 Endogenous Enforcement (M3)

Under endogenous enforcement, at least one of the parties in conflict tries to manage the potential shift in power that generates the commitment problem and the risk of war. The reduction or elimination of the power shift correspondingly reduces or eliminates the commitment problem and the risk of war.

American calculations on the oil embargo in 1940-41 echoed endogenous enforcement through tactical restraint. In July 1940, the Roosevelt administration decided against imposing an oil embargo on Japan for fear that it would provoke rather than deter the Japanese.⁷¹ The argument resonated with Franklin Roosevelt and the U.S. Navy in 1940-41; Roosevelt had in fact articulated the argument publicly in July 1941.⁷² He had also reminded the British ambassador Lord Halifax that the U.S. could not fight both a

⁷¹ Scott Sagan, "The Origins of the Pacific War," in *The Origin and Prevention of Major Wars*, ed. Robert Rotberg, Theodore Rabb, and Robert Gilpin (Cambridge, UK: Cambridge University Press, 1989), 328; Waldo Heinrichs, *Threshold of War: Franklin D. Roosevelt and American Entry into World War II* (New York: Oxford University Press, 1988), 134; Jonathan Utley, *Going to War with Japan, 1937-1941* (Knoxville, TN: University of Tennessee Press, 1985), 83-101. Utley argued that the imposition of the oil embargo on Japan in July 1941 was contrary to Roosevelt's intention. Marc Trachtenberg suggested otherwise. Utley, *Going to War with Japan*, 151-6; Marc Trachtenberg, *The Craft of International History: A Guide to Method* (Princeton, NJ: Princeton University Press, 2006), chapter 4.

⁷² Sagan, *Pacific War*, 334.

European war and a Pacific war at the same time.⁷³ In late November 1941, U.S. diplomats devised a modus vivendi proposal that would lift the oil embargo in a three-month “experiment” if Japan withdrew its troops from southern Indochina. Washington believed that the modus vivendi, which would stem the rapid deterioration in Japan’s relative power, could help to put off a war in at least the near future.⁷⁴

Armed clashes between Chinese and Indian troops led to high tension at the Sino-Indian border in mid-1959. After a deadly clash at Kongka Pass in October, Mao told Premier Zhou Enlai to propose a mutual withdrawal of 20 kilometers each from the line of actual control; he also suggested that China should unilaterally withdraw even if India did not agree.⁷⁵ This implied a halt at the present status quo, and that India’s claims and bargaining power would not be weakened through Chinese encroachment at the border. Sino-Indian tension declined over the subsequent months until India launched its “forward policy” in November 1961.

Conversely, the lack of strategic or tactical restraint often leads to trouble. This is the logic of endogenous enforcement in its reversed manifestation: if one of the parties in conflict takes action that can potentially shift the relative power in its favor, the commitment problem is sharpened and the risk of war increased.

⁷³ Heinrichs, *Threshold of War*, 134.

⁷⁴ Iriye, *Origins of the Second World War*, 180. The British, Chinese, and Dutch governments swiftly pressurized Washington to drop the proposal.

⁷⁵ M. Taylor Fravel, *Strong Borders, Secure Nation: Cooperation and Conflict in China's Territorial Disputes* (Princeton, NJ: Princeton University Press, 2008), 84-5; Garver, “China’s Decision for War with India,” 106.

Hence, in 1941, the U.S. oil embargo motivated Japan's decision to launch the Pacific War, as discussed in Chapter 2. The oil embargo would greatly strengthen American bargaining power in the future: When Japanese ships stop moving, the U.S. would be in a very strong bargaining position. But precisely so, the Japanese were desperate to stem the power shift *before* their ships ran out of fuel. Washington recognized this. Hence, in late November 1941, it tried to moderate the power shift and calm Japanese desperation by devising the *modus vivendi*. But the *modus vivendi* fell apart, for reasons to be discussed later in this chapter.

From November 1961, India's implemented its "forward policy" that progressively weaved a growing network of border posts and sentries in disputed territories not controlled by China. The occupation by *fait accompli* would strengthen territorial claims and relative bargaining power in India's favor – and in China's disfavor. China responded in kind to defend its territorial claims. Both sides refused to yield until China went to war in October 1962.

In 1964-65, North Vietnam escalated its operations against the south, gambling that if South Vietnam collapsed before the U.S. directly intervened in Vietnam, the U.S. would have no choice but to withdraw.⁷⁶ North Vietnamese leaders had hoped to avoid a direct

⁷⁶ Ang Cheng Guan, *The Vietnam War from the Other Side: The Vietnamese Communists' Perspective* (London: RoutledgeCurzon, 2002), 75; Herring, *America's Longest War*, 132, 146; Logevall, *Origins of the Vietnam War*, 62; DeGroot, *Noble Cause*, 130. DeGroot also highlighted internal disagreement at the Ninth Plenum of the Vietnamese Communist Party in 1963: "An aggressive faction wanted quick deployment of the PAVN [People's Army of Vietnam], in order to topple the crumbling Saigon regime. The other side was satisfied with the pace of the revolution and warned that intervention by the PAVN would annoy Moscow, worry Beijing, and alarm Washington." DeGroot, *Noble Cause*, 126.

confrontation with the Americans.⁷⁷ But the rapid deterioration in the military situation alarmed Washington and it responded by intensifying its operations to prevent the disintegration of South Vietnam. The 1962 and 1965 cases suggest a common lesson: Attempts by one side to achieve a favorable power shift do not succeed easily. They tend to trigger countermoves by the other side, leading to a conflict spiral that makes both sides worse off than before.⁷⁸

Vietnam formally aligned with the Soviet Union in November 1978 and launched a large-scale invasion of Cambodia in December 1978. If Vietnam successfully liquidate Pol Pot's forces and consolidate its domination of Indochina, the relative power balance would shift against China in favor of the Soviet-Vietnamese front, with China trapped between a Soviet superpower in the north and a Vietnamese regional hegemon in the south. By going to war, Beijing hoped to prevent the full liquidation of Pol Pot's forces and to break the Soviet-Vietnamese encirclement of China.⁷⁹

The cases reflect the logic of endogenous enforcement in its reversed manifestation, which is a derivative from the commitment theory of war. Let us now turn to this theory.

⁷⁷ Ibid. See Nguyen Lien-Hang, *Hanoi's War: An International History of the War for Peace in Vietnam* (Chapel Hill, NC: University of North Carolina Press, 2012), chapter 2.

⁷⁸ Robert Jervis, *Perception and Misperception in International Politics* (Princeton, NJ: Princeton University Press, 1976), chapter 3.

⁷⁹ Ross, *Indochina Tangle*, chapter 7; Steven Hood, *Dragons Entangled: Indochina and the China-Vietnam War* (Armonk, NY: M. E. Sharpe, 1992), 50-7; William Duiker, *Vietnam: Revolution in Transition* (Boulder, CO: Westview Press, 1995), 208; Zhang, "Deng Xiaoping and China's Decision," 13, 28.

1.4 Commitment Theory of War

This theory suggests that war is more likely when one side expects a significant increase in its future vulnerability that the opponent cannot commit *not* to exploit. The theory combines the window theory with the commitment-problem explanation. The cause of war lies in the perceived potential vulnerability generated by an impending power shift, which the opponent cannot commit *not* to exploit. This theory resonates in the 1931 and 1937 cases discussed in Chapter 3, and in several other cases in East Asia.

Tokyo launched the Russo-Japanese War because Russian ambitions in Manchuria were seen as a major threat to Korea and the security of Japan itself.⁸⁰ If Russia took Korea, Russia would command what many Japanese saw as the “dagger pointed at the heart of Japan.”⁸¹ Meanwhile, Russia’s progress with the Trans-Siberian and Chinese Eastern railways gave her a growing military advantage, since the railways would be able to expedite the deployment of Russian troops to the east in large numbers.⁸² The imperial conference on 12 January 1904 agreed that “Russia had made no adequate concession over Korea and had even refused to enter into negotiations over Manchuria, while she was at the same time trying to build up her military strength there.”⁸³ By end January, when Russia failed to respond to the final Japanese proposal after three weeks, Tokyo

⁸⁰ Shumpei Okamoto, *The Japanese Oligarchy and the Russo-Japanese War* (New York: Columbia University Press, 1970), 102.

⁸¹ Michael Auslin, “Japanese Strategy, Geopolitics and the Origins of the War, 1792-1895,” in *The Russo-Japanese War in Global Perspective: World War Zero*, ed. John Steinberg et al. (Leiden: Brill Academic Publishers, 2005), 18. For a summary of the historical background on Japan’s geopolitical thinking, see Auslin, “Japanese Strategy,” 17-21.

⁸² Nish, *Origins of the Russo-Japanese War*, 203.

⁸³ *Ibid.*, 206.

concluded that Russia was stalling for time, and that further delay would only strengthen Russia's military position.⁸⁴ Tokyo refused international mediation on the same grounds: "Russia, if she asks for mediation, does so only to gain time in order to consolidate her position in the east. Consequently mediation, if it were now to be arranged, would result in [an] advantage to Russia."⁸⁵

Japanese leaders knew that war would be very costly and risky.⁸⁶ But Tokyo decided for war despite the costs and risks involved. Ito Hirobumi, who was previously the leader of the pro-peace faction, explained in retrospect in February 1904: "There is no question but that Russia's aim was from the start to increase her military and naval forces and then reject Japan's demands. In this way she could fulfill her ambitions in Manchuria and Korea without interference. This being so, if Japan does not now go to war and defend her threatened interests, she will eventually have to kowtow to the Russian governor of one of her frontier provinces."⁸⁷

China went to war in 1950 to prevent the Korean peninsula from falling into American control. China feared that it would lose a critical strategic buffer and be exposed to a constant threat of an American attack from the northeast.⁸⁸ Once the U.N. forces took

⁸⁴ Okamoto, *Japanese Oligarchy*, 100; Nish, *Origins of the Russo-Japanese War*, 211

⁸⁵ Nish, *Origins of the Russo-Japanese War*, 223.

⁸⁶ Okamoto, *Japanese Oligarchy*, 101. The army saw a fifty-fifty chance; the navy estimated that half its forces would be lost.

⁸⁷ Nish, *Origins of the Russo-Japanese War*, 207.

⁸⁸ Shen and Li, *After Leaning to One Side*, 48; Chen, *China's Road to the Korean War*, 159, 184; Thomas Christensen, *Worse Than a Monolith: Alliance Politics and Problems of Coercive Diplomacy in Asia* (Princeton, NJ: Princeton University Press, 2011), 91. At the crucial Politburo meeting on 4 October, Mao highlighted: "If the U.S. imperialists win, they will be complacent and pose a threat for us. We cannot but help Korea in the form of volunteers." Premier Zhou Enlai

North Korea, the Americans would command a much stronger military position that it could use against China at any time of its choosing. Diplomatic efforts to reassure China about the U.N. unification of Korea cut no ice in Beijing.⁸⁹ As Premier Zhou Enlai later explained, “The information we obtained was that they [the Americans] wanted to leave us alone at first. They will not deal with China until they cross the 38th Parallel. We saw through this fraud, and so we made a statement on September 30: We cannot ignore the invasion of Korea by the imperialist aggressors of the United States.”⁹⁰

In 1949-50, Kim wanted to attack South Korea sooner than later partly because South Korea could become stronger in the future. Once the land reform bill of June 1949 was implemented in the following summer, the already bleak prospects of guerilla action in South Korea would decline even further.⁹¹ Evidence of closer interactions between South Korea and Japan also triggered fears that a powerful Korea-Japan-U.S. front would emerge once Japan regained its political independence and military power.⁹²

In 1964-65, South Vietnam was becoming weaker and weaker, while the communists in the North were becoming more and more vigorous. In this context, as Secretary of

agreed: “If the American imperialists push North Korea down, it will be no good for peace and they will be more arrogant.” Shen and Li, *After Leaning to One Side*, 35. These records are based on the recollections of the meeting participants. See Christensen, *Worse Than a Monolith*, 90.

⁸⁹ Christensen, *Worse Than a Monolith*, 87.

⁹⁰ Shen and Li, *After Leaning to One Side*, 49. Niu Jun argued that U.S. intervention in Korea and the increase in U.S. military presence in the Taiwan Strait and Southeast Asia had “led Chinese leaders to think that the United States was about to engage in a strategic expansion against China”. Niu Jun, “The Birth of the People’s Republic of China and the Road to the Korean War,” in *The Cambridge History of the Cold War*, Vol. 1, ed. Melvyn Leffler and Odd Arne Westad (New York: Cambridge University Press, 2010), 238.

⁹¹ William Stueck, *The Korean War: An International History* (Princeton, NJ: Princeton University Press, 1995), 31.

⁹² Stueck, *Korean War: An International History*, 31-2; Christensen, *Worse Than a Monolith*, 54-5.

Defense Robert McNamara argued, any settlement to divide or neutralize South Vietnam “would inevitably mean a new government in Saigon that would in short order be communist dominated.”⁹³ Hence, the U.S. continuously deepened its commitments in Vietnam, hoping to reverse or at least stabilize the situation. By the summer of 1965, however, the military situation in South Vietnam continued to worsen despite Operation Rolling Thunder and the introduction of U.S. ground forces.⁹⁴ In late July 1965, Lyndon Johnson decided for war after a series of urgent meetings held under the framework of a drastically deteriorating situation on the ground.⁹⁵

In 1965, the concern was not simply over a shift in relative power, but also over the *speed* of the shift. When relative power is shifting *rapidly* in one’s disfavor, the decision for war takes on a new urgency. This connects to Observation 2 derived from the 1941 case analyzed in Chapter 2. The perception of a rapid deterioration in the military balance also figured prominently on the eve of the Third Sino-Japanese War (Chapter 3). Likewise, Chinese leaders decided for war in 1950 in the shadow of a rapid power shift. Mao told the Politburo on 2 October that “the question now is not whether or not but how fast we should send troops to Korea. One day’s difference will be crucial to the whole situation.”⁹⁶ On 5 October, Mao argued that it would be too late for Chinese intervention

⁹³ Schulzinger, *Time for War*, 133.

⁹⁴ Herring, *America’s Longest War*, 151; Logevall, *Origins of the Vietnam War*, 81.

⁹⁵ Gary Hess, *Presidential Decisions for War: Korea, Vietnam, the Persian Gulf, and Iraq* (Baltimore, MD: John Hopkins University Press, 2009), 101.

⁹⁶ Chen, *China’s Road to the Korean War*, 173.

once U.N. forces reached the Yalu River. He told Marshal Peng Dehuai: “We have to move forward immediately.”⁹⁷

In 1940, Thailand invaded the disputed territories in French Indochina with the belief that a Japanese move was imminent.⁹⁸ If Japan took those territories, Thailand would have much less bargaining power vis-à-vis the Japanese to recover their lost territories. Premier Phibun felt that he had to act quickly to forestall the Japanese. Without an impending Japanese move, Phibun said that he would still order an invasion to satisfy public sentiments, but the timeframe could be pushed back to “March or April of next year at the latest”.⁹⁹

Costly defense

Earlier, we have considered a manifestation of endogenous enforcement through strategic or tactical restraint by the rising power. Endogenous enforcement can also be achieved through defensive measures by the declining power. For example, if the U.S. cannot commit not to attack China from North Korea, China could reduce the commitment problem by increasing its defenses at its northeast border, instead of going to war with the U.S. in 1950. Why did China choose war?

⁹⁷ Ibid, 183.

⁹⁸ Wyatt, *Thailand*, 245; Aldrich, *Key to the South*, 282.

⁹⁹ Aldrich, *Key to the South*, 282.

A key reason is that a unilateral reduction of the commitment problem would be too costly. As Zhou Enlai later explained, even if a U.S. attack on China was not forthcoming in the short term, the cost of maintaining effective defenses along the Yalu river and relocating key industries from vulnerable positions would be too high for China in the long term.¹⁰⁰ “How many troops are needed to guard the Yalu River of one thousand kilometers? Moreover, we have to wait there year after year without knowing when the enemy will come.”¹⁰¹ A similar argument was heard in Washington before U.S. troops were authorized to cross the 38th parallel. If the retreating North Korean forces were annihilated, the costs of maintaining security in South Korea would be much reduced in the long-term.¹⁰²

Observation 26: The risk of war increases when leaders expect the short-term costs of active offense to be lower than the long-term costs of passive defense.¹⁰³

Likewise, Hanoi invaded Cambodia in 1978 with the intention to achieve long-term security at its western flank.¹⁰⁴ Constant attacks by Cambodian troops at the border were costly to Vietnam and had escalated in scale since 1977. Diplomatic negotiations were

¹⁰⁰ Christensen, *Worse Than a Monolith*, 91.

¹⁰¹ Hao Yufan and Zhai Zhihai, “China’s Decision to Enter the Korean War: History Revisited,” *China Quarterly* 121 (1990): 104.

¹⁰² Christensen, *Worse Than a Monolith*, 64.

¹⁰³ See also Robert Powell’s model of costly deterrence, which suggests that states choose to fight earlier than later when the expected cost of deterrence in the long term exceeds the expected cost of eliminating the threat altogether in the present. Robert Powell, “War as a Commitment Problem,” *International Organization* 60, No. 1 (2006): 192-4.

¹⁰⁴ Duiker, *Vietnam: Revolution in Transition*, 204.

moving nowhere. A decisive attack on Cambodia would liquidate the Pol Pot regime before it strengthened itself further with Chinese assistance.¹⁰⁵

One of Deng's justifications for waging the 1979 Sino-Vietnamese War applied the same logic. Deng was about to launch the "Four Modernizations," a momentous set of major domestic reforms. Yet Vietnam's unabated aggression and hegemonic aspirations would distract China's domestic reforms.¹⁰⁶ Deng argued that China must have a safe and stable environment while the Four Modernizations were implemented, without the Soviets threatening China from the north and the Vietnamese from the south, which would "wedge us in."¹⁰⁷ This case suggests that a commitment problem can be worsened when leaders expect an impending *domestic* vulnerability in the future:

Observation 27: The risk of war increases in the present when leaders expect domestic changes that would increase the costs of war in the future.

These cases suggest conditions under which endogenous enforcement through defensive measures is infeasible. But a puzzle remains for the absence of endogenous enforcement through tactical restraint: Rising powers have an incentive to reduce the commitment problem and the risk of war by adopting some measure of strategic or tactical restraint. Hence, the U.S. should have relaxed the oil embargo with its *modus vivendi* in 1941;

¹⁰⁵ SarDesai, *Vietnam*, 161.

¹⁰⁶ Chen, *China's War with Vietnam*, 87-8.

¹⁰⁷ Deng made the argument to the Chinese Communist Party Central Committee on 16 March 1979. Zhang, "Deng Xiaoping and China's Decision," 19.

India should have suspended the “forward policy” in 1962; and North Vietnam should have exercised restraint in 1965. The rest of this chapter explains why they didn’t do so.

1.5 Costly Signaling (M4)

Chapter 4 tested the signaling mechanisms using randomized experiments. Experimental tests are used because, as discussed earlier, observational evidence based on historical records does not offer a fair test of the signaling mechanisms. Most records do not show decision-makers articulating explicit credibility assessments for the signals they received. Even when credibility assessments are on record, they are always confounded with the effects of other pieces of information, including the signals observed previously or simultaneously. In particular, in international politics, there are few pure cases of a sunk-cost signal and the conditions for type-separation are often opaque. But the historical cases provide a sense of the external validity of the signaling mechanisms. They also point us to contradictions and puzzles that suggest possible refinements to those mechanisms.

Mechanisms M4, M5 and M6 share a truism and an assumption. The truism: War is prevented if one side is successfully deterred by the other. The assumption: Deterrence is more likely to succeed if private information on resolve can be credibly signaled. Under Mechanism M4, a state can credibly signal its resolve by sinking some costs or risks that would discourage unresolved states from sending the costly signal.

The within-case and cross-case evidence for M4 is limited in the case universe. There are a number of cases that contradict the logic in M4. In 1962, China's policy of armed coexistence – which blocked India's "forward policy" and led to several armed clashes – showed China's willingness to bear the costs and risks of defending its territorial claims. However, the costly signals did not convince New Delhi. Instead, the Indian government decided to make its "forward policy" even more aggressive.¹⁰⁸ Indian sources and subsequent defeats showed that India had made no contingency preparations for a large-scale Chinese attack.¹⁰⁹ John Graver concluded that "[t]here was a virtual consensus among Indian leaders that China would not respond with military force to Indian advances, and that if it did, any military response would be extremely limited. A Chinese resort to large-scale military force was deemed impossible."¹¹⁰ Allen Whiting reported that the belief that China would not fight was so strong in New Delhi that it seemed unshakable.¹¹¹ This was clear to Beijing by October 1962. In a meeting just before the war, Mao was told that a key reason for India's increased aggressiveness was its belief that China was "bluffing".¹¹² Premier Zhou Enlai allegedly argued that "to fight a bit ... would cause some people to understand things more clearly."¹¹³

By February 1979, China's deployment of forces had given substance to its repeated verbal warnings to Vietnam. More than 400,000 troops were mobilized along the Sino-

¹⁰⁸ Garver, "China's Decision for War with India," 108-9.

¹⁰⁹ Whiting, *Chinese Calculus of Deterrence*, 93. Steven Hoffmann reported that "surprise was immense" in New Delhi when China attacked: "Neither the decision-makers nor those advising them had been really prepared for war." Hoffmann, *India and the China Crisis*, 163-4.

¹¹⁰ Garver, "China's Decision for War with India," 109.

¹¹¹ Whiting, *Chinese Calculus of Deterrence*, 168.

¹¹² Garver, "China's Decision for War with India," 120.

¹¹³ *Ibid*, 115.

Vietnamese and Sino-Laotian borders, and the Chinese South Sea Fleet with three hundred ships was deployed to the Vietnamese coast.¹¹⁴ But the Vietnamese appeared unmoved. It also seemed that they did not think China would invade at that point in time. Just one day before the Chinese invasion, Vietnamese leaders including Premier Pham Van Dong and Chief of Staff Van Tien Dung left the capital for a four-day visit to Cambodia.¹¹⁵

In some cases, it seems that costly signals promoted rather than deterred escalation by the opponent. This was the case during the July 1937 crisis in the Third Sino-Japanese War, as analyzed in Chapter 3. This was also the case in Vietnam during 1964-65. On both occasions after the U.S. flexed its muscles at Tonkin Gulf and after Operation Rolling Thunder, the North Vietnamese escalated their military operations in South Vietnam.¹¹⁶

The 1929 Sino-Soviet crisis, however, provides some support for M4. In July 1929, the mobilization of Soviet troops at the Manchurian border was a costly signal that convinced Marshal Zhang Xueliang of the Soviet resolve to use force.¹¹⁷ In response, Zhang tried to push for a diplomatic settlement. But Chiang Kaishek remained unconvinced. He judged that the Soviet threat was too costly to implement and hence not credible.¹¹⁸

¹¹⁴ Nguyen-Vo, *Khmer-Viet Relations*, 132; see also Chen, *China's War with Vietnam*, 82.

¹¹⁵ Chanda, *Brother Enemy*, 356. At the same time, the Indian foreign minister was visiting China to mend relations after the 1962 war. Chanda suggested that Vietnam did not think China would launch a war and compromise the diplomatic thaw with India.

¹¹⁶ Logevall, *Origins of the Vietnam War*, 67, 76-80.

¹¹⁷ Xue, *Zhongsu Guanxi Shi*, 235; Zeng, Huang and Jin, *Zhonghua Minguo Shi*, 209.

¹¹⁸ Zeng, Huang and Jin, *Zhonghua Minguo Shi*, 210.

Overall, the case universe provides no clear evidence that sinking costs make signals more credible and deterrence more successful. This is consistent with the experimental results from Chapter 5: high-resolve signalers are more likely to send costly signals, but receivers do not necessarily respond according to the signalers' expectation, despite the sunk costs. Experiments also show that receivers are highly responsive to Mechanism M5: the credibility of a threat is strongly influenced by the perceived costs of implementing the threat. Does the case evidence support the experimental findings?

1.6 Costly Implementation (M5)

The experimental results find strong support in the case universe. Several cases show that the threat to use force becomes less credible if the opponent believes that the threat is costly to implement.

In 1904, Russia thought that Japan would not dare to wage war. Discounting the capabilities of the Japanese military, many Russian leaders did not see the need to take Tokyo seriously.¹¹⁹ During the 1929 Sino-Soviet crisis, Nanjing called the Soviet ultimatum received on 13 July an empty threat.¹²⁰ The Nanjing government believed that the Soviet Union would not resort to war given its international isolation and domestic

¹¹⁹ Nish, *Origins of the Russo-Japanese War*, 241, 256; Nicholas Riasanovsky and Mark Steinberg, *A History of Russia* (New York: Oxford University Press, 2005), 374; David Schimmelpenninck Van der Oye, "The Immediate Origins of the War," in *The Russo-Japanese War in Global Perspective: World War Zero*, ed. John Steinberg et al. (Leiden: Brill Academic Publishers, 2005), 42. It is possible that racist thinking had biased the Russian assessment. See, e.g., Ian Nish, "Stretching out to the Yalu: A Contested Frontier, 1900-1903," in *The Russo-Japanese War in Global Perspective: World War Zero*, ed. John Steinberg et al. (Leiden: Brill Academic Publishers, 2005), 64.

¹²⁰ Zeng, Huang and Jin, *Zhonghua Minguo Shi*, 207.

difficulties.¹²¹ Chiang Kaishek saw that the Soviets had “the intent to maintain [their claims] but no intent to send troops to protect [their claims]”.¹²² He telegraphed Marshal Zhang Xueliang: “In view of [the Soviet] domestic situation and international relations, [the Soviet Union will] not necessarily dare to declare war against us. The Central government has already set its policy on this matter, and [we] must protect our national sovereignty and absolutely not yield to Soviet pressure”¹²³

In July 1938, after Soviet troops moved into Changkufeng Hill, Japanese military leaders who supported the use of force thought that the Soviets were unlikely to retaliate with a strong offensive. Colonel Inada Masazumi, the architect behind the offensive plan, argued that a war with Japan would be undesirable for Stalin, who was focusing his resources on implementing his economic program and dealing with the German threat in the west.¹²⁴ Witnessing Stalin's purges in the Soviet armed forces that peaked in the summer of 1938, the Japanese military had also discounted the Red Army as a capable fighting force.¹²⁵

In October 1950, the U.S. believed that overt Chinese intervention in Korea was unlikely. The belief was despite at least three clear warnings from Beijing and the deployment of

¹²¹ Oliver Edmund Clubb, *China and Russia: The Great Game* (New York: Columbia University Press, 1971), 260.

¹²² Chiang Kaishek's diary entry on 18 July, in Zeng, Huang and Jin, *Zhonghua Minguo Shi*, 208.

¹²³ Zeng, Huang and Jin, *Zhonghua Minguo Shi*, 208.

¹²⁴ However, Inada was wise enough to caveat that nothing was certain due to “the unpredictable attitude of the Soviet Union.” Kikuoka, *Changkufeng Incident*, 82-3.

¹²⁵ Goldman, *Red Army Victory*, 55-6.

hundreds of thousands of Chinese troops in Manchuria.¹²⁶ Washington assumed that war would be too costly to China, given her limited military capabilities, her internal security problems, her economic reconstruction needs after a long civil war, and her reluctance to be further dependent on the Soviets.¹²⁷ The CIA estimated that if China entered the Korean War, “the regime's entire domestic program and economy would be jeopardized,” and “anti-Communist forces would be encouraged and the regime's very existence would be endangered.”¹²⁸ Dean Acheson thought that fighting the U.S. “would be sheer madness” for China given her circumstances.¹²⁹ Indeed, Chinese leaders had serious reservations when they decided to enter the Korean War.¹³⁰ But these reservations were overridden by the arguments discussed in the earlier sections.

In 1962, India believed firmly that China would not wage war. Beijing attributed the belief to the domestic and international difficulties that China faced.¹³¹ Zhou Enlai stated that “[the Indians] reckoned that our famine was very serious, Tibet was empty, the

¹²⁶ Chen, *China's Road to the Korean War*, 163-4, 170; Stueck, *Rethinking the Korean War*, 89; Christensen, *Worse Than a Monolith*, 67-8.

¹²⁷ Stueck, *Rethinking the Korean War*, 100-1; Foot, *The Wrong War*, 80-2, 86.

¹²⁸ Chen, *China's Road to the Korean War*, 169

¹²⁹ *Ibid.*

¹³⁰ Participants at the crucial 4 October Politburo meeting raised several major reservations about entering the Korean War. See Chen, *China's Road to the Korean War*, 182; Shen and Li, *After Leaning to One Side*, 40; Hao and Zhai, “China's Decision,” 105. Shen and Li summarized the reasons for their objection: “China still needed time to heal its wounds from the successive wars in the past decades; there were areas in China yet to be liberated, and those newly liberated areas had not yet carried out land reforms; the weapons and equipments of the PLA lagged far behind those of the American armed forces, and China did not have command of the air and the seas; and having just endured protracted wars in China, some party cadres and PLA soldiers were war-weary and longed for peace.” Shen and Li, *After Leaning to One Side*, 40. Mao's telegram to the Soviet ambassador on 3 October also highlighted serious concerns within the Politburo on the domestic impact of a Sino-U.S. war. Niu Jun, “Birth of the People's Republic,” 239.

¹³¹ Garver, “China's Decision for War with India,” 114.

rebellion unsettled.”¹³² According to General Lei Yingfu, Indian leaders thought China would not fight a war at a time when she was facing domestic economic problems as well as international threats from the Soviets and the Americans.¹³³ Hence, despite repeated warnings and costly signals from China, India escalated her “forward policy” because she thought China’s threat of war would be too costly to implement.

1.7 Contradictory Signaling (M6)

Mechanisms M4 and M5 focus on the individual signal. The level of inquiry is at the unit level. Mechanism M6 takes a step back and considers the series of signals sent. The level of inquiry is at the serial level. The idea is deceptively simple: A series of signals with a salient contradiction is noisier and less credible.¹³⁴ Hence, deterrence is less likely to succeed when signals of strong resolve are mixed with signals of weak resolve.

Outside a controlled experimental setting, a real-world crisis often has sufficient richness and noisiness for one to locate some kind of contradiction. However, M6 requires a *salient* contradiction: the contradictory signal must be registered and considered by decision-makers. Such evidence is not easily found in the historical records. The Korean War, however, offers a suggestive case.

¹³² Fravel, *Strong Borders*, 195.

¹³³ Ibid.

¹³⁴ See the first footnote in the section on contradictory signaling in Chapter 1.

In 1949, the U.S. withdrew its troops from South Korea but later made permanent the provisional U.S. advisory team in South Korea, with increased commitments in American aid.¹³⁵ On various occasions, U.S. leaders had referred to Korea as a testing ground between communism and democracy.¹³⁶ But in January 1950, Secretary of State Dean Acheson's National Press Club speech publicly omitted South Korea from the U.S. defense perimeter. It is known that Stalin and Mao had subsequently discussed Acheson's speech, and diplomatic historians have argued that it affected Stalin and Mao as they assessed the credibility of the U.S. commitment to defend South Korea.¹³⁷ By April 1950, Stalin concluded that "the prevailing mood [in the U.S.] is not to interfere."¹³⁸

The 1950 case also extends M6 by highlighting the importance of a salient *rational contradiction*: If the opponent did not choose war earlier when it was less costly, why would the opponent choose war now that it has become more costly? In October 1950, the rational contradiction supported the belief in Washington that overt Chinese intervention in Korea was unlikely. The CIA concluded on 28 September that since China did not intervene earlier in Korea when it was favorable for them to intervene, "like the

¹³⁵ Doris Condit, *History of the Office of the Secretary of Defense, Volume II: The Test of War, 1950-1953* (Washington, DC: Historical Office, Office of the Secretary of Defense, 1984), 45; Stueck, *Rethinking the Korean War*, 75

¹³⁶ Stueck, *Korean War: An International History*, 35.

¹³⁷ Christensen, *Worse Than a Monolith*, 53. For a contrary view, see James Matray, "Dean Acheson's National Press Club Speech Reexamined," *Journal of Conflict Studies* 22, No. 1 (2002): 28-55. William Stueck argued that the U.S. commitment to South Korea was undercut by a Europe-first strategy and by bureaucratic pressures within Washington. Peter Lowe suggested that Washington did not want to give a firm commitment to defend South Korea for fear that the South Koreans would take advantage of the commitment to attack North Korea. Stueck, *Rethinking the Korean War*, 74-5; Lowe, *The Korean War*, 14.

¹³⁸ Stueck, *Rethinking the Korean War*, 73.

USSR, [China] will not openly intervene in North Korea.”¹³⁹ The National Security Agency retrospectively stated: When China did not intervene after the U.S. landed at Inchon and forced back the North Koreans, a large-scale Chinese intervention appeared to be much less likely, even to intelligence analysts who were concerned earlier about the deployments of Chinese forces in the northeast.¹⁴⁰ The logic is as follows: If China did not wage war earlier when it was easier and less costly, China is unlikely to wage war now that it is harder and more costly. This suggests an extension to M6:

Observation 28: The threat to use force becomes less credible if its implementation involves a *rational contradiction* with the choices made earlier in the crisis (i.e. the signaler chose not to use force when it was less costly to do so).

¹³⁹ Chen, *China's Road to the Korean War*, 169.

¹⁴⁰ Christensen, *Worse Than a Monolith*, 68. The U.S. ambassador to Moscow argued that Chinese intervention had become less likely because the best time for intervention “was logically when UN forces were desperately defending the small area of Taegu-Pusan, when the influx of overwhelming numbers of Chinese ground forces would have proved the decisive factor.” Chen, *China's Road to the Korean War*, 169.

PART 2: THEORY GENERATION

Part 2 of this chapter infers alternative theories of war that fall outside the rationalist mechanisms assessed in Part 1. Four sets of hypotheses are extracted from the East Asian cases. The first two operate at the international level, retaining the unitary-state assumption. The last two sets of hypotheses remove the assumption and focus on domestic causes of war.

2.1 Two-Dimensional Games

In 1941, both Tokyo and Washington recognized the potential costs of war and preferred to reach a negotiated settlement if possible. The final breakdown of the negotiations occurred in late November 1941 after Washington decided to abandon the modus vivendi proposal. By lifting the oil embargo if Japan withdrew from southern Indochina, the modus vivendi would stem the expected relative power decline that drove Tokyo to desperation.¹⁴¹ Why did Washington discard the modus vivendi? The key reason was because it was impossible to isolate a Japanese-American bargain within a purely bilateral setting.¹⁴² A separate peace with Japan based on the modus vivendi would threaten the ABCD (American-British-Chinese-Dutch) entente – the cornerstone of U.S. strategy in World War II. As Secretary of State Cordell Hull later recalled, Washington’s abandonment of the modus vivendi was driven especially by “the serious risk of collapse

¹⁴¹ A detailed narrative on the modus vivendi is found in David Lu, *From the Marco Polo Bridge to Pearl Harbor: Japan's Entry into World War II* (Washington, DC: Public Affairs Press, 1961), 224-9.

¹⁴² Iriye, *Origins of the Second World War*, 178-81.

of Chinese morale and resistance and even of disintegration of China.”¹⁴³ Washington simply could not concede if Japan did not abandon its China policy. Chiang Kaishek, according to Hull, “sent numerous hysterical cable messages to different Cabinet officers” to kill the proposal.¹⁴⁴ Lord Halifax, the British ambassador, also sent severe criticisms of the proposal.¹⁴⁵ Winston Churchill reminded the White House: “What about Chiang Kai-shek? Is he not having a very thin diet? [Should China collapse] our joint dangers would enormously increase.”¹⁴⁶

A peaceful settlement based on the *modus vivendi* might have reduced the risk of war on the Pacific front, but it might also unravel the ABCD entente and worsen the U.S. position on the European front. Washington might have made enough concessions to avoid a war if the bargain with Japan could be isolated in a purely bilateral setting. But it couldn't. The 1941 case shows that bargaining failure may arise because concessions made in one bargaining dyad can shift relative power in a second bargaining dyad. Although a bargainer may be willing to make sufficient concessions in a solo bargaining game, those concessions may be infeasible because of their negative impact on the bargainer's position in another bargaining game.¹⁴⁷

¹⁴³ Lu, *From the Marco Polo Bridge*, 229.

¹⁴⁴ LaFeber, *The Clash*, 207-8.

¹⁴⁵ Heinrichs, *Threshold of War*, 211.

¹⁴⁶ *Ibid.*

¹⁴⁷ This may be modelled as one *n*-player bargaining game with two bilateral subgames.

Hypothesis 1: Bargaining failure leading to war is more likely when concessions made in one bargaining dyad negatively affects bargaining power in another bargaining dyad.

Hypothesis 1 suggests a rationalist explanation for war that is separate and distinct from the three rationalist explanations proposed by James Fearon. Fearon showed that with perfect information, full divisibility of the prize, and no commitment problems, two rational actors should always reach a peaceful division of a prize that is preferable to war. Fearon's model is based on the prize being valued on a single common dimension. It is straightforward to show that a peaceful division cannot be guaranteed in the model if the valuation of each increment of concession is not fully reflected on the same common scale. That is, if a bargainer values one point of concession in the bilateral game as more than just one point, a mutually preferable division may not exist. This occurs when the prize is valued on two dimensions with concessions made in one dimension directly affecting the relative bargaining power in the other dimension. A model of two-dimensional bargaining¹⁴⁸ cannot guarantee a peaceful division of the prize even if we assume perfect information, full divisibility of the prize, and no commitment problems. Future work may design a game-theoretic experiment to test the relationship between two-dimensional bargaining and the incidence of conflict.

Two-dimensional bargaining can lead to war even if we restrict it based on similar rationalist assumptions proposed by Fearon. It points us to a distinct rationalist

¹⁴⁸ In "two-dimensional bargaining", at least one player plays two different but related bargaining games. For example: a bilateral game between A and B in the first dimension, and another bilateral game between A and C in the second dimension.

explanation that is historically significant. Hence, it challenges Fearon's famous assertion that the twin problems of information and commitment comprise "the full set of rationalist explanations that are both theoretically coherent and empirically plausible."¹⁴⁹

2.2 Opportunistic Wars

Hypothesis 2: War is more likely when conquest is *suddenly* easy and cheap.

Stephen Van Evera proposed that war is more likely when conquest is easy, and conquest is easy in an offense-dominant situation.¹⁵⁰ Offense dominance is a consequence of military doctrine and technology, geography, as well as diplomatic and domestic social factors.¹⁵¹ Except for geography, these causes can be endogenously shaped through state choice.¹⁵² Extending Van Evera's theory, Hypothesis 2 focuses on a period in which the cost of conquest suddenly falls due to an exogenous shock. The exogenous shock incentivizes opportunistic aggression. There is a sharp danger of war when there is a sudden collapse in the cost of conquest.

There are at least four opportunistic wars in the East Asian case universe: the Sino-Russian War in 1900, the German-Japanese War in 1914, the Franco-Thai War in 1940,

¹⁴⁹ James Fearon, "Rationalist Explanations for War," *International Organization* 49, No. 3 (1995): 380.

¹⁵⁰ Stephen Van Evera, *Causes of War: Power and the Roots of International Conflict* (Ithaca, NY: Cornell University Press, 1999), chapter 6.

¹⁵¹ Van Evera, *Causes of War*, 122. See also Jervis, "Security Dilemma."

¹⁵² Geography almost never changes. Military technology may be partially exogenous, but a sudden revolution in military technology is rare.

and the Vietnamese-Cambodian War in 1975. While Saigon was collapsing in 1975 at the final stage of the Vietnam War, Cambodian forces exploited the situation and attacked several Vietnamese-held islands in the Gulf of Thailand.¹⁵³ Likewise in 1940, France's surrender and the war in Europe exposed French Indonesia to weakness, making a Thai invasion much more feasible than before.

In 1914, Japan exploited the outbreak of war in Europe to conquer German territories in Shandong and the Pacific Micronesia.¹⁵⁴ Foreign Minister Kato Komei stated that joining the war would offer "the opportunity to sweep up bases in Eastern waters and to advance the Empire's position in the world."¹⁵⁵ Sir Edward Grey, the British Prime Minister, pointed out: "In the Great War [the Japanese] took some advantage of the opportunity to strengthen their position with China in East Asia. Europe was prostrated in war, the attention and at least the energy of the United States was absorbed in it. The opportunity for Japan was immense and unique. What Western nation with a population feeling the need for territorial outlets would have used such an opportunity with more or even as much restraint?"¹⁵⁶

¹⁵³ David Chandler, *History of Cambodia* (Boulder, CO: Westview Press, 2008), 269; Kiernan, *Pol Pot Regime*, 104-5.

¹⁵⁴ Roger Hackett, *Yamagata Aritomo in the Rise of Modern Japan, 1838-1922* (Cambridge, MA: Harvard University Press, 1971), 283; Frederick Dickinson, *War and National Reinvention: Japan in the Great War, 1914-1919* (Cambridge, MA: Harvard University Asia Center, 1999), 36-9; Takeuchi Tatsuji, *War and Diplomacy in the Japanese Empire* (New York: Doubleday, Doran & Co, 1935), 169-70.

¹⁵⁵ Conrad Totman, *A History of Japan* (Oxford: Blackwell, 2005), 428.

¹⁵⁶ Peter Lowe, "Great Britain and Japan's Entrance into the Great War, 1914-1915," in *The Anglo-Japanese Alliance, 1902-1922*, ed. Phillips Payson O'Brien (London: RoutledgeCurzon, 2004), 172.

In 1900, China was under attack by the Great Powers in the Boxer War, and the riots in Manchuria provided a context for Russian action. General Kuropatkin, the Russian War Minister, told Count Witte, the Finance Minister, that he was “very pleased with [the Boxer rebellion], for it gave us ‘grounds for taking Manchuria’” and turning it “into something like Bukhara”.¹⁵⁷ Russia dispatched a large troop and occupied Manchuria.¹⁵⁸

Opportunistic wars can be viewed as an extension from the commitment theory of war, if the conqueror believes that the relative power balance will shift in its disfavor in the future. Hence, a peaceful bargain is infeasible because the weakened party has the incentive to renege once it becomes strong. But it is important to emphasize that opportunistic wars can occur even if leaders are uncertain or neutral on the future power trend. They may choose war simply because conquest was previously difficult and costly, but suddenly becomes easy and cheap.

2.3 Principal-Agent Problems

The unitary-actor assumption masks the domestic factors leading to war. I relaxed this assumption earlier when I described the domestic manifestation of inadvertent enforcement. This is one variant of the principal-agent problem: The interests of leaders making the decision for war may not be fully coincident with the national interest.

¹⁵⁷ Sidney Harcave, ed., *The Memoirs of Count Witte*, Vol. 1 (Armonk, New York: M. E. Sharpe, 1990), 279. See also Tan, *Boxer Catastrophe*, 159; Cheng, *Sino-Russian Relations*, 65.

¹⁵⁸ Nish, *Origins of the Russo-Japanese War*, 85; William Duiker, *Cultures in Collision: The Boxer Rebellion* (San Rafael, CA: Presidio Press, 1978), 187.

Another variant of the principal-agent problem occurs when the preference of the central government diverges from the preference of its agent on the ground. The East Asian case universe suggests three pathways by which this principal-agent problem causes war.

The first pathway is *direct escalation* by the agent:

Hypothesis 3: War is more likely if the agent has a preference for escalation and a significant latitude of action.

Due to information and resource constraints, the central authority must necessarily allow for some degree of agent latitude on the ground. The agent may then exploit this latitude to escalate conflict on the ground. Of course, the center may limit this freedom of action by crafting explicit policies and guidelines. But no policy can cover all possible contingencies and every policy can allow for at least some interpretative flexibility.

The 1900 Boxer War suggests a case of agent-driven escalation. The Great Powers were not disposed toward fighting a war in China, but their colonial representatives on the ground took preemptive decisions that escalated hostilities. Xiang Lanxin chained together a series of fatal decisions taken by the allied authorities in China: “the calling up of the Legation guards to defend [foreign nationals] who were not at all threatened with physical danger; the Seymour Expedition to relieve people who were not yet besieged; the reckless [hunt for the Boxers] in Beijing to take Chinese jurisdiction into their own hands and, finally, the seizure of the Dagu Forts to relieve the relievers – the Seymour

Column that never made it to Beijing.”¹⁵⁹ The allied attack on the Dagou Forts triggered Empress Dowager Cixi’s decision for war in June 1900.

The 1938 Changkufeng War was an agent-driven war. The Emperor had already rejected the proposal to launch an offensive at Changkufeng.¹⁶⁰ Most Japanese forces had already pulled back from the Tumen River by 28 July. But Division Commander Suetaka Kamezo longed to “teach the Russians a lesson”.¹⁶¹ He remained at the frontline and spotted a Soviet patrol on a hill near Changkufeng on 29 July. Suetaka interpreted this as a Soviet intrusion into Japanese territory based on Japanese maps. He crossed the Tumen River, eliminated the intruders, and attacked Changkufeng.¹⁶²

That war broke out despite the imperial prohibition is remarkable. Yet the prohibition was not transmitted to the ground as an explicit blanket prohibition on the use of force. It tied Suetaka’s hands – but not totally. It allowed the hawks on the ground to interpret the limits of the prohibition liberally. To Suetaka, the imperial prohibition did not apply in a “separate” border incident away from Changkufeng where the Japanese Army was duty-bound to defend the frontier.¹⁶³ Indeed, while Tokyo’s cable on 23 July stressed that the use of force would not be authorized, it also added the caveat: “unless some new

¹⁵⁹ Xiang, *Boxer War*, 355. See also Paul Cohen, *History in Three Keys: The Boxers as Event, Experience, and Myth* (New York: Columbia University Press, 1997), 48-51. Xiang argued that during the crisis “Lord Salisbury had no choice but to delegate full authority to his representatives on the spot, for the fluid situation in China had become too difficult for London to control on daily basis.” Xiang, *Boxer War*, 241.

¹⁶⁰ Coox, *Anatomy of a Small War*, 57-70.

¹⁶¹ Goldman, *Red Army Victory*, 65.

¹⁶² Coox, *Nomonhan*, 133; Hata, “The Japanese-Soviet Confrontation,” 146-8; Coox, *Anatomy of a Small War*, 95-111.

¹⁶³ Coox, *Nomonhan*, 132.

developments occur, such as a great deterioration in the situation entailing sheer necessity.”¹⁶⁴ What Suetaka saw was a prohibiting order from the top that left room for his hawkish interpretations on the ground.

On the eve of the Nomonhan War in 1939, Tokyo was shocked by the Kwantung Army’s hidden plan to bomb the Soviet air base at Tamsag Bulak in Outer Mongolia. Tokyo immediately sent a prohibiting order, highlighting that the air offensive “will lead to the gradual extension of bombing assaults by both sides behind the opposing frontiers ... For operational liaison purposes, Lt. Colonel Arisue is being flown to Kwantung Army Headquarters on 25 June.”¹⁶⁵ The Kwantung Army chose to interpret Tokyo’s order, which was courteously crafted in the formal Japanese language, as a “suggestion”. Knowing that such an interpretation would not stand once Lieutenant-Colonel Arisue arrived, the Kwantung Army expedited its plans and quickly launched its air offensive against the Soviets.¹⁶⁶

Franco-Thai negotiations collapsed in 1940 largely due to the uncompromising attitude of Admiral Decoux and Paul Baudouin, who were appointed governor general of Indochina and Foreign Secretary by the new Vichy regime.¹⁶⁷ Franco-Thai negotiations stalled with Baudouin’s inflexibility and Decoux’s strong aversion to any alienation of colonial

¹⁶⁴ Ibid, 131. According to Goldman, there were officers in the AGS hoping that the local commander would take the initiative and attack, without the AGS having to take responsibility. Goldman, *Red Army Victory*, 63.

¹⁶⁵ Goldman, *Red Army Victory*, 107.

¹⁶⁶ Ibid, 108; Hata, “The Japanese-Soviet Confrontation,” 164-5. Goldman pointed out that the Kwantung Army was given a broad mandate under the Imperial Order to “defend Manchukuo”. Goldman, *Red Army Victory*, 110-1. See also Coox, *Nomonhan*, 266-83.

¹⁶⁷ Flood, “Franco-Thai Border Dispute,” 321-5.

territories. Richard Aldrich suggested that Baudouin was “heavily influenced by his personal financial interests in Indo-China and his recent governorship of the powerful Banque d'Indochine,” and he “therefore pursued a 'policy de banque' that was if anything less flexible than previous French policy.”¹⁶⁸

The second pathway is *information misrepresentation* by the agent to the center:

Hypothesis 4: War is more likely if the agent has a preference for escalation and misrepresents information to the center.

The agent knows more about the ground situation than the center, which depends on the agent for accurate information to guide its policies. If the preferences of the agent and the center are not aligned, the agent has an incentive to select the information transmitted to the center to bias decision-making at the center. Hence, the principal-agent problem may worsen interstate information asymmetry and promote war-causing miscalculations.

Division Commander Suetaka triggered a war at Changkufeng while manipulating the information he sent to Tokyo.¹⁶⁹ He painted a picture of aggressive intrusions by the Soviets into Japanese territory. The Emperor later gave his ex-post sanction for the use of force after being assured that the army was acting to defend the border within Japanese

¹⁶⁸ Aldrich, *Key to the South*, 263.

¹⁶⁹ Coox, *Nomonhan*, 133. To prevent the Army HQ from stopping the attack, reporting was prohibited until after the fact. Hata, “The Japanese-Soviet Confrontation,” 148.

boundaries.¹⁷⁰ Moscow interpreted the situation accurately on 2 August: “Tokyo is trying to avoid complications, but the local military is presenting its government with faits accomplis; in order to obtain approval it sends false information about our imaginary moves across the frontier as attacks on Japanese forces, etc.”¹⁷¹

In June 1939, the Kwantung Army submitted its proposal to the Army HQ at Tokyo prior to the Nomonhan operations. The proposal was a vague outline that withheld crucial operational details. Even so, officials from the Army Ministry had strong reservations.¹⁷² But Army Minister Itagaki Seishiro¹⁷³ endorsed the plan: “There is no need for us to be nervous about the operations of a force of only one division’s strength.”¹⁷⁴ The Army HQ approved the counterattack with only partial information about its scope. Actual operations were much more aggressive, including an air offensive against Soviet forces in Mongolia – a major detail deliberately omitted from the plan presented to Tokyo.¹⁷⁵ Tokyo, as described earlier, was horrified when it later found out about the air offensive.

The allied attack on the Dagu Forts was the trigger behind Empress Dowager Cixi’s decision for war in 1900. It turned out, however, that China had declared war without knowing that the forts – which were important to Beijing’s defense – were already lost. Viceroy Yu Lu, who was responsible for the Dagu Forts, did not report the truth. His

¹⁷⁰ Coox, *Nomonhan*, 134; Hata, “The Japanese-Soviet Confrontation,” 149.

¹⁷¹ Jonathan Haslam, *The Soviet Union and the Threat from the East, 1933-1941: Moscow, Tokyo and the Prelude to the Pacific War* (University of Pittsburgh Press, 1992), 118.

¹⁷² Hata, “The Japanese-Soviet Confrontation,” 164.

¹⁷³ This was the same Itagaki who plotted the Mukden conspiracy in 1931 when he was a senior staff officer at the Kwantung Army HQ.

¹⁷⁴ Goldman, *Red Army Victory*, 107.

¹⁷⁵ Hata, “The Japanese-Soviet Confrontation,” 164; Goldman, *Red Army Victory*, 107.

memorial on 21 June was a positive report of Chinese successes and the cooperation between the Boxers and Imperial troops in resisting the allied forces.¹⁷⁶ The Imperial declaration of war was issued on the same day after the arrival of Yu Lu's "news".¹⁷⁷ According to Ji Chang, a secretary at the Grand Council, "Viceroy Yu reported ... that [General] Luo Rongguang was engaging the foreign troops. In fact the Forts had already been lost and the Viceroy dared not tell the truth. He only reported the truth two days later."¹⁷⁸ By then, Beijing had already declared war on the basis of Yu Lu's misleading memorials.¹⁷⁹

The third pathway is *signaling distortion* by the agent to the opponent:

Hypothesis 5: War is more likely if the agent has a preference for escalation and transmits biased signals to the opponent.

If the agent has revisionist preferences, it may send signals that reflect revisionist intent contrary to the center's preference for the status quo. This creates noise in the signaling environment and may lead the opponent state to misinterpret the center's true intent.

¹⁷⁶ Esherick, *Origins of the Boxer Uprising*, 303; Xiang, *Boxer War*, 328; Tan, *Boxer Catastrophe*, 75; Immanuel Hsu, *The Rise of Modern China* (New York: Oxford University Press, 2000), 395.

¹⁷⁷ Tan, *Boxer Catastrophe*, 75. Empress Dowager sent an urgent message to Yu Lu on 20 June: "Your report of June 17 has not been followed by further report. Did the war break out at the Dagu forts? What have the foreign soldiers done in the past few days? How many Boxers have you recruited? Are they helpful at the front? What is the general situation concerning the Boxers?" Xiang, *Boxer War*, 327.

¹⁷⁸ Xiang, *Boxer War*, 308.

¹⁷⁹ Note that in this case, the agent misrepresented information not because of a preference for escalation, but because of fear.

Here, the principal-agent problem may worsen the information asymmetry and promote war-causing miscalculations by the opponent.

In 1900, Empress Dowager Cixi misperceived the attacks at the Dagu Forts as part of a foreign attempt to depose her and restore Emperor Guangxu to power.¹⁸⁰ How did the misperception arise? The foreign diplomatic corps in Beijing fed that misperception by adopting an attitude of non-recognition towards Cixi's regime after her coup d'état in 1898. The collective stance of the diplomatic corps was not authorized by their home governments, who did not establish any policy to depose the empress dowager.¹⁸¹ But the biased signals sent by the diplomatic corps reinforced Cixi's belief that the allied attacks aimed to topple her regime and restore Emperor Guangxu. Cixi's false belief may explain why she took the leap and went to war with eight Great Powers simultaneously in 1900.

Hypotheses 3, 4 and 5 all assume that the agent prefers escalation but the center doesn't. If this preference assignment is reversed, Hypotheses 3, 4 and 5 become hypotheses for peace. A principal-agent problem does not necessarily promote war: the peace-making moves in the 1929 case and the 1937 case were also consequences of a principal-agent problem. In both cases, the agent on the ground (Marshal Zhang Xueliang in 1929 and General Song Zheyuan in 1937) had exploited their latitude of action to push for a peaceful settlement, contrary to the hard-line taken by the central government.¹⁸²

¹⁸⁰ Xiang, *Boxer War*, 256, 294-6, 355; Tan, *Boxer Catastrophe*, 73; Esherick, *Origins of the Boxer Uprising*, 302.

¹⁸¹ Xiang, *Boxer War*, 353-4.

¹⁸² Zeng, Huang and Jin, *Zhonghua Minguo Shi*, 210-5. See Chapter 3 for the 1937 case.

Sometimes, it is the agent that is the dove. In this case, the principal-agent problem is not so much a problem as a blessing in disguise.

2.4 Three Types of Audience Costs

Audience costs are the political losses from backing down after escalating in an international crisis. The theory of audience costs suggests that the higher the audience costs, the more credible the signal of resolve. Thus, a peaceful outcome is more likely because deterrence is more likely to succeed. But audience costs can become so high that they exceed the costs of war. In such situations, war becomes a rational choice.¹⁸³

The literature suggests two types of audience costs. *International audience costs* are the international political losses from backing down after escalating. The losses are conferred by an international audience. But it is not clear what exactly these losses are. Suggested candidates include the losses in a state's international credibility, reputation, or national honor. *Domestic audience costs* are the domestic political losses from backing down after escalating. It is assumed that citizens or domestic elites will punish the leader for backing down and making the nation suffer international audience costs.¹⁸⁴ I propose a different form of audience costs: *personal* audience costs. These are the political costs to a career politician or bureaucrat for opposing the position taken by powerful colleagues or

¹⁸³ James Fearon, "Domestic Political Audiences and the Escalation of International Disputes," *American Political Science Review* 88, No. 3 (1994): 577-92.

¹⁸⁴ There is a major empirical debate on the existence and effect of domestic audience costs in international politics. Theoretically, it is not clear why leaders should be punished if backing down is a rational decision – if it would prevent the nation from suffering the costs and risks of war. See the relevant footnotes under the section for M5 in Chapter 1.

superiors. Personal audience costs involve a different definition of audience costs and a different causal mechanism, which will be explained later in this section.

There is much debate over whether domestic audience costs really exist.¹⁸⁵ I ask a related but different question: Do perceptions of high audience costs make leaders more likely to choose war? I begin first with international audience costs:

Hypothesis 6: War is more likely if leaders expect high *international audience costs* and the opponent does not back down.

When do leaders believe international audience costs to be so high as to exceed the cost of war? The relevant cases concentrate on states with magnified concerns over their international credibility: they fear that backing down in the present crisis will make the international audience discount the credibility of their resolve in future crises. Credibility of resolve appears especially important to states with multiple commitments around the world. They believe that backing down in one crisis in one part of the world will increase their risks of being challenged in other crises in other parts of the world. The more commitments they have, the higher their valuation of their international credibility. The higher the valuation, the larger the expected audience costs, and the greater the incentive for war.

¹⁸⁵ See the footnotes in the section for M5 in Chapter 1. Chapter 4 had used a survey experiment to examine if audience costs make the threat to use force more credible.

The U.S. entered the Korean War in June 1950 with a keen eye to international audience costs. Washington saw the Korean War as “a struggle for credibility, to prove that the liberal democracy of people unused to sustained effort abroad could rise to the challenge of international communism.”¹⁸⁶ Dean Acheson pointed out on 28 June that while Korea might not have major strategic significance in itself, it was “vital ... as a symbol [of the] strength and determination of [the] west”, and shying away from the Korean War would encourage “new aggressive actions elsewhere” and demoralize “countries adjacent to [the] Soviet orbit.”¹⁸⁷

Washington viewed the importance of the Vietnam War not in terms of Vietnam itself, but in terms of U.S. credibility and global commitments.¹⁸⁸ U.S. leaders thought that backing down in Vietnam would demoralize their allies and embolden their adversaries to challenge U.S. commitments in other parts of the world.¹⁸⁹ Assistant Secretary of Defense John McNaughton weighed American aims in Vietnam in early 1965 as “70% – To avoid a humiliating US defeat (to our reputation as a guarantor).”¹⁹⁰ In the final meetings before the war decision was finalized, Secretary of the Navy Paul Nitze warned that “the shape of the world will change” were the U.S. to acknowledge that “we couldn't beat the VC.”¹⁹¹ Secretary of the Army Stanley Resor warned that “we can't go back on our commitment. Our allies are watching carefully.”¹⁹² Lyndon Johnson's announcement

¹⁸⁶ Stueck, *Korean War: An International History*, 43.

¹⁸⁷ Ibid; Stueck, *Rethinking the Korean War*, 81.

¹⁸⁸ Berman and Routh, “Why the United States Fought in Vietnam,” 187.

¹⁸⁹ Logevall, *Choosing War*, xiv.

¹⁹⁰ DeGroot, *Noble Cause*, 125.

¹⁹¹ Schulzinger, *Time for War*, 175.

¹⁹² Ibid.

of war on 28 July stated: “There are great stakes in the balance [If the United States was] driven from the field in Vietnam, then no nation can ever again have the same confidence in America's promise, or in American protection”.¹⁹³

China's war in 1979 can also be interpreted in the light of international audience costs. Vietnam had aligned with the Soviet Union and invaded Cambodia, despite China's coercive diplomacy and repeated warnings. Deng later admitted that he did not want other countries to think China was soft in countering the Soviet-Vietnamese alliance.¹⁹⁴ Chinese leaders had openly expressed their desire to “teach Vietnam a lesson” for its repeated affronts.¹⁹⁵ Vice Premier Li Xiannian said that the Chinese attack was “a slap in the face of [Vietnam] to warn and punish them.”¹⁹⁶

Hypothesis 7: War is more likely if leaders expect high *domestic audience costs* and the opponent does not back down.

The evidence for Hypothesis 7 was discussed in Section 1.2. The section also highlighted Observation 24: Nationalistic fervor increases the domestic audience costs for leaders and reduces their incentive to back down.

Hypothesis 8: War is more likely when a pro-peace position involves high *personal audience costs*.

¹⁹³ Hess, *Presidential Decisions*, 105.

¹⁹⁴ Zhang, “Deng Xiaoping and China's Decision,” 13.

¹⁹⁵ Ross, *Indochina Tangle*, 224; Zhang, “China's 1979 War,” 861.

¹⁹⁶ Zhang, “Deng Xiaoping and China's Decision,” 12.

Fearon's theory defines audience costs as the political losses for backing down after escalating in an international crisis. Audience costs are generated by a reversal in position. The unit of analysis is the state or the political leadership. The audience comprises other states or domestic constituents.

Personal audience costs are a different form of audience costs. These are the political losses to a career politician or bureaucrat for opposing the position taken by powerful colleagues or superiors. Unlike Fearon's theory, personal audience costs need not be generated by a *reversal* in position; they are generated by taking an *opposing* position. The unit of analysis is the politician or official within the political leadership. The audience comprises other politicians and officials within the leadership, including (and especially) the national leader. The concept of personal audience costs connects to both James Fearon's theory of audience costs, as well as the organizational theories of non-self-evaluation by Aaron Wildavsky and Stephen Van Evera.¹⁹⁷

Personal audience costs suggest a different mechanism for war based on a deficiency in the decision-making environment. A pro-war bias emerges when a pro-peace position

¹⁹⁷ Aaron Wildavsky, "The Self-Evaluating Organization," *Public Administration Review* 32, No. 5 (1972): 509-20; Stephen Van Evera, "Why States Believe Foolish Ideas: Non-Self-Evaluation By States and Societies," in *Perspectives on Structural Realism*, ed. Andrew Hanami (New York: Palgrave Macmillan, 2003), 163-98. The articles argued that organizations (Wildavsky) and states (Van Evera) are poor in self-evaluation. Van Evera highlighted government bureaucracies as one of the sources for non-self-evaluation within states: "At a minimum, agencies with evaluative responsibilities are not invited to evaluate – they are kept out of the loop, their opinions unsought. At a maximum, government agencies actively suppress their own internal evaluative units and are discouraged from evaluating the beliefs and policies of other agencies." Van Evera, "Why States Believe Foolish Ideas," 165.

involves high personal audience costs. With personal audience costs, officials do not take positions based solely on their understanding of national interest and the available facts. They may avoid anti-war positions even though they believe in those positions. They have an incentive to withhold opinions or information that promote an anti-war position. Thus, personal audience costs create a deficient decision environment. They promote incomplete evaluation, incomplete information, and the incomplete evaluation of incomplete information. A pro-war consensus may emerge from the deficient decision setting. But that consensus may be an artifact of personal audience costs. The prospect of peace is slim in a decision setting where no one dares to stake an anti-war position.

In June 1900, Empress Dowager Cixi decided to go to war with the eight most powerful nations of the world all at once. Yet there was a consensus at the Imperial Court that supported Cixi's decision. Did all the officials believe that China should go to war? According to the diary of Rong Qing, an official at the 19 June imperial audience, he had complained to Grand Councilor Rong Lu that it did not make sense to fight the eight Great Powers. The Grand Councilor replied: "Your words reflect my heart. But if you try to argue with them again, you would be considered a traitor."¹⁹⁸

In June 1939, the Kwantung Army submitted to the Army HQ its attack plan that would send Japanese forces into Outer Mongolian territories under Soviet protection. It was common knowledge that a cross-border offensive was illegal without the Emperor's approval – but no one at the Army HQ meeting pointed this out. It is possible that no one

¹⁹⁸ Xiang, *Boxer War*, 309. See also Xiang, *Boxer War*, 291.

wanted to be the one to sabotage the plan. It was also difficult for any officer to overtly overturn a proposal already approved by General Ueda Kenkichi, the commander of the Kwantung Army and one of the most senior figures in the military hierarchy.¹⁹⁹ Even Army Minister Itagaki Seishiro said at the meeting that the plan should be approved out of respect for General Ueda.²⁰⁰ But the chief of the AGS Soviet Intelligence Subsection remained disturbed after the meeting. He went to Colonel Inada Masazumi, the chief of the AGS Operations Section, arguing that “there is a big difference between people on the spot and those who see the big picture. You must guide the Kwantung Army.” But his superior was unwilling to oppose the Kwantung Army, insisting that the Kwantung Army officers were all “great men,” so “let’s not say unnecessary things to them.”²⁰¹

Personal audience costs are generated by taking an *opposing* position. The costs are further increased if taking an opposing position involves *reversing* a position taken earlier. Here the intuition is closer to Fearon’s audience costs, which are incurred when a leader backs down from a hawkish position.

Japanese-American negotiations collapsed in November 1941 on the issue of Japan’s withdrawal from China. In fact, by October 1941, both Prime Minister Konoe and the navy believed that Japan should withdraw from China rather than go to war with the U.S.²⁰² The navy leadership agreed internally that “it would be the height of folly to fight

¹⁹⁹ Coox, *Nomonhan*, 269.

²⁰⁰ Goldman, *Red Army Victory*, 107.

²⁰¹ Coox, *Nomonhan*, 262. This evidence comes from a postwar interview with the chief of the AGS Soviet Intelligence Subsection.

²⁰² Iriye, *Origins of the Second World War*, 140-1; Sagan, *Pacific War*, 340.

with the United States on the issue of withdrawing troops from China”.²⁰³ But the navy refused to directly oppose the army in the liaison conferences. Navy Minister Oikawa Koshiro said after the war that he could not reverse his position because “if we were to say that we were not able to carry out operations against the United States, it would have meant we had been lying to the Emperor when presenting operational plans for war.”²⁰⁴ Furthermore, the army was vehemently opposed to a withdrawal from China: reversing its course in China after years of national sacrifices would damage the army’s domestic position and prestige.²⁰⁵ Army leaders would rather go to war than to quit China. Hence, the army’s pro-war position was almost unshakable. As Sato Kenryo, the military affairs section chief, observed: “It would have required enormous courage to voice an anti-war view when pro-war fever was boiling.”²⁰⁶

Three points should be emphasized. (1) Personal audience costs are political costs and do not involve any psychological assumptions. Nonetheless, their effects can be reinforced if there are psychological costs involved in the opposition of the prevailing consensus. The psychological processes suggested in Irving Janis’s groupthink theory can make things worse.²⁰⁷ (2) Historical evidence for personal audience costs requires records that show a

²⁰³ Iriye, *Origins of the Second World War*, 164.

²⁰⁴ Sagan, *Pacific War*, 340.

²⁰⁵ Iriye, *Origins of the Second World War*, 141; see also Sagan, *Pacific War*, 341.

²⁰⁶ Alvin Coox, “The Pacific War,” in *The Cambridge History of Japan*, Vol. 6, ed. Peter Duus (Cambridge, UK: Cambridge University Press, 1988), 334.

²⁰⁷ In Irving Janis’ definition, “The main principle of groupthink, which I offer in the spirit of Parkinson’s Law, is this: The more amiability and esprit de corps there is among the members of a policy-making ingroup, the greater the danger that independent critical thinking will be replaced by groupthink, which is likely to result in irrational and dehumanizing actions directed against outgroups.” Irving Janis, “Groupthink,” *Psychology Today* 5, No. 6 (1971): 44. Janis argued that groupthink influenced the U.S. decision to escalate and enter the Vietnam War. Irving Janis, *Victims of Groupthink* (Boston, MA: Houghton, Mifflin, 1972).

contradiction between one's private belief and public position. But if the private belief is politically costly, it is unlikely to appear on the public record. Hence, contemporary historical records are likely to underestimate the existence and effect of personal audience costs. (3) Broadening and extending M4 to the individual level, opposition to the prevailing opinion – if it involves high political costs – should be more compelling to its audience. Just like a costly signal, a costly opinion should be taken more seriously than a cheap opinion. Whether or not the theoretical analogy holds remains an open question.²⁰⁸

OVERVIEW

Five general observations can be made about the East Asian case universe:

- (1) **(Observation 29)** In at least eight cases, there is evidence that leaders had perceived a significant increase in their potential vulnerability when they took the road to war. Leaders in these cases had pushed for immediate action, escalation, or war based on this perception. The cases include the wars started in 1904, 1931, 1937, 1940, 1941, 1950, 1965 and 1979. This implies that in at least half of the case universe, the commitment theory of war applies.
- (2) **(Observation 30)** In at least eight cases, there is evidence that one side had misperceived its opponent's resolve on the eve of war. Russia underestimated Japanese resolve in 1904; China underestimated Soviet resolve in 1929; and Japan

²⁰⁸ It should be noted that: (1) Neither the experiments in Chapter 4 nor the historical cases in Chapters 3 and 5 provide clear support for M4. (2) A credible signal of one's *intention* through a costly threat may differ qualitatively from a credible signal of one's *belief* through a costly opinion.

underestimated Chinese resolve in 1937. Similarly, Japan underestimated the Soviets in 1939; the U.S. underestimated China in 1950; India underestimated China in 1962; Cambodia underestimated Vietnam in 1978; and Vietnam underestimated China in 1979. In at least five of these cases, there is evidence that signaling failed because the signaled threats to use force were believed to be too costly to implement.

(3) **(Observation 31)** In eight cases, Imperial Japan was one of the parties at war. The Japanese fought against the Chinese in 1900, against the Russians in 1904, against the Germans in 1914, against the Chinese again in 1931 and 1937, against the Soviets in 1938 and 1939, and against the Americans in 1941. As the sole rising imperial power in the region, Imperial Japan fought eight out of the eleven pre-1945 wars in East Asia.

(4) **(Observation 32)** In four cases, leaders took the road to war after an exogenous shock that suddenly made conquest easy. The four opportunistic wars were launched by the Russians in 1900, the Japanese in 1914, the Thais in 1940, and the Cambodians in 1975. These wars were consequences of a sudden exogenous change in the international environment.

(5) **(Observation 33)** Most of the wars appear to involve both rationalist and non-rationalist causes. The two exceptions are the Boxer War in 1900 and the Cambodian-Vietnamese War in 1977-8, in which Imperial China and the Pol Pot

regime had severely strained the assumption of a rational unitary-state actor. The problem lies not so much in the rationality assumption as in the unitary-state assumption. Neither case can be understood without relaxing the latter assumption.

Chapter 6

Conclusion

It is time to stocktake and synthesize. The stocktaking will describe the ideas and results produced in this thesis. The synthesis will generalize and tie together the findings.

The stocktaking is already done. Throughout the thesis, I have crystallized the theoretical ideas and empirical findings at their place of occurrence. I itemized them under separate headings highlighted in bold. I summarized the mechanisms and hypotheses under the headings “Mechanism” and “Hypothesis”; the model predictions under the heading “Prediction”; the case-based observations under the heading “Observation”; and the experimental results under the heading “Result”. There are six mechanisms, eight hypotheses, eleven model predictions, thirty-three case-based observations, and twenty experimental results. The reader can swiftly construct a bullet-point summary of the thesis by flipping the pages and noting the “Mechanisms”, “Hypotheses”, “Predictions”, “Observations” and “Results” where they occur.

Let us turn instead to the synthesis: Which theoretical ideas are supported by the evidence? Which should be revised or rejected? What are the main lessons learnt?

Table 6.1 provides a quick refresher:

Table 6.1: Review of Key Empirical Results

	Experimental support for the mechanism	Historical evidence for the mechanism
<i>2 Rationalist Explanations</i>		
Commitment problem	Yes	Yes
Private information	No	Omitted
<i>3 Enforcement Mechanisms</i>		
Exogenous Enforcement	Yes	Yes
Endogenous Enforcement	Yes	Yes
Inadvertent Enforcement	Yes	Yes
<i>3 Signaling Mechanisms</i>		
Costly Signaling	Mixed	Inconclusive
Costly Implementation	Yes	Yes
Contradictory Signaling	Yes	Limited

See the notes in Table 1.1.

First, war is more likely when one side expects a significant increase in its future vulnerability that the opponent cannot commit *not* to exploit. The commitment theory of war – which combines the window theory and the commitment problem – resonates across several East Asian cases. Experiments based on the commitment-problem game show that the commitment problem tripled the incidence of conflict. Convergent findings from three different methods – formal, historical, and experimental – underline a similar lesson: Power shifts are perilous when an enforcement mechanism is missing. The strategic form of the commitment problem isolated in the models and experiments is extremely general, with potential implications for other forms of conflict that also involve

commitment issues – such as civil wars, domestic conflict, or frictions in international trade and finance.¹

Second, exogenous enforcement can reduce the risk of war. Leaders in crises are less war-prone when they see an external actor that can act as an enforcer. Conversely, they are more war-prone when they believe that exogenous enforcement is unlikely. The experiment and model in Chapter 2 highlight the powerful peace-promoting effect of the exogenous enforcement mechanism in its perfect form, with the computer as the external enforcer. When exogenous enforcement is removed, the incidence of war tripled in the experiment. Evidence from the case universe also highlights the peace-promoting effects of exogenous enforcement. But the case evidence also suggests that divergent beliefs about the existence of exogenous enforcement can *increase* the risk of war. When one side adopts a hard-line in the belief that war is unlikely because exogenous enforcement is likely, the other side may respond aggressively if it believes – in contrast – that exogenous enforcement *is* unlikely.

Third, inadvertent enforcement discourages war-prone behavior. Leaders try to avoid war when they expect war to create a strategic advantage for their third-party rival. This is an unintended consequence of third-party rivalry (at the international or domestic level) enforcing the peace despite the contradictory intent of the third party. Several historical

¹ See, e.g., Barbara Walter, *Committing to Peace: The Successful Settlement of Civil Wars* (Princeton, NJ: Princeton University Press, 2002); Daron Acemoglu and James Robinson, *Economic Origins of Dictatorship and Democracy* (New York: Cambridge University Press, 2006); Avner Greif, Paul Milgrom, and Barry Weingast, “Coordination, Commitment, and Enforcement: The Case of the Merchant Guild,” *Journal of Political Economy* 102, No. 4 (1994): 745-76; J. Lawrence Broz, “Political System Transparency and Monetary Commitment Regimes,” *International Organization* 56, No. 4 (2002): 861-87.

cases show that inadvertent enforcement can motivate war-avoidant behavior even in a militaristic environment where such behavior is least expected. In the experiments, inadvertent enforcement calms the commitment problem with decisive impact: the overall incidence of conflict is more than three times as high when inadvertent enforcement is absent than when it is present. The war-restraining effects of inadvertent enforcement – and the war-promoting effects of its removal – may also be found in several historical cases in East Asia. But the case universe also yields some contradictory evidence that highlight three caveats.² Firstly, states can sometimes use pacts or alliances to endogenously remove inadvertent enforcement. Secondly, inadvertent enforcement is weakened when leaders believe that war can be finished rapidly before the third party organizes itself for an attack. Thirdly, inadvertent enforcement can be overridden when the fear of a two-front war in the present is overridden by the fear of a worse two-front war in the future.

Fourth, endogenous enforcement through strategic restraint has peace-promoting effects. When a rising power exercises strategic or tactical restraint to moderate the power shift, the commitment problem and the risk of war are reduced as a consequence. Experimental results suggest that endogenous enforcement is not a rare phenomenon: in half of the cases, players in the lab willingly sacrificed their future military advantage to eliminate the potential power shift. This eliminated their power advantage in the future, but it also shut down the commitment problem in the present. The incidence of conflict is sharply reduced when endogenous enforcement is achieved: the overall incidence of conflict

² These caveats represent cases in which at least one of the assumptions behind inadvertent enforcement (Chapter 1) does not hold.

more than tripled when endogenous enforcement is absent than when it is present. Historical evidence from the case universe is also consistent with the observable implications of endogenous enforcement.

Fifth, there is no clear evidence that sinking costs makes the threat to use force more credible and deterrence more successful. The historical evidence for the sunk-cost signaling mechanism is limited, with a number of contradictory cases. It should be noted, of course, that there are few pure cases of a sunk-cost signal in international politics. The credibility effect of a sunk-cost signal may be confounded with the effects of other signals sent previously or simultaneously. To eliminate the confounders, I test the sunk-cost signaling mechanism with randomized experiments. It turns out that the positive credibility effect hypothesized by the mechanism is hard to detect even in an idealized confound-free setting. The experiments also reveal a surprising mismatch in behavior between signalers and receivers. Signalers randomly assigned with high resolve are much more likely to use a signal with sunk cost, but receivers do not necessarily respond in line with the signalers' expectation, despite the sunk cost suffered. The experiments suggest an unexpected asymmetry between the beliefs and behaviors of signalers and receivers – an asymmetry that leads to wasted resources and mutually undesirable outcomes.

Sixth, the credibility of a threat is strongly influenced by the perceived costs of implementing that threat. The historical evidence shows that across several cases, the threat to use force is less credible when the opponent believes that the threat is costly to implement. Experimental results corroborate the finding. In actual international crises as

well as in randomized survey experiments, observers discounted the threat to use force when they knew that it would be costly for the threatener to do what it threatened to do.

Seventh, credibility is compromised when signals of strong resolve are mixed with signals of weak resolve in the series of signals sent. A series of signals with contradictory messages is noisier and less credible. In the case universe, clear-cut historical cases are scarce, but the Korean War provides a suggestive case. The case also suggests how rational contradictions influence the calculation of credibility: the threat to use force becomes less credible if its implementation involves a rational contradiction with the choices made earlier in the crisis (“If China did not intervene earlier when it was less costly, why would China intervene now that intervention is *more* costly?”). A survey experiment also shows a negative relationship between contradictory signaling and signal credibility. The experiment uses stylized historical counterfactuals based on the Korean War case. We can never rerun history to answer counterfactual questions. But we can design experiments to test the theoretical expectations behind the counterfactuals.

What alternative ideas germinate from the East Asian case universe?

First, bargaining failure leading to war is more likely when concessions made in one bargaining dyad negatively affects relative bargaining power in another bargaining dyad. In such a situation, a mutually preferable bargain may not exist between two states bargaining over a valuable prize. This describes the bargaining failure between Japan and the U.S. on the eve of the Pacific War. When bargaining is “two-dimensional”, a peaceful

division of the prize cannot be guaranteed, even if we assume perfect information, full divisibility of the prize, and no commitment problems. This points us to a separate and distinct rationalist explanation for war. This also means that contrary to the conventional assumption, the information problem and the commitment problem do *not* exhaust “the full set of rationalist explanations that are both theoretically coherent and empirically plausible.”³

Second, war is more likely when conquest is suddenly easy and cheap. There is a sharp danger of war when the cost of conquest suddenly falls due to an exogenous shock. There are at least four “opportunistic wars” in the East Asian case universe. In these cases, leaders took the road to war after an exogenous change in the international environment that suddenly made conquest easy.

Third, several East Asian wars have roots in some form of domestic principal-agent problem. One manifestation of the principal-agent problem is when the national leader's political interests diverge from the national interest. Another manifestation is when the preference of the central government diverges from the preference of its agent on the ground. The East Asian cases suggest three pathways by which this principal-agent problem causes war: direct escalation by the agent, information misrepresentation by the agent to the center, and signaling distortion by the agent to the opponent state.

³ James Fearon, “Rationalist Explanations for War,” *International Organization* 49, No. 3 (1995): 380.

Fourth, there are three types of audience costs that may make war more likely. The first two types share the same mechanism: War is more likely if leaders expect high international or domestic audience costs and their opponent does not back down. The third type is personal audience costs – the political costs to a career politician or bureaucrat for opposing the position taken by powerful colleagues or superiors. War is more likely when an anti-war position involves personal audience costs. With personal audience costs, officials do not take positions based solely on their understanding of national interest and the available facts. They may avoid anti-war positions even though they believe in those positions. The problem of personal audience costs might be particularly pronounced in autocratic and militaristic regimes.

The four alternative ideas are inferred from the historical cases. They are empirically resonant, but require further tests. Future work may focus on clarifying their logic formally and testing their mechanisms experimentally.

War is an extremely complex phenomenon. Rationalist theories for war cannot fully explain how and why wars occur. A war is almost always mixed with both rationalist and non-rationalist elements. As scientists, it is important for us to specify and test the different mechanisms separately to understand the causal logic in complex cases. In this thesis, I specified six different mechanisms that underpin rationalist explanations for war. I studied their effects empirically through a series of randomized experiments, case studies of decision processes, and a comparative analysis of East Asian wars fought in the 20th century. The historical analysis reveals the power and limitations of the rationalist

theories. The experimental analysis provides confound-clean tests of the conflict mechanisms at the heart of these theories. Together, they establish a rationalist baseline for understanding the causes of war. On this baseline, non-rationalist factors can be introduced and tested to expand our knowledge on how and why wars occur.

Appendices

APPENDIX A1

Experimental Instructions for Chapter 2

The experiment is programmed and conducted with z-Tree (Fischbacher 2007). Subjects viewed the instructions on their computer screens. The instructions are reproduced fully as follows.

Welcome:

Welcome to the experiment!

The experiment will take one hour. If you follow the instructions and make good decisions, you might earn a considerable amount of money. Hence it is important that you read the instructions very carefully. All the money you earn is yours to keep, and will be paid to you, in cash, at the end of the experiment. Your confidentiality is assured.

Please do not communicate with other players during the experiment. If you have questions or need assistance, raise your hand and a monitor will come to you.

You should NOT look at the decisions of others, or talk or laugh or exclaim aloud in the experiment. You will be warned if you violate the rule the first time. If you violate the rule a second time, you will be asked to leave and you will not be paid.

Overview:

The experiment is divided into 5 SCENARIOS with a total of 17 ROUNDS. Each round is independent. The experiment begins with Scenario 1 and ends with Scenario 5.

In the experiment, you will be randomly divided into groups of 2 players. Your opponent will change at random after every round. You will NOT be matched with the same opponent twice in the same scenario.

Earnings:

Your dollar earnings for the experiment are determined as follows.

The computer will randomly choose 9 rounds out of the 17 rounds. Then it will sum up your total point earnings in all 9 chosen rounds. Each round has a total possible value of 10 points.

We will pay you

\$0.50 PER POINT

That is half a dollar, or fifty cents, for EVERY SINGLE POINT you earn. For example, 20 points will give you \$10.00 in cash.

The more points you win, the more money you receive.

Transition:

Let us begin the experiment.

Role:

In the experiment, you are a national leader facing an international crisis.

In this crisis, your country is bargaining with another country for a valuable Prize.

As a national leader, you will want to get the BEST DEAL for your country.

Pairing:

You will be randomly assigned as either Country A or Country B in each round.

Your opponent will change at random after every round. You will NOT be matched with the same opponent twice in the same scenario.

Transition:

Scenario 1 will start in a few seconds.

SCENARIOS 1 – 2

Payoffs:

You are a national leader bargaining with another country over a valuable Prize.

Both countries know that:

- The Prize is worth **10 points**
- If there is a war, both countries will pay the **costs of war**

[For the public-information group]

The cost of war is PUBLIC information:

- **You know your Opponent's cost of war**
- **Your Opponent knows your cost of war**

[For the private-information group]

The cost of war is PRIVATE information:

- **You do NOT know your Opponent's cost of war**
- **Your Opponent does NOT know your cost of war**

Costs:

To generate the costs of war, the computer will assign one of the values {0, 1, 2, 3, 4} to you.

Then, it will assign one of the values {0, 1, 2, 3, 4} to your Opponent.

Once generated, the cost of war will be FIXED for Scenarios 1 and 2.

Costs:

The computer has generated the costs of war.

The cost is FIXED for all rounds in Scenarios 1 and 2.

[For the public-information group]

Your Cost of War = **2 points**

Opponent's Cost of War = **2 points**

[For the private-information group]

Your Cost of War = **2 points**

Opponent's Cost of War = **Unknown**

War Costs:

War is always costly.

Whenever war is triggered, **you lose 2 points.**

[For the public-information group]

Whenever war is triggered, **your opponent loses 2 points.**

[For the private-information group]

Whenever war is triggered, **your opponent loses [unknown].**

Power Shift:

The game has 2 stages:

STAGE 1

In Stage 1, Country A and Country B are **EQUALLY POWERFUL**.

If they fight a war, each country will be able to seize 50% of the prize for itself. Hence, Country A gets 5 points and Country B gets 5 points.

But because war is costly, both countries will also **LOSE POINTS** based on its own cost of war.

STAGE 2

In Stage 2, Country A becomes **MORE POWERFUL** than Country B.

If they fight a war, Country A will be able to seize 70% of the prize for itself. Hence, Country A gets 7 points and Country B gets 3 points.

But because war is costly, both countries will also **LOSE POINTS** based on its own cost of war.

Stages:

The game has 2 stages. Here is how the game works:

STAGE 1:

Country A suggests how much of the prize it wants for itself.

- The Prize is worth 10 points. Country A can demand 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 points of the Prize.
- Country A's suggestion in Stage 1 is not implemented until it is confirmed in Stage 2.

Country B agrees or disagrees

- If Country B agrees, you **PROCEED TO STAGE 2**.
- If Country B disagrees, war is triggered and **THE GAME ENDS HERE**.

STAGE 2:

Country A confirms the number of points it wants.

If you are playing in the **ENFORCEMENT CONDITION**:

- The computer **WILL NOT ALLOW** Country A to make any changes in Stage 2.
- Hence, the amount which Country A suggested in Stage 1 **WILL BE THE SAME** amount that is confirmed in Stage 2.

If you are playing in the **NO-ENFORCEMENT CONDITION**:

- The computer **WILL ALLOW** Country A to make changes in Stage 2.
- Hence, the amount which Country A suggested in Stage 1 **MAY OR MAY NOT BE THE SAME** amount that is confirmed in Stage 2.

Country B agrees or disagrees

PAYOFFS:

If Country B agrees in both Stages 1 and 2,

- Country A will get the number of points it demanded.
- Country B will get what is left of the Prize (i.e. 10 points minus the points Country A has taken).
- Both countries **DO NOT LOSE POINTS** in the costs of war.

*If Country B disagrees (war) in **Stage 1**,*

- Country A gets **5 points minus its cost of war**. Country B gets **5 points minus its cost of war**.

*If Country B disagrees (war) in **Stage 2**,*

- Country A gets **7 points minus its cost of war**. Country B gets **3 points minus its cost of war**.

Note:

Note that:

- Country B can wage war in Stage 1 or Stage 2.
- Both countries can avoid the costs of war if they can reach an agreement in Stage 2.

Questions:

To ensure that you have read the instructions carefully, here are a few questions:

- (1) If war is triggered in Stage 1, each country gets 5 points minus its cost of war. Given your cost of war, how many points do you get if war is triggered in Stage 1? Please type your answer here: ____
- (2) If war is triggered in Stage 2, Country A gets 7 points minus its cost of war and Country B gets 3 points minus its cost of war. If you are COUNTRY B, how many points do you get if war is triggered in Stage 2? ____
- (3) You will be randomly assigned as either Country A or Country B in each round. Guess the % chance (probability) that you will be assigned as Country A. ____
- (4) One of the values {0, 1, 2, 3, 4} is selected as your Opponent's cost of war. Guess the % chance (probability) that your Opponent will get the value 1. The value 2? The value 3? ____

Flag:

You are currently in Scenario [1/2].

Prelude:

You are in Scenario [1/2]. Here is the scenario summary:

STAGE 1:

Country A suggests how much of the prize it wants for itself.

Country B agrees (go to Stage 2) **or disagrees** (war is triggered).

STAGE 2:

Country A confirms how much of the prize it wants for itself.

Country B agrees (bargain is made) **or disagrees** (war is triggered).

PAYOFFS:

If bargain is made,

- Country A gets what it confirmed in Stage 2.
- Country B gets what is left of the Prize.
- Both countries avoid the costs of war.

If war is triggered in Stage 1,

- Country A gets **5 points minus its cost of war.**
- Country B gets **5 points minus its cost of war.**

If war is triggered in Stage 2,

- Country A gets **7 points minus its cost of war.**
- Country B gets **3 points minus its cost of war.**

Group Assignment:

There is [ENFORCEMENT / NO ENFORCEMENT] in this scenario.

- The computer [WILL NOT ALLOW / WILL ALLOW] Country A to make changes in Stage 2.
- Hence, the amount which Country A suggested in Stage 1 [will be the SAME / MAY OR MAY NOT BE THE SAME] amount that is confirmed in Stage 2.

Country Assignment:

In this round, you are randomly assigned as a leader in Country [A/B].

Treatment Summary:

There is [ENFORCEMENT / NO ENFORCEMENT].

- Country A [cannot / can] make changes in Stage 2.
- The amount Country A suggested in Stage 1 [will be the SAME / MAY OR MAY NOT BE THE SAME] amount confirmed in Stage 2.

The cost of war is [PRIVATE / PUBLIC] information.

[Private-information group]

- Your Cost of War = **2 points** and Opponent's Cost of War = **Unknown.**
- You do NOT know your Opponent's cost of war
- Your Opponent does NOT know your cost of war

[Public-information group]

- Your Cost of War = **2 points** and Opponent's Cost of War = **2 points.**
- You know your Opponent's cost of war
- Your Opponent knows your cost of war

Actual Play:

[STAGE 1]

[Country A (decision)]

You are Country A.

Suggest how much of the prize you want by choosing ONE of the possibilities below.

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

You [CAN / CANNOT] CHANGE the amount in Stage 2.

Your Cost of War = 2 points. Opponent's Cost of War = [2 points / Unknown].

THIS STAGE – Both countries are equally powerful.

If war is triggered, you get **5 points minus your cost of war**.

Country B gets **5 points minus its cost of war**.

NEXT STAGE – You are more powerful.

If war is triggered, you get **7 points minus your cost of war**.

Country B gets **3 points minus its cost of war**.

[Country A (waiting)]

Please wait while Country B makes a decision.

[Country B (waiting)]

Please wait while Country A makes a decision.

[For Country B (decision)]

Country A has suggested the following:

“A gets [], B gets []”

If you agree, you proceed to Stage 2 in which Country A will confirm its demand.

Country A [CAN / CANNOT] CHANGE the amount in Stage 2.

If you disagree, war is triggered.

Your Cost of War = 2 points. Opponent's Cost of War = [2 points / Unknown].

THIS STAGE – Both countries are equally powerful.

If war is triggered, Country A gets **5 points minus its cost of war.**

You get **5 points minus your cost of war.**

NEXT STAGE – Country A is more powerful.

If war is triggered, Country A gets **7 points minus its cost of war.**

You get **3 points minus your cost of war.**

< **AGREE – Go to Stage 2** >

< **DISAGREE – Go to War** >

[If Country B disagrees]

Country B disagreed. War is triggered.

[STAGE 2]

[Country A (decision)] [Enforcement group]

You are Country A.

In Stage 1, you suggested the following:

“A gets [], B gets []”

You cannot change this decision. Click < **CONFIRM** > to continue.

< **CONFIRM** >

[Country A (decision)] [No-enforcement group]

You are Country A.

In Stage 1, you suggested the following:

“A gets [], B gets []”

You can change or confirm the suggested amount by choosing ONE of the possibilities below:

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

Your Cost of War = 2 points. Opponent's Cost of War = 2 points.

THIS STAGE – You are more powerful.
If war is triggered, you get **7 points minus your cost of war.**
Country B gets **3 points minus its cost of war.**

[For Country B (decision)]

Country A has confirmed the following:

“A gets [], B gets []”

If you agree, the bargain is made.

If you disagree, war is triggered.

Your Cost of War = 2 points. Opponent's Cost of War = [2 points / Unknown].

THIS STAGE – Country A is more powerful.
If war is triggered, Country A gets **7 points minus its cost of war.**
You get **3 points minus your cost of war.**

< AGREE – Bargain is Made >

< DISAGREE – Go to War >

Outcome:

[If Country B agrees]

Country B agreed.

Country A gets the points it demanded.

Country B gets what is left of the Prize.

[If Country B disagrees]

Country B disagreed. War is triggered.

Country A gets 7 points minus its cost of war.

Country B gets 3 points minus its cost of war.

Transition [After Round 5]:

Scenario 1 is over. Scenario 2 will start in a few seconds.

Transition [Start of Round 6]:

Scenario 2 is similar to Scenario 1 except for a change in the ENFORCEMENT CONDITION.

Repeat Screens from "Prelude" to "Outcome" in Rounds 2-10.
--

SCENARIO 3

Players are randomly divided into the enforcement and no-enforcement groups.

Transition:

Scenario 2 is over. Scenario 3 will start in a few seconds.

Comparison:

One major difference between Scenario 3 and the earlier scenarios is:

COUNTRY B CAN DECIDE TO WAGE WAR AT ANY TIME

Payoffs:

You are a national leader bargaining with another country over a valuable Prize.

Both countries know that:

- The Prize is worth 10 points
- If there is a war, both countries will pay the costs of war
- Cost of war is **2 points for all countries**. This cost is fixed throughout the remaining rounds. All countries know that the cost of war is 2 points for all countries.

War Costs:

War is always costly.

Whenever war is triggered, **you lose 2 points**.

Whenever war is triggered, **your opponent loses 2 points**.

Group Assignment:

There is [ENFORCEMENT / NO ENFORCEMENT] in this scenario.

- The computer [WILL NOT ALLOW / WILL ALLOW] Country A to make any changes in Stage 2.

- Hence, the amount which Country A suggested in Stage 1 [will be the SAME / MAY OR MAY NOT BE THE SAME] amount that is confirmed in Stage 2.

Stages:

The game has 2 stages. Here is how the game works:

STAGE 1:

Country A suggests how much of the prize it wants for itself. The Prize is worth 10 points.

Country A can demand 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 points of the Prize.

Country B has 3 options:

- WAIT to see Country A's demand, and then AGREE to go to Stage 2.
- WAIT to see Country A's demand, and then DISAGREE (go to war).
- DISAGREE (go to war) without seeing Country A's demand.

STAGE 2:

Country A confirms the number of points it wants.

- Since there is [ENFORCEMENT / NO ENFORCEMENT], the computer [WILL NOT / WILL] ALLOW Country A to make changes in Stage 2.
- Hence, the amount which Country A suggested in Stage 1 [WILL BE THE SAME / MAY OR MAY NOT BE THE SAME] amount that is confirmed in Stage 2.

Country B has 3 options:

- WAIT to see Country A's demand, and then AGREE to the bargain.
- WAIT to see Country A's demand, and then DISAGREE (go to war).
- DISAGREE (go to war) without seeing Country A's demand.

PAYOFFS:

If Country B agrees in both Stages 1 and 2,

- Country A will get the number of points it confirmed in Stage 2.
- Country B will get what is left of the Prize (i.e. 10 points minus the points Country A has taken).
- Both countries DO NOT LOSE POINTS in the costs of war.

If Country B disagrees (war) in Stage 1,

- Country A gets **5 points minus its cost of war**. Country B gets **5 points minus its cost of war**.

If Country B disagrees (war) in Stage 2,

- Country A gets **7 points minus its cost of war**. Country B gets **3 points minus its cost of war**.

Comparison:

To summarize, one major difference between Scenario 3 and the earlier scenarios is:

COUNTRY B CAN DECIDE TO WAGE WAR AT ANY TIME

Note:

Note that:

- Country B can wage war at any time.
- Both countries can avoid the costs of war if they can reach an agreement in Stage 2.

Prelude:

You are in Scenario 3. Here is the scenario summary:

STAGE 1:

Country A suggests how much of the prize it wants for itself.

Country B has 3 options:

- WAIT to see Country A's demand, and then AGREE to go to Stage 2.
- WAIT to see Country A's demand, and then DISAGREE (go to war).
- DISAGREE (go to war) without seeing Country A's demand.

STAGE 2:

Country A confirms how much of the prize it wants for itself.

Country B has 3 options:

- WAIT to see Country A's demand, and then AGREE to the bargain.
- WAIT to see Country A's demand, and then DISAGREE (go to war).
- DISAGREE (go to war) without seeing Country A's demand.

PAYOFFS:

If bargain is made,

- Country A gets what it confirmed in Stage 2.
- Country B gets what is left of the Prize.
- Both countries avoid the costs of war.

Country B can trigger war at any time:

If war is triggered in **Stage 1**,

- Country A gets **5 points minus its cost of war.**
- Country B gets **5 points minus its cost of war.**

If war is triggered in **Stage 2**,

- Country A gets **7 points minus its cost of war.**
- Country B gets **3 points minus its cost of war.**

Flag:

You are currently in Scenario 3.

Country Assignment:

In this round, you are randomly assigned as a leader in Country [A/B].

Treatment Summary:

There is [ENFORCEMENT / NO ENFORCEMENT].

- Country A [**cannot / can**] **make changes** in Stage 2.
- The amount Country A suggested in Stage 1 [will be the SAME / MAY OR MAY NOT BE THE SAME] amount confirmed in Stage 2.

The cost of war is PUBLIC information.

- Your Cost of War = **2 points** and Opponent's Cost of War = **2 points.**
- You know your Opponent's cost of war
- Your Opponent knows your cost of war

Actual Play:

[STAGE 1]

[Country A (decision)]

You are Country A.

Suggest how much of the prize you want by choosing ONE of the possibilities below.

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

You [CAN / CANNOT] CHANGE the amount in Stage 2.

Your Cost of War = 2 points. Opponent's Cost of War = 2 points.

Country B can trigger war at any time.

THIS STAGE – Both countries are equally powerful.

If war is triggered, you get **5 points minus your cost of war.**

Country B gets **5 points minus its cost of war.**

NEXT STAGE – You are more powerful.

If war is triggered, you get **7 points minus your cost of war.**

Country B gets **3 points minus its cost of war.**

[Country A (waiting)]

Please wait while Country B makes a decision.

[Country B (waiting)]

Country A has not made its decision yet.

You can either WAIT to see Country A's demand, or DISAGREE without seeing Country A's demand.

To wait, just do nothing until Country A has made its decision. To disagree, click the button below.

THIS STAGE – Both countries are equally powerful.
If war is triggered, Country A gets **5 points minus its cost of war.**
You get **5 points minus your cost of war.**

NEXT STAGE – Country A is more powerful.
If war is triggered, Country A gets **7 points minus its cost of war.**
You get **3 points minus your cost of war.**

There is [ENFORCEMENT / NO ENFORCEMENT] in this scenario.

< DISAGREE – Go to War >

[For Country B (decision)]

Country A has suggested the following:

“A gets [], B gets []”

If you agree, you proceed to Stage 2 in which Country A will confirm its demand.

Country A [CAN / CANNOT] CHANGE the amount in Stage 2.

If you disagree, war is triggered.

Your Cost of War = 2 points. Opponent’s Cost of War = 2 points.

THIS STAGE – Both countries are equally powerful.
If war is triggered, Country A gets **5 points minus its cost of war.**
You get **5 points minus your cost of war.**

NEXT STAGE – Country A is more powerful.
If war is triggered, Country A gets **7 points minus its cost of war.**
You get **3 points minus your cost of war.**

< AGREE – Go to Stage 2 >

< DISAGREE – Go to War >

[STAGE 2]

[Country A (decision)] [Enforcement group]

You are Country A.

In Stage 1, you suggested the following:

“A gets [], B gets []”

You cannot change this decision. Click < CONFIRM > to continue.

< CONFIRM >

[Country A (decision)] [No-enforcement group]

You are Country A.

In Stage 1, you suggested the following:

“A gets [], B gets []”

You can change or confirm the suggested amount by choosing ONE of the possibilities below:

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

Your Cost of War = 2 points. Opponent’s Cost of War = 2 points.

Country B can trigger war at any time.

THIS STAGE – You are more powerful.

If war is triggered, you get 7 points minus your cost of war.

Country B gets 3 points minus its cost of war.

[Country B (waiting)]

Country A has not made its decision yet.

You can either **WAIT** to see Country A's demand, or **DISAGREE** without seeing Country A's demand.

To wait, just do nothing until Country A has made its decision. To disagree, click the button below.

THIS STAGE – Country A is more powerful.
If war is triggered, Country A gets **7 points minus its cost of war**.
You get **3 points minus your cost of war**.

There is [ENFORCEMENT / NO ENFORCEMENT] in this scenario.

< **DISAGREE – Go to War** >

[For Country B (decision)]

Country A has confirmed the following:

“A gets [], B gets []”

If you agree, the bargain is made.

If you disagree, war is triggered.

Your Cost of War = 2 points. Opponent's Cost of War = [2 points / Unknown].

THIS STAGE – Country A is more powerful.
If war is triggered, Country A gets **7 points minus its cost of war**.
You get **3 points minus your cost of war**.

< **AGREE – Bargain is Made** >

< **DISAGREE – Go to War** >

Outcome:

[If Country B agrees]

Country B agreed.

Country A gets the points it demanded.

Country B gets what is left of the Prize.

[If Country B disagrees]

Country B disagreed. War is triggered.

Country A gets 7 points minus its cost of war.

Country B gets 3 points minus its cost of war.

Repeat Screens from “Prelude” to “Outcome” in Rounds 12-15

SCENARIO 4

Players are randomly divided into a “30-seconds” group and a “60-seconds” group.

Transition:

Scenario 3 is over. Scenario 4 will start in a few seconds.

Comparison:

One major difference between Scenario 3 and Scenario 4 is:

PAYOFFS FOR WAR DEPEND ON THE TIMER CLOCK

Payoffs:

You are a national leader bargaining with another country over a valuable Prize.

Both countries know that:

- The Prize is worth 10 points
- If there is a war, both countries will pay the costs of war
- Cost of war is 2 points for all countries. This cost is fixed throughout the remaining rounds. All countries know that the cost of war is 2 points for all countries.

War Costs:

War is always costly.

Whenever war is triggered, **you lose 2 points.**

Whenever war is triggered, **your opponent loses 2 points.**

Scenario Condition I:

There is **NO ENFORCEMENT** in this scenario.

- The computer **WILL ALLOW** Country A to make changes in Stage 2.
- Hence, the amount which Country A suggested in Stage 1 **MAY OR MAY NOT BE THE SAME** amount that is confirmed in Stage 2.

Scenario Condition II:

In this scenario:

COUNTRY B CAN DECIDE TO WAGE WAR AT ANY TIME.

Scenario Condition III:

In this scenario, the payoffs for war depend on the **TIMER CLOCK**.

UNTIL THE [30TH /60TH] SECOND, Country A and Country B are **EQUALLY POWERFUL**. If war is triggered,

- Country A gets 5 points minus its cost of war.
- Country B gets 5 points minus its cost of war.

AFTER THE [30TH /60TH] SECOND, Country A becomes **MORE POWERFUL**. If war is triggered,

- Country A gets 7 points minus its cost of war.
- Country B gets 3 points minus its cost of war.

Elaboration I:

Unlike earlier scenarios, the payoffs for war do **NOT** depend on whether you are in Stage 1 or Stage 2.

- As long as Country B disagrees (war) **AT OR BEFORE the [30TH /60TH] SECOND**, Country A gets 5 points minus its cost of war and Country B gets 5 points minus its cost of war. It **DOES NOT MATTER** whether war occurs in Stage 1 or Stage 2.

- As long as Country B disagrees (war) **AFTER the [30TH /60TH] SECOND**, Country A gets 7 points minus its cost of war and Country B gets 3 points minus its cost of war. It **DOES NOT MATTER** whether war occurs in Stage 1 or Stage 2.

The Timer Clock **does NOT stop ticking** until the end of the round. It resets only at the end of the round. It **does NOT reset** for each stage.

Stages:

The game has 2 stages. Here is how the game works:

STAGE 1:

Country A suggests how much of the prize it wants for itself. The Prize is worth 10 points.

Country A can demand 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 points of the Prize.

Country B has 3 options:

- WAIT to see Country A's demand, and then AGREE to go to Stage 2.
- WAIT to see Country A's demand, and then DISAGREE (go to war).
- DISAGREE (go to war) without seeing Country A's demand.

STAGE 2:

Country A confirms the number of points it wants.

- Since there is NO ENFORCEMENT, the computer WILL ALLOW Country A to make changes in Stage 2.
- Hence, the amount which Country A suggested in Stage 1 MAY OR MAY NOT BE THE SAME amount that is confirmed in Stage 2.

Country B has 3 options:

- WAIT to see Country A's demand, and then AGREE to the bargain.
- WAIT to see Country A's demand, and then DISAGREE (go to war).
- DISAGREE (go to war) without seeing Country A's demand.

PAYOFFS:

If Country B agrees in both Stages 1 and 2,

- Country A will get the number of points it confirmed in Stage 2.
- Country B will get what is left of the Prize (i.e. 10 points minus the points Country A has taken).
- Both countries **DO NOT LOSE POINTS** in the costs of war.

If Country B disagrees (war) at or before the [30th / 60th] second,

- Country A gets **5 points minus its cost of war**. Country B gets **5 points minus its cost of war**.

If Country B disagrees (war) after the [30th / 60th] second,

- Country A gets **7 points minus its cost of war**. Country B gets **3 points minus its cost of war**.

Comparison:

To summarize, one major difference between Scenario 3 and Scenario 4 is:

Payoffs for war depend on the TIMER CLOCK

Note:

Note that:

- Country B can wage war at any time.
- Both countries can avoid the costs of war if they can reach an agreement in Stage 2.

Prelude:

You are in Scenario 4. Here is the scenario summary:

STAGE 1:

Country A SUGGESTS how much of the prize it wants for itself.

Country B has 3 options:

- WAIT to see Country A's demand, and then AGREE to go to Stage 2.
- WAIT to see Country A's demand, and then DISAGREE (go to war).
- DISAGREE (go to war) without seeing Country A's demand.

STAGE 2:

Country A CONFIRMS how much of the prize it wants for itself.

Country B has 3 options:

- WAIT to see Country A's demand, and then AGREE to the bargain.
- WAIT to see Country A's demand, and then DISAGREE (go to war).
- DISAGREE (go to war) without seeing Country A's demand.

PAYOFFS:

If bargain is made,

- Country A gets what it confirmed in Stage 2.
- Country B gets what is left of the Prize.
- Both countries avoid the costs of war.

Country B can trigger war at any time:

If war is triggered in *at or before the [30th / 60th] second,*

- Country A gets **5 points minus its cost of war.**
- Country B gets **5 points minus its cost of war.**

If war is triggered in *after the [30th / 60th] second,*

- Country A gets **7 points minus its cost of war.**
- Country B gets **3 points minus its cost of war.**

Reminder:

Remember:

Payoffs for War Depend on the Timer Clock Regardless of Whether You are in Stage 1 or 2.

The Timer Clock does NOT stop ticking until the end of the round.
It does NOT reset for each stage.

Flag:

You are currently in Scenario 4.

Country Assignment:

In this round, you are randomly assigned as a leader in Country [A/B].

Treatment Summary:

There is NO ENFORCEMENT.

- Country A **can make changes** in Stage 2.

- The amount Country A suggested in Stage 1 MAY OR MAY NOT BE THE SAME amount confirmed in Stage 2.

The cost of war is PUBLIC information.

- Your Cost of War = **2 points** and Opponent's Cost of War = **2 points**.
- You know your Opponent's cost of war
- Your Opponent knows your cost of war

Actual Play:

[STAGE 1]

[Country A (decision)]

TIMER CLOCK: [sec]

You are Country A.

Suggest how much of the prize you want by choosing ONE of the possibilities below.

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

You CAN CHANGE the amount in Stage 2.

Your Cost of War = 2 points. Opponent's Cost of War = 2 points.

Country B can trigger war at any time.

AT/BEFORE [30TH / 60TH] SECOND – Both countries are equally powerful.

If war is triggered, you get 5 points minus your cost of war.

Country B gets 5 points minus its cost of war.

AFTER [30TH / 60TH] SECOND – You are more powerful.

If war is triggered, you get **7 points minus your cost of war.**
Country B gets **3 points minus its cost of war.**

[Country A (waiting)]

Please wait while Country B makes a decision.

[Country B (waiting)]

Country A has not made its decision yet.

You can either WAIT to see Country A's demand, or DISAGREE without seeing Country A's demand.

To wait, just do nothing until Country A has made its decision. To disagree, click the button below.

AT/BEFORE [30TH / 60TH] SECOND – Both countries are equally powerful.
If war is triggered, Country A gets **5 points minus its cost of war.**
You get **5 points minus your cost of war.**

AFTER [30TH / 60TH] SECOND – Country A is more powerful.
If war is triggered, Country A gets **7 points minus its cost of war.**
You get **3 points minus your cost of war.**

There is NO ENFORCEMENT in this scenario.

< DISAGREE – Go to War >

[For Country B (decision)]

Country A has suggested the following:

“A gets [], B gets []”

If you agree, you proceed to Stage 2 in which Country A will confirm its demand.

Country A CAN CHANGE the amount in Stage 2.

If you disagree, war is triggered.

Your Cost of War = 2 points. Opponent's Cost of War = 2 points.

AT/BEFORE [30TH / 60TH] SECOND – Both countries are equally powerful.

If war is triggered, Country A gets **5 points minus its cost of war.**
You get **5 points minus your cost of war.**

AFTER [30TH / 60TH] SECOND – Country A is more powerful.
If war is triggered, Country A gets **7 points minus its cost of war.**
You get **3 points minus your cost of war.**

< **AGREE** – Go to Stage 2 >
< **DISAGREE** – Go to War >

[STAGE 2]

[Country A (decision)]

You are Country A.

In Stage 1, you suggested the following:

“A gets [], B gets []”

You can change or confirm the suggested amount
by choosing ONE of the possibilities below:

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

Your Cost of War = 2 points. Opponent’s Cost of War = 2 points.

Country B can trigger war at any time.

AT/BEFORE [30TH / 60TH] SECOND – Both countries are equally powerful.
If war is triggered, you get **5 points minus your cost of war.**
Country B gets **5 points minus its cost of war.**

AFTER [30TH / 60TH] SECOND – You are more powerful.

If war is triggered, you get **7 points minus your cost of war.**
Country B gets **3 points minus its cost of war.**

[Country B (waiting)]

Country A has not made its decision yet.

You can either **WAIT** to see Country A's demand, or **DISAGREE** without seeing Country A's demand.

To wait, just do nothing until Country A has made its decision. To disagree, click the button below.

AT/BEFORE [30TH / 60TH] SECOND – Both countries are equally powerful.

If war is triggered, Country A gets **5 points minus its cost of war.**
You get **5 points minus your cost of war.**

AFTER [30TH / 60TH] SECOND – Country A is more powerful.
If war is triggered, Country A gets **7 points minus its cost of war.**
You get **3 points minus your cost of war.**

< DISAGREE – Go to War >

[For Country B (decision)]

Country A has confirmed the following:

“A gets [], B gets []”

If you agree, the bargain is made.

If you disagree, war is triggered.

Your Cost of War = 2 points. Opponent's Cost of War = [2 points / Unknown].

AT/BEFORE [30TH / 60TH] SECOND – Both countries are equally powerful.

If war is triggered, Country A gets **5 points minus its cost of war.**
You get **5 points minus your cost of war.**

AFTER [30TH / 60TH] SECOND – Country A is more powerful.
If war is triggered, Country A gets **7 points minus its cost of war.**
You get **3 points minus your cost of war.**

< AGREE – Bargain is Made >

< DISAGREE – Go to War >

Outcome:

[If Country B agrees]

Country B agreed.

Country A gets the points it demanded.

Country B gets what is left of the Prize.

[If Country B disagrees]

Country B disagreed. War is triggered.

Country A gets 7 points minus its cost of war.

Country B gets 3 points minus its cost of war.

Risk-Aversion Test

Role:

In this scenario, you are a national leader facing a series of decisions between two options.

As a national leader, you will want to get the BEST DEAL for your country.

Payoffs:

You must make a series of decisions between two options.

OPTION A will pay an amount of points with certainty.

OPTION B will pay 0 point with 50% probability and 10 points with 50% probability.

The computer will randomly select one of the decision problems and count it as your points for this round.

Decision:

OPTION A	OPTION B	Your Choice
2.50 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
3.00 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
3.50 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
4.00 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
4.50 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
5.00 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
5.50 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
6.00 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
6.50 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
7.00 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B
7.50 points	0 with 50% probability, 10 points with 50% probability	A <input type="checkbox"/> <input type="checkbox"/> B

Outcome:

Decision was randomly selected from the list of decisions you just made. In that decision, you had the choice between points and the risky gamble of 10 points with 50% chance and 0 points with 50% chance.

You chose the [first option / risky gamble]. Your score for this round is points.

Press the OK button to end the experiment.

< OK >

Results

The experiment has ended.

The computer will now calculate your total earnings for the experiment. It will randomly choose 9 rounds out of the 17 rounds you played. Then it will sum up your total point earnings in all 9 chosen rounds. You earn \$0.50 per point.

Earnings:

Period	Profit
[]	[] points
<i>[Repeat first row for 17 rows]</i>	

Your scores in each round are shown here. The computer has randomly chosen the following 9 rounds for payment:

Rounds: [...]

Your total score from the 9 randomly-chosen rounds is [] points, which is [] in cash.

Transition to Questionnaire:

Thank you for participating in this experiment!

Please allow us 5-10 minutes to collate the earnings data and prepare the cash payment for you. Meanwhile, please take the time to fill in a short questionnaire.

APPENDIX A2

Experimental Instructions for Chapter 4 (Experiments 1, 2 and 4)

The experiment is programmed and conducted with z-Tree (Fischbacher 2007). Subjects viewed the instructions on their computer screens. The instructions are reproduced fully as follows.

Welcome:

Welcome to the experiment!

The experiment will take one hour. If you follow the instructions and make good decisions, you might earn a considerable amount of money. Hence it is important that you read the instructions very carefully. All the money you earn is yours to keep, and will be paid to you, in cash, at the end of the experiment. Your confidentiality is assured.

Please do not communicate with other players during the experiment. If you have questions or need assistance, raise your hand and an assistant will come to you.

You should NOT look at the decisions of others, or talk or laugh or exclaim aloud in the experiment. You will be warned if you violate the rule the first time. If you violate the rule a second time, you will be asked to leave and you will not be paid.

Overview:

The experiment has multiple SCENARIOS with a total of 15 rounds.

Each scenario is DIFFERENT and INDEPENDENT.

Each scenario has one or more rounds. Each round is independent.
The outcome in each round does NOT affect the outcome in another round.

Earnings:

Your dollar earnings for the experiment are determined as follows.

The computer will randomly choose 9 rounds out of the 15 rounds.
Then it will sum up your total point earnings in all 9 chosen rounds.
Each round has a total possible value of 10 points.

We will pay you

\$0.50 PER POINT

That is half a dollar, or fifty cents, for EVERY SINGLE POINT you earn.
For example, every 10 points you earn will be \$5.00 in cash.

HOW MUCH YOU EARN DEPENDS ON YOUR PERFORMANCE.

The more points you win, the more money you receive.

Matching:

Participants in this experiment will be randomly matched to each other as opponents.

Your opponent will change at random after every round.

Transition:

Scenario 1 will start in a few seconds.

Role:

You are a national leader facing an international crisis.

In this crisis, your country is bargaining with another country for a valuable Prize.

Pairing:

You will be randomly assigned as ONE of the following countries:

- Country A
- Country B
- *[For those in the inadvertent-enforcement condition, add: Country C]*

Country A and Country B are bargaining over a valuable Prize. Everyone knows that:

- The Prize is worth **10 points**
- If there is a war, each country loses **2 points in the costs of war**

[For those in the inadvertent-enforcement condition, add: Country C is not involved in the bargaining]

Power Shift:

The game has 2 stages:

STAGE 1

In Stage 1, Country A and Country B are **EQUALLY POWERFUL.**

If they fight a war, each country can seize 50% of the prize for itself.
Hence, Country A gets 5 points and Country B gets 5 points.

But because war is costly, both countries will lose 2 points in the cost of war. Hence:

- Country A ends up with **3 points**
- Country B ends up with **3 points**

STAGE 2

In Stage 2, Country A becomes **MORE POWERFUL** than Country B.

If they fight a war, Country A can seize 70% of the prize for itself.
Hence, Country A gets 7 points and Country B gets 3 points.

But because war is costly, both countries will lose 2 points in the cost of war. Hence:

- Country A ends up with **5 points**
- Country B ends up with **1 point**

[For those in the inadvertent-enforcement treatment]

However, your war payoffs will change in Country C enters the war.

Country C can choose to attack Countries A and B there is a war between A and B in stage 1 or stage 2.

If Country C decides to attack, it will win a decisive victory and take over everything earned by A and B. (This is because A and B will already be weakened from fighting each other.) Hence:

- A ends up with **0 points**
- B ends up with **0 points**
- C ends up with **6 points**

If Country C decides NOT to attack, it gets 0 points. The payoffs for A and B remain unchanged.

Of course, Country C cannot choose to enter a war if there is **NO** war between A and B. Hence, C cannot choose to attack when there is no war between A and B.

Stages:

The game has 2 stages. Here is how the game works:

STAGE 1:

Country A suggests how much of the prize it wants for itself.

- The Prize is worth 10 points. Country A can demand 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 points of the Prize.
- Country A's suggestion in Stage 1 is **not implemented** until it is confirmed in Stage 2.

Country B agrees or disagrees

- If Country B agrees, you PROCEED TO STAGE 2.
- If Country B disagrees, war is triggered. *[For those in the inadvertent-enforcement treatment, add: Country C decides whether or not to attack.]* THE GAME ENDS HERE.

STAGE 2:

Country A confirms the number of points it wants.

- Country A **CAN CHANGE its demand** in Stage 2.
- Hence, the amount which Country A suggested in Stage 1 **may or may not be the same** as the amount confirmed in Stage 2.

Country B agrees or disagrees

PAYOFFS:

If Country B agrees in both Stages 1 and 2,

- Country A will get the number of points it confirmed in Stage 2.
- Country B will the remainder of the Prize (i.e. 10 points minus the points A has taken).
- Both countries do not lose points in the costs of war.
- *[For those in the inadvertent-enforcement treatment, add: Country C cannot choose to attack because there is no war.]*

If Country B disagrees (war) in Stage 1,

- *[For inadvertent-enforcement treatment: A gets 0 points, B gets 0 points, C gets 6 points, if C chooses to attack.]*
- A gets 3 points, B gets 3 points *[For inadvertent-enforcement treatment, add: C gets 0 points, if C chooses NOT to attack]*

If Country B disagrees (war) in Stage 2,

- *[For inadvertent-enforcement treatment: A gets 0 points, B gets 0 points, C gets 6 points, if C chooses to attack.]*
- *A gets 5 points, B gets 1 point [For inadvertent-enforcement treatment, add: C gets 0 points, if C chooses NOT to attack]*

Questions: [For control group]

To ensure that you have read the instructions carefully, here are a few questions:

If war is triggered in Stage 1: A gets **3 points**, B gets **3 points**

- True
- False

If war is triggered in Stage 2: A gets **5 points**, B gets **1 point**

- True
- False

Country A CAN CHANGE its demand in Stage 2. Hence, the amount that A suggests in Stage 1 may or may not be the same amount that A confirms in Stage 2.

- True
- False

Questions: [For inadvertent-enforcement treatment group]

To ensure that you have read the instructions carefully, here are a few questions:

If war is triggered (in Stage 1 or Stage 2): A gets **0 points**, B gets **0 points**, C gets **6 points**, if Country C chooses to attack

- True
- False

If war is triggered (in Stage 1 or Stage 2): C always gets **0 points** if it chooses NOT to attack

- True
- False

Country A CAN CHANGE its demand in Stage 2. Hence, the amount that A suggests in Stage 1 may or may not be the same amount that A confirms in Stage 2.

- True
- False

Country C CANNOT choose to attack if there is NO war between Countries A and B

- True
- False

[Answers are revealed]

*** Scenario:**

You are currently in Scenario 1

*** Player Assignment:**

In this round, you are randomly assigned as a leader in Country [A/B/C]

Prelude Summary:

Here is a summary of the game. You will make your decision in the next screen.

STAGE 1:

Country A suggests how much of the prize it wants for itself.

Country B agrees (go to Stage 2) **or disagrees** (war is triggered).

[For those in the inadvertent-enforcement treatment, add:

If war is triggered, Country C decides whether or not to attack.]

STAGE 2:

Country A confirms how much of the prize it wants for itself.

Country B agrees (a bargain is made) **or disagrees** (war is triggered).

[For those in the inadvertent-enforcement treatment, add:

If war is triggered, Country C decides whether or not to attack.]

PAYOFFS:

If a bargain is made,

- A gets what it confirmed in Stage 2.
- B gets the remainder of the Prize.
- Both countries avoid the costs of war.
- *[For those in the inadvertent-enforcement treatment, add: Country C cannot attack because there is no war.]*

If war is triggered in **Stage 1**,

- *[For inadvertent-enforcement treatment: A gets 0 points, B gets 0 points, C gets 6 points, if C chooses to attack.]*
- A gets 3 points, B gets 3 points *[For inadvertent-enforcement treatment, add: C gets 0 points, if C chooses NOT to attack]*

If war is triggered in **Stage 2**,

- **[For inadvertent-enforcement treatment: A gets 0 points, B gets 0 points, C gets 6 points, if C chooses to attack.]**
- **A gets 5 points, B gets 1 point [For inadvertent-enforcement treatment, add: C gets 0 points, if C chooses NOT to attack]**

Please take a moment to think about what you will do

*** Actual Play:**

[STAGE 1]

[Country A (decision)]

You are Country A in Stage 1.

Suggest how much of the prize you want by choosing **ONE** of the possibilities below.

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

You CAN CHANGE the amount in Stage 2.

THIS STAGE – Both countries are equally powerful.

If war is triggered, you get **3 points** and Country B gets **3 points** after deducting the costs of war

NEXT STAGE – You are more powerful.

If war is triggered, you get **5 points** and Country B gets **1 point** after deducting the costs of war

[For inadvertent-enforcement treatment, add:

However, if war is triggered, Country C can choose to attack.

If Country C attacks: You and Country B get 0 points. Country C gets 6 points.

If Country C does not attack: You and Country B are unaffected. Country C gets **0 points**. /

[Country A (waiting)]

Please wait while Country B makes a decision.

[Country B (waiting)]

Please wait while Country A makes a decision.

[For Country B (decision)]

This is Stage 1.

Country A has suggested the following:

“A gets [], B gets []”

If you agree, you proceed to Stage 2 in which Country A will confirm its demand.

Country A CAN CHANGE the amount in Stage 2.

If you disagree, war is triggered in this stage.

How would you respond to Country A?

< AGREE – Go to Stage 2 >

< DISAGREE – Go to War >

THIS STAGE – Both countries are equally powerful.

If war is triggered, you get **3 points** and Country A gets **3 points** after deducting the cost of war

NEXT STAGE – Country A becomes more powerful.

If war is triggered, you get **1 point** and Country A gets **5 points** after deducting the cost of war.

[For inadvertent-enforcement treatment, add:

However, **if war is triggered, Country C can choose to attack.**

If Country C attacks: You and Country A get **0 points**. Country C gets **6 points**.

If Country C does not attack: You and Country A are unaffected. Country C gets **0 points**./

[For Country C (decision)]

Please choose whether or not to attack if Countries A and B fight each other.

< **Attack** >
< **Do not attack** >

If you **attack**
You get **6 points**

If you **do not attack**,
You get **0 points**

[STAGE 2]

[Country A (decision)]

You are Country A.

In Stage 1, you suggested the following:

“A gets [], B gets []”

You can change or confirm the amount by choosing ONE of the possibilities below:

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

THIS STAGE – You are more powerful.

If war is triggered, you get **5 points** and Country B gets **1 point** after deducting the costs of war

[For inadvertent-enforcement treatment, add:

However, if war is triggered, Country C can choose to attack.

If Country C attacks: You and Country B get **0 points**. Country C gets **6 points**.

If Country C does not attack: You and Country B are unaffected. Country C gets **0 points**.]

[For Country B (decision)]

Country A has confirmed the following:

“A gets [], B gets []”

If you agree, the bargain is made.

If you disagree, war is triggered.

< AGREE – Bargain is Made >

< DISAGREE – Go to War >

THIS STAGE – Country A is more powerful.

If war is triggered, you get **1 point** and Country A gets **5 points** after deducting the cost of war.

[For inadvertent-enforcement treatment, add:

However, **if war is triggered, Country C can choose to attack.**

If Country C attacks: You and Country A get **0 points**. Country C gets **6 points**.

If Country C does not attack: You and Country A are unaffected. Country C gets **0 points**.]

*** Outcome:**

[If Country B agreed]

Country B agreed.

Country A gets the points it demanded.

Country B gets what is left of the Prize.

[If Country B disagreed (Control Group)]

Country B disagreed. War is triggered in Stage [1/2].

Country A gets [3/5] points.

Country B gets [3/1] points.

[If Country B disagreed and C did not attack under inadvertent enforcement]

Country B disagreed. War is triggered in Stage [1/2].

Country A gets [3/5] points.

Country B gets [3/1] points.

Country C did not attack.

[If Country B disagreed and C attacked under inadvertent enforcement]

Country B disagreed. War is triggered.

Country C attacks. Country C gets 6 points.

Country A gets 0 points.

Country B gets 0 points.

[For Player C under inadvertent enforcement]

[You chose to attack. You get 6 points if there is a war between Countries A and B, and 0 points if otherwise. / You chose not to attack. You get 0 points.]

Repeat Sections with * in Rounds 2-5.

Transition [After Round 5]:

Scenario 1 is over.

Scenario 2 will start in a few seconds.

Transition [Start of Round 6]:

Scenario 2 has the same bargaining procedure as Scenario 1:

In Stage 1:

- **A suggests** how much of the prize it wants for itself.
- **B agrees** (go to Stage 2) or **disagrees** (war is triggered; game ends in Stage 1).

In Stage 2:

- **A confirms** how much of the prize it wants for itself.
- **B agrees** (a bargain is made) or **disagrees** (war is triggered).

Third-Player:

In Scenario 2:

There is a **THIRD COUNTRY (Country C)** in the game

Country C can attack Country B, but it CANNOT attack Country A

You will be randomly assigned as one of the following: **Country A, B, or C**

Asymmetric Inadvertent Enforcement:

Country C can choose to attack Country B if there is a war between A and B in stage 1 or stage 2.

If Country C decides to attack, it will win a decisive victory and take over everything earned by Country B. (This is because B will already be weakened from fighting with A.) A is unaffected. Hence:

- A ends up with **3 points**
- B ends up with **0 points**
- C ends up with **3 points**

If Country C decides NOT to attack, it gets 0 points. The payoffs for A and B remain unchanged.

Note that:

Country C can only attack Country B. It CANNOT attack Country A.

C can only attack B **if there is a war between A and B.**

C cannot attack B when there is **NO** war between A and B.

Payoffs:

The payoffs for war are as follows:

If Country B disagrees (war) in Stage 1,

- A gets **3 points**, B gets **0 points**, C gets **3 points**, if Country C chooses to attack B
- A gets **3 points**, B gets **3 points**, C gets **0 points**, if Country C chooses NOT to attack B

If Country B disagrees (war) in Stage 2,

- A gets **5 points**, B gets **0 points**, C gets **1 point**, if Country C chooses to attack B
- A gets **5 points**, B gets **1 point**, C gets **0 points**, if Country C chooses NOT to attack B

If Country B agrees in both Stages 1 and 2,

- Country C cannot choose to attack because there is no war

Summary:

To summarize:

C always gets 0 points if it chooses NOT to attack

B always gets 0 points if it is attacked by C

C can attack B, but it cannot attack A. Hence, A is NOT affected.

**** Scenario:**

You are currently in Scenario 2

**** Player Assignment:**

In this round, you are randomly assigned as a leader in Country [A/B/C]

**** Actual Play:**

[STAGE 1]

[Country A (decision)]

You are Country A in Stage 1.

Suggest how much of the prize you want by choosing ONE of the possibilities below.

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

You CAN CHANGE the amount in Stage 2.

THIS STAGE – Both countries are equally powerful.
If war is triggered, you get **3 points** and Country B gets **3 points** after deducting the costs of war

NEXT STAGE – You are more powerful.
If war is triggered, you get **5 points** and Country B gets **1 point** after deducting the costs of war

However, **if war is triggered, Country C can choose to attack Country B.**
If Country C attacks B: Country B gets **0 points**. Country C gets **3 points** (in stage 1) or **1 point** (in stage 2). You are unaffected.
If Country C does not attack B: You and Country B are unaffected. Country C gets **0 points**.

[Country A (waiting)]

Please wait while Country B makes a decision.

[Country B (waiting)]

Please wait while Country A makes a decision.

[For Country B (decision)]

This is Stage 1.

Country A has suggested the following:

“A gets [], B gets []”

If you agree, you proceed to Stage 2 in which Country A will confirm its demand.

Country A CAN CHANGE the amount in Stage 2.

If you disagree, war is triggered in this stage.

How would you respond to Country A?

< AGREE – Go to Stage 2 >

< DISAGREE – Go to War >

THIS STAGE – Both countries are equally powerful.
If war is triggered, you get **3 points** and Country A gets **3 points** after deducting the cost of war

NEXT STAGE – Country A becomes more powerful.

If war is triggered, you get **1 point** and Country A gets **5 points** after deducting the cost of war.

However, **if war is triggered, Country C can choose to attack you.**

If Country C attacks you: You get **0 points**. Country C gets **3 points** (in stage 1) or **1 point** (in stage 2). Country A is unaffected.

If Country C does not attack you: You and Country A are unaffected. Country C gets **0 points**.

[For Country C (decision)]

Please choose whether or not to attack Country B if Countries A and B fight each other.

< **Attack** >
< **Do not attack** >

If you **attack**
You get **3 points** (in Stage 1) or **1 point** (in Stage 2)

If you **do not attack**,
You get **0 points**

[STAGE 2]

[Country A (decision)]

You are Country A.

In Stage 1, you suggested the following:

“A gets [], B gets []”

You can change or confirm the amount by choosing ONE of the possibilities below:

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

THIS STAGE – You are more powerful.

If war is triggered, you get **5 points** and Country B gets **1 point** after deducting the costs of war

However, **if war is triggered, Country C can choose to attack Country B.**

If Country C attacks B: Country B gets **0 points**. Country C gets **1 point** (in stage 2). You are unaffected.

If Country C does not attack B: You and Country B are unaffected. Country C gets **0 points**.

[For Country B (decision)]

Country A has confirmed the following:

“A gets [], B gets []”

If you agree, the bargain is made.

If you disagree, war is triggered.

< AGREE – Bargain is Made >

< DISAGREE – Go to War >

THIS STAGE – Country A is more powerful.

If war is triggered, you get **1 point** and Country A gets **5 points** after deducting the cost of war.

However, **if war is triggered, Country C can choose to attack you.**

If Country C attacks you: You get **0 points**. Country C gets **1 point** (in stage 2). Country A is unaffected.

If Country C does not attack you: You and Country A are unaffected. Country C gets **0 points**.

**** Outcome:**

[If Country B agreed]

Country B agreed.

Country A gets the points it demanded.

Country B gets what is left of the Prize.

[If Country B disagreed and C did not attack]

Country B disagreed. War is triggered in Stage [1/2].

Country A gets [3/5] points.

Country B gets [3/1] points.

Country C did not attack.

[If Country B disagreed and C attacked]

Country B disagreed. War is triggered in Stage [1/2].

Country C enters the war and attacks Country B. Country C gets [3/1] points.

Country A gets [3/5] points.

Country B gets 0 points.

[For Player C if C did not attack]

You chose not to attack. You get 0 points.

Repeat Sections with ** in Rounds 7-8.
--

[Players play different games in Rounds 9-14]

Rounds 9-11: Endogenous-Enforcement Game

Transition [After Round 8]:

Scenario 2 is over.

Scenario 3 will start in a few seconds.

Transition [Start of Round 9]:

In Scenario 3:

The relative power between Countries A and B in STAGE 2 depends on
whether Country A cuts its military investment in STAGE 1

There are only 2 countries (A and B) in the game.

Power Shift:

In Stage 1, Country A and Country B are **EQUALLY POWERFUL**.

If they fight a war, each country can seize 50% of the prize (i.e. 5 out of 10 points):

- A ends up with **3 points** (5 points minus 2 points in cost of war)
- B ends up with **3 points** (5 points minus 2 points in cost of war)

In Stage 2, the relative power between Countries A and B depends on whether Country A cuts its military investment in Stage 1.

If A chooses to **CUT** its military investment, A and B remain **EQUALLY POWERFUL** in Stage 2:

If they fight a war in Stage 2, each country can seize 50% of the prize:

- A ends up with **3 points** (5 points minus 2 points)
- B ends up with **3 points** (5 points minus 2 points)

If A chooses to **NOT CUT** its military investment, A becomes **MORE POWERFUL** than B in Stage 2:

If they fight a war in Stage 2, A can seize 70% of the prize:

- A ends up with **5 points** (7 points minus 2 points)
- B ends up with **1 point** (3 points minus 2 points)

Stages:

The game has 2 stages. Here is how the game works:

STAGE 1:

Country A suggests how much of the prize it wants for itself.

- The Prize is worth 10 points. Country A can demand 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 points of the Prize.
- Country A's suggestion in Stage 1 is not implemented until it is confirmed in Stage 2.

Country A decides whether or not to cut its military investment. A's decision is observed by Country B.

Country B agrees or disagrees

- If Country B agrees, you PROCEED TO STAGE 2.
- If Country B disagrees, war is triggered. THE GAME ENDS HERE.

STAGE 2:

Country A confirms the number of points it wants.

- Country A CAN CHANGE its demand in Stage 2.
- Hence, the amount which Country A suggested in Stage 1 may or may not be the same as the amount confirmed in Stage 2.

Country B agrees or disagrees

PAYOFFS:

If Country B agrees in both Stages 1 and 2,

- Country A gets the number of points it confirmed in Stage 2.
- Country B gets the remainder of the Prize.
- Both countries do not lose points in the costs of war.

If Country B disagrees (war) in Stage 1,

- A gets **3 points**, B gets **3 points**

If Country B disagrees (war) in Stage 2,

- A gets **3 points** and B gets **3 points** if A chose to **CUT** its military investment in Stage 1.
- A gets **5 points** and B gets **1 point** if A chose to **NOT CUT** its military investment in Stage 1.

Questions:

To ensure that you have read the instructions carefully, here are a few questions:

In Stage 1, B observes both A's suggested offer and whether A cuts its military investment.

- True
- False

If Country A does NOT CUT its military investment in Stage 1, it becomes MORE POWERFUL than Country B in Stage 2. B's war payoff falls from 3 points (in Stage 1) to 1 point (in Stage 2).

- True
- False

If Country A CUTS its military investment in Stage 1, there is NO CHANGE IN RELATIVE POWER in Stage 2. B's war payoff remains unchanged (3 points) in both Stage 1 and Stage 2.

- True
- False

[Answers are revealed]

****** Scenario:***

You are currently in Scenario 3

****** Player Assignment:***

In this round, you are randomly assigned as a leader in Country [A/B]

Prelude Summary:

Here is a summary of the game. You will make your decision in the next screen.

STAGE 1:

Country A suggests how much of the prize it wants for itself.

Country A decides whether or not to cut its military investment. A's decision is observed by B.

Country B agrees (go to Stage 2) or disagrees (war is triggered; the game ends in Stage 1).

STAGE 2:

Country A confirms how much of the prize it wants for itself.

Country B agrees (a bargain is made) or disagrees (war is triggered).

PAYOFFS:

If a bargain is made,

- A gets what it confirmed in Stage 2.
- B gets the remainder of the Prize.
- Both countries avoid the costs of war.

If war is triggered in **Stage 1**,

- A gets **3 points**, B gets **3 points**. The game ends in Stage 1.

If war is triggered in **Stage 2**,

- A gets **3 points** and B gets **3 points** if A chose to **CUT** its military investment in Stage 1.
- A gets **5 points** and B gets **1 point** if A chose to **NOT CUT** its military investment in Stage 1.

Please take a moment to think about what you will do

***** Actual Play:**

[STAGE 1]

[Country A (decision)]

You are Country A in Stage 1.

Suggest how much of the prize you want by choosing ONE of the possibilities below.

A gets 10, B gets 0
A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3

A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

You CAN CHANGE the amount in Stage 2.

THIS STAGE – Both countries are equally powerful.

If war is triggered, you get **3 points** and Country B gets **3 points** after deducting the costs of war

NEXT STAGE – Your relative power depends on whether you cut military investment in Stage 1.

If war is triggered:

You get **3 points** and Country B gets **3 points** if you chose to **CUT** military investment
 You get **5 points** and Country B gets **1 point** if you chose to **NOT CUT** military investment

You will choose your military investment in the next screen.

[Country A (military investment)]

You are Country A in Stage 1.

Your choice of military investment in Stage 1 affects your war payoff in Stage 2, but **only if you reach Stage 2 and war is triggered in Stage 2. In this case:**

You get **3 points** and Country B gets **3 points** if you chose to **CUT** military investment
 You get **5 points** and Country B gets **1 point** if you chose to **NOT CUT** military investment

Please make your decision:

< **CUT military investment** >
 < **Do NOT CUT military investment** >

After observing your decisions, Country B can either agree or disagree.
 If B agrees, you go to Stage 2. If B disagrees, war is triggered in Stage 1 and the game ends in Stage 1.

[Country B (waiting)]

Please wait while Country A makes a decision.

[For Country B (decision)]

This is Stage 1.

Country A has suggested the following:

“A gets [], B gets []”

Country A chose to [CUT / NOT CUT] its military investment

If you agree, you proceed to Stage 2 in which Country A will confirm its demand.

Country A CAN CHANGE the amount in Stage 2.

If you disagree, war is triggered in this stage.

How would you respond to Country A?

< AGREE – Go to Stage 2 >

< DISAGREE – Go to War >

THIS STAGE – Both countries are equally powerful.

If war is triggered, you get **3 points** and Country A gets **3 points** after deducting the costs of war

NEXT STAGE – Your relative power depends on whether A cut military investment in Stage 1.

If war is triggered: You get **3 points** and Country A gets **3 points** if A chose to **CUT** its military investment. You get **1 point** and Country A gets **5 points** if A chose to **NOT CUT** its military investment.

[STAGE 2]

[Country A (decision)]

You are Country A in Stage 2.

In Stage 1, you suggested the following:

“A gets [], B gets []”

You can change or confirm the amount by choosing ONE of the possibilities below:

A gets 10, B gets 0

A gets 9, B gets 1
A gets 8, B gets 2
A gets 7, B gets 3
A gets 6, B gets 4
A gets 5, B gets 5
A gets 4, B gets 6
A gets 3, B gets 7
A gets 2, B gets 8
A gets 1, B gets 9
A gets 0, B gets 10

[As you had CUT military investment, both A and B remain equally powerful. / As you had CUT military investment, you have become powerful than B.]

If war is triggered: You get **[3 points /5 points]** and B gets **[3 points / 1 point]**

[For Country B (decision)]

Country A has confirmed the following:

“A gets [], B gets []”

Country A chose to **CUT its military investment**

If you agree, the bargain is made.

If you disagree, war is triggered.

How would you respond to Country A?

< AGREE – Bargain is Made >

< DISAGREE – Go to War >

THIS STAGE – If war is triggered:

You get **3 points** and Country A gets **3 points** if A chose to **CUT** its military investment.

You get **1 point** and Country A gets **5 points** if A chose to **NOT CUT** its military investment.

***** Outcome:**

[If Country B agreed]

Country B agreed.

Country A gets the points it demanded.

Country B gets what is left of the Prize.

[If Country B disagreed]

Country B disagreed. War is triggered in Stage [1/2].

Country A gets [3/5] points.

Country B gets [3/1] points.

Repeat Sections with *** in Rounds 10-11.

Rounds 12-14: Sunk-Cost Signaling Game

Transition [After Round 11]:

Scenario 3 is over.

Scenario 4 will start in a few seconds.

Transition [Start of Round 12]:

You are in **SCENARIO 4**

You will be randomly assigned as either **Player A** or **Player B**.

In this game, both players are interested in a valuable prize.

Player A has staked its claim on the Prize.

The game is as follows:

- **Player A chooses one of two threats** to send to Player B, warning B to stay out of a conflict over A's claim.
- **Player B observes the threat and decides** whether to stay out or to confront A's claim.

Types:

The Computer will randomly assign **Player A** as one of two types:

TRUE type (50% chance)

- The **TRUE** type will **ALWAYS** fulfill its threat to fight Player B if B confronts its claim.

FAKE type (50% chance)

- The **FAKE** type will **NEVER** fulfill its threat to fight Player B if B confronts its claim.

Note that **Player A knows** its own type, but **Player B does NOT know** if it has encountered a **TRUE** or **FAKE** Player A.

Knowledge:

The following facts are known to both players:

- The prize is worth much more to TRUE Player A than to FAKE Player A.
- If Player A is TRUE, it will ALWAYS fulfill its threat to fight Player B.

In this case, B gets a **low payoff** if B confronts A.

- If Player A is FAKE, it will NEVER fulfill its threat to fight Player B.

In this case, B gets a **high payoff** if B confronts A.

Sequence:

All players know that the game is as follows:

1) Player A chooses one of two threats to send to Player B:

- **Threat X**, which costs **0 points** for Player A to send
- **Threat Y**, which costs **2 points** for Player A to send

All threats have the **same content** (“A will fight B if B confronts A”) but have **different costs** to A.

The threat (and its cost) A chooses will be known to B.

Player A **knows** whether it is TRUE or FAKE, but Player B **does not know** if it has met a TRUE or FAKE Player A.

2) Player B decides whether to stay out or to confront A's claim:

- If B **stays out**, B gets **6 points** regardless of Player A's type.
- If B **confronts**, B gets
 - **10 points** if Player A is FAKE
 - **2 points** if Player A is TRUE

Note that B gets a high payoff if it confronts a FAKE Player A but a low payoff if it confronts a TRUE Player A.

Player A's payoff depends on its TYPE, and on whether B confronts A:

- If A is **TRUE**, A gets

- **10 points minus the cost of threat** if B stays out
- **4 points minus the cost of threat** if B confronts
- If A is **FAKE**, A gets
 - **3 points minus the cost of threat** if B stays out
 - **2 points minus the cost of threat** if B confronts

Note that the Prize is worth much more to a TRUE Player A than to a FAKE Player A.

Questions:

To ensure that you have read the instructions carefully, here are a few questions:

The prize is worth much more to TRUE Player A than to FAKE Player A.

- True
- False

TRUE Player A will ALWAYS fulfill its threat to fight Player B if B confronts, while FAKE Player A will NEVER fulfill its threat.

- True
- False

If Player A is FAKE, it gets either 3 points minus the cost of threat if B stays out, or 2 points minus the cost of threat if B confronts. What is the most that a FAKE Player A can get if it sends Threat Y (cost = 2 points)?

- 1 point
- 2 points
- 3 points

If Player B confronts Player A ...

- B gets 10 points if Player A is FAKE, but 2 points if Player A is TRUE
- B gets 2 points if Player A is FAKE, but 10 points if Player A is TRUE
- B gets 10 points if Player A is FAKE and 10 points if Player A is TRUE

[Answers are revealed]

****** Scenario:**

You are currently in Scenario 4

****** Player and Resolve Assignment:**

You are randomly assigned as [Player B / Player A of the TRUE Type / Player A of the FAKE Type] in this round.

Prelude:

Here is a summary of the game. You will make your decision in the next screen.

1) Player A chooses one of two threats to send to Player B:

- **Threat X**, which costs **0 points** for Player A to send
- **Threat Y**, which costs **2 points** for Player A to send

All threats carry the **same content** (“A will fight B if B confronts A”) but have **different costs** to Player A. The threat (and its cost) A chooses will be known to B.

Player A **knows** whether it is TRUE or FAKE, but Player B **does not know** if it has met a TRUE or FAKE Player A.

2) Player B decides whether to stay out or to confront:

- If B **stays out**, B gets **6 points** regardless of Player A’s type.
- If B **confronts**, B gets
 - **10 points** if Player A is FAKE
 - **2 points** if Player A is TRUE

Player A’s payoff depends on its TYPE, and on whether B confronts A:

- If A is **TRUE**, A gets
 - **10 points minus the cost of threat**, if B stays out
 - **4 points minus the cost of threat**, if B confronts
- If A is **FAKE**, A gets
 - **3 points minus the cost of threat**, if B stays out
 - **2 points minus the cost of threat**, if B confronts

Please take a moment to think about what you will do

****** Decision [Player A]:**

You are **Player A [(TRUE type) / (FAKE type)]**

Choose which threat to send to Player B:

< **Threat X (cost = 0 points)** >

< **Threat Y (cost = 2 points)** >

All threats carry the **same content** but have **different costs**.
The threat (and its cost) you choose will be known to Player B.

If you are a **TRUE** type, you get
10 points minus the cost of threat, if B stays out
4 points minus the cost of threat, if B confronts

If you are a **FAKE** type, you get
3 points minus the cost of threat, if B stays out
2 points minus the cost of threat, if B confronts

If B **stays out**, B gets **6 points** regardless of your type.
If B **confronts**, B gets **2 points** if you are **TRUE** and **10 points** if you are **FAKE**.

****** Decision [Player B]:**

You are **Player B**

Player A has sent the following threat:

Threat [X / Y], which costs Player A [**0 / 2**] **points**.

Player A threatens to fight you if you confront its claim.

Do you choose to stay out or to confront?

< **Stay Out** >

< **Confront** >

If you **stay out**, you get **6 points** regardless of Player A's type.
If you **confront**, you get **2 points** if Player A is **TRUE** and **10 points** if Player A is **FAKE**.

If A is a **TRUE** type, A gets:
10 points minus the cost of threat if you stay out
4 points minus the cost of threat if you confront

If A is a FAKE type, A gets:
3 points minus the cost of threat if you stay out
2 points minus the cost of threat if you confront

****** Assessment [Player B]:**

Do you think that Player A is a TRUE type?

- < **Very Likely** >
- < **Somewhat Likely** >
- < **More Likely than Unlikely** >
- < **Neither Likely Nor Unlikely** >
- < **More Unlikely than Likely** >
- < **Somewhat Unlikely** >
- < **Very Unlikely** >

Repeat Sections with **** in Rounds 13-14.

Round 15: Risk-Aversion Game

Transition:

Scenario 4 is over. Scenario 5 is the final scenario. It has only 1 round.

Scenario:

You are currently in Scenario 5

Payoffs:

In this scenario, you must make a series of decisions between two options.

OPTION A will pay an amount of points with certainty.

OPTION B will pay 0 point with 50% probability and 10 points with 50% probability.

The computer will randomly select one of the decision problems and count it as your points for this round.

Decision:

OPTION A	OPTION B	Your Choice
2 points	0 with 50% Chance, 10 points with 50% Chance	A <input type="checkbox"/> <input type="checkbox"/> B
3 points	0 with 50% Chance, 10 points with 50% Chance	A <input type="checkbox"/> <input type="checkbox"/> B
4 points	0 with 50% Chance, 10 points with 50% Chance	A <input type="checkbox"/> <input type="checkbox"/> B
5 points	0 with 50% Chance, 10 points with 50% Chance	A <input type="checkbox"/> <input type="checkbox"/> B
6 points	0 with 50% Chance, 10 points with 50% Chance	A <input type="checkbox"/> <input type="checkbox"/> B
7 points	0 with 50% Chance, 10 points with 50% Chance	A <input type="checkbox"/> <input type="checkbox"/> B
8 points	0 with 50% Chance, 10 points with 50% Chance	A <input type="checkbox"/> <input type="checkbox"/> B

[Respondents answer five background questions on age, gender, citizenship, student status and course of study]

Results:

The experiment has ended.

The computer will now calculate your total earnings for the experiment.
It will randomly choose 9 rounds out of the 15 rounds you played.

Then it will sum up your total point earnings in the 9 chosen rounds.

You earn \$0.50 per point.

Earnings:

Round	Profit
[]	[] points
<i>[Repeat first row for 15 rows]</i>	

Your scores in each round are shown here. The computer has randomly chosen the following 9 rounds for payment:

Rounds: [...]

Your total score from the 9 randomly-chosen rounds is [] points, which is [] in cash.

Bye:

Thank you for participating in this experiment!

Please allow us 5 minutes to collate the earnings data and prepare the cash payment for you.

Meanwhile, please take the time to fill in a short questionnaire.

APPENDIX A3

Experimental Instructions for Chapter 4 (Experiment 3)¹

Procedure:

You will be playing 3 decision-making games.

Each game is different and independent.

The computer will randomly choose 1 out of the 3 games you played, and count your point earnings in that game as bonus payment. Each game has a total possible value of 10 points.

We will pay you

\$0.10 PER POINT

You will be paid your point earnings as a bonus through Mechanical Turk.

HOW MUCH BONUS YOU EARN DEPENDS ON YOUR PERFORMANCE.

The more points you win, the more money you receive.

Matching:

The game outcome depends on your decision **and** your opponent's decision.

To generate the outcome, **your decision will be randomly matched with the decision of another player** in the opponent role.

Your opponent will be **randomly drawn** from participants who have already played the same game in the opponent role on Mechanical Turk.

Transition:

This is Game 1.

Role:

¹ The instructions are largely similar for Experiment 4 except for the differences highlighted in Section 3.3 in Chapter 4. References or screens specific to the Internet-based version (such as "Mechanical Turk" and the "Matching", "Pause" and "Payment" screens) are removed or substituted in the laboratory version.

You will be randomly assigned as either **Player A** or **Player B**.

In this game, both players are interested in a valuable prize.

Player A has staked its claim on the Prize.

The game is as follows:

- **Player A chooses one of two threats** to send to Player B, warning B to stay out of a conflict over A's claim.
- **Player B observes the threat and decides** whether to stay out or to confront A's claim.

Types:

The Computer will randomly assign **Player A** as one of two types:

TRUE type (50% chance)

- The TRUE type will **ALWAYS** fulfill its threat to fight Player B if B confronts its claim.

FAKE type (50% chance)

- The FAKE type will **NEVER** fulfill its threat to fight Player B if B confronts its claim.

Note that **Player A knows** its own type, but **Player B does NOT know** if it has encountered a TRUE or FAKE Player A.

Knowledge:

The following facts are known to both players:

- The prize is worth much more to TRUE Player A than to FAKE Player A.
- If Player A is TRUE, it will **always** fulfill its threat to fight B.

In this case, B gets a **low payoff** if B confronts A.

- If Player A is FAKE, it will **never** fulfill its threat to fight B.

In this case, B gets a **high payoff** if B confronts A.

Pause:

The next screen provides instructions on how the game works. **Please take the time to read the game instructions carefully.**

For game instruction screens, **there will be a brief pause** of about 30 seconds so that you can read carefully. At the end of the pause, an arrow will appear at the bottom of the screen. Once the arrow appears, you may move forward by clicking on the arrow.

Sequence:

All players know that the game is as follows:

1) Player A chooses one of two threats to send to Player B:

- **Threat X**, which costs **0 points** for Player A to send
- **Threat Y**, which costs **2 points** for Player A to send

All threats have the **same content** (“A will fight B if B confronts A”) but have **different costs** to A.

The threat (and its cost) A chooses will be known to B.

Player A **knows** whether it is TRUE or FAKE, but Player B **does not know** if it has met a TRUE or FAKE Player A.

2) Player B decides whether to stay out or to confront A's claim:

- If B **stays out**, B gets **6 points** regardless of Player A's type.
- If B **confronts**, B gets
 - **10 points** if Player A is FAKE
 - **2 points** if Player A is TRUE

Note that B gets a high payoff if it confronts a FAKE Player A but a low payoff if it confronts a TRUE Player A.

Player A's payoff depends on its TYPE, and on whether B confronts A:

- If A is **TRUE**, A gets

- **10 points minus the cost of threat**, if B stays out
- **4 points minus the cost of threat**, if B confronts
- If A is **FAKE**, A gets
 - **3 points minus the cost of threat**, if B stays out
 - **2 points minus the cost of threat**, if B confronts

Note that the Prize is worth much more to a TRUE Player A than to a FAKE Player A.

Questions:

To ensure that you have read the instructions carefully, here are a few questions:

The prize is **worth much more** to TRUE Player A than to FAKE Player A.
[True / False] ²

TRUE Player A will **ALWAYS** fulfill its threat to fight Player B if B confronts, while FAKE Player A will **NEVER** fulfill its threat.
[True / False]

If Player A is **FAKE**, it gets either 3 points minus the cost of threat if B stays out, or 2 points minus the cost of threat if B confronts. What is the most that a **FAKE Player A** can get if it sends **Threat Y (cost = 2 points)**?
[1 point / 2 points / 3 points] ³

If Player B **confronts** Player A ...
[B gets 10 points if Player A is FAKE, but 2 points if Player A is TRUE / B gets 2 points if Player A is FAKE, but 10 points if Player A is TRUE / B gets 10 points if Player A is FAKE and 10 points if Player A is TRUE]

[Answers are revealed]

Player and Resolve Assignment:

² For the randomized group of receivers without type-separation (see Section III in paper), this question is replaced by: "Player B **does not know** what are the payoffs for TRUE Player A or FAKE Player A. But B **knows** that both TRUE and FAKE Player A will get a higher payoff if B stays out than if B confronts. [True / False]" The modified question appears as the second question.

³ Omitted for the randomized group of receivers without type-separation.

You are randomly assigned as [**Player B / Player A of the TRUE Type / Player A of the FAKE Type**] in this game.

Prelude:

Here is a summary of the game. You will make your decision in the next screen.

1) Player A chooses one of two threats to send to Player B:

- **Threat X**, which costs **0 points** for Player A to send
- **Threat Y**, which costs **2 points** for Player A to send

All threats carry the **same content** (“A will fight B if B confronts A”) but have **different costs** to Player A. The threat (and its cost) A chooses will be known to B.

Player A **knows** whether it is TRUE or FAKE, but Player B **does not know** if it has met a TRUE or FAKE Player A.

2) Player B decides whether to stay out or to confront:

- If B **stays out**, B gets **6 points** regardless of Player A’s type.
- If B **confronts**, B gets
 - **10 points** if Player A is FAKE
 - **2 points** if Player A is TRUE

Player A’s payoff depends on its TYPE, and on whether B confronts A: ⁴

- If A is **TRUE**, A gets
 - **10 points minus the cost of threat**, if B stays out
 - **4 points minus the cost of threat**, if B confronts
- If A is **FAKE**, A gets
 - **3 points minus the cost of threat**, if B stays out
 - **2 points minus the cost of threat**, if B confronts

Please take a moment to think about what you will do

⁴ For the randomized group of receivers without type-separation: “Player B **does not know** what are the payoffs for TRUE Player A or FAKE Player A. However, Player B **knows** that: Regardless of whether it is TRUE or FAKE, Player A gets a higher payoff if B stays out than if B confronts.”

*** Decision [Player A]:**

You are Player A [(TRUE type) / (FAKE type)]

Choose which threat to send to Player B:

< **Threat x** (cost = **0 points**) >

< **Threat z** (cost = **2 points**) >

All threats carry the **same content** but have **different costs**. The threat (and its cost) you choose will be known to Player B.

If you are a **TRUE** type, you get

10 points minus the cost of threat, if B stays out

4 points minus the cost of threat, if B confronts

If you are a **FAKE** type, you get

3 points minus the cost of threat, if B stays out

2 points minus the cost of threat, if B confronts

If B **stays out**, B gets **6 points** regardless of your type. If B **confronts**, B gets **2 points** if you are **TRUE** and **10 points** if you are **FAKE**.

*** Open-Response [Player A]:**

Please tell us briefly why you sent [**Threat X (with cost of 0 points / Threat Y (with cost of 2 points))**].

*** Decision [Player B]:**

You are Player B in Game 1

Player A has sent the following threat:

Threat [X / Y], which costs Player A [**0 / 2**] **points**.

Player A threatens to fight you if you confront its claim.

Do you choose to stay out or to confront?

< Stay Out >

< Confront >

- If you **stay out**, you get **6 points** regardless of Player A's type.
- If you **confront**, you get **2 points** if Player A is **TRUE** and **10 points** if Player A is **FAKE**.

If A is a TRUE type, A gets:⁵

10 points minus the cost of threat, if you stay out
4 points minus the cost of threat, if you confront

If A is a FAKE type, A gets:

3 points minus the cost of threat, if you stay out
2 points minus the cost of threat, if you confront

*** Assessment [Player B]:**

Do you think that Player A is a TRUE type?

< Likely / Unlikely / Neither Likely Nor Unlikely >

If likely: How likely do you think that Player A is a TRUE type?
 [Very Likely / Somewhat Likely]

If unlikely: How unlikely do you think that Player A is a TRUE type?
 [Very Unlikely / Somewhat Unlikely]

If neither: Do you lean toward thinking that Player A is a TRUE type or a FAKE type, or don't you lean either way?
 [Lean toward thinking that Player A is a TRUE type / Lean toward thinking that Player A is a FAKE type / Lean neither way]

*** Open-Response [Player B]:**

Please tell us briefly why you [believe / do NOT believe] Player A is a TRUE type.

Outcome:

Game 1 has ended.

⁵ For the randomized group of receivers without type-separation: "You do not know Player A's payoffs."

The Computer will generate the outcome by matching your decision with the decision of a randomly selected opponent who played the game on Mechanical Turk.

[Game 2]

Transition to Game 3:

This is the final game: Game 3.

In this game, you face a series of decisions between two options.

Payoffs:

You must make a series of decisions between two options.

OPTION A will pay an amount of points with certainty.

OPTION B will pay 0 point with 50% probability and 10 points with 50% probability.

The computer will randomly select one of the decision problems and count it as your points for this round.

*** Decision:**

OPTION A: 2 points with certainty
OPTION B: 0 points with 50% Chance, **10 points** with 50% Chance [A/B]

OPTION A: 3 points with certainty
OPTION B: 0 points with 50% Chance, **10 points** with 50% Chance [A/B]

OPTION A: 4 points with certainty
OPTION B: 0 points with 50% Chance, **10 points** with 50% Chance [A/B]

OPTION A: 5 points with certainty
OPTION B: 0 points with 50% Chance, **10 points** with 50% Chance [A/B]

OPTION A: 6 points with certainty
OPTION B: 0 points with 50% Chance, **10 points** with 50% Chance [A/B]

OPTION A: 7 points with certainty

OPTION B: 0 points with 50% Chance, 10 points with 50% Chance [A/B]

OPTION A: 8 points with certainty

OPTION B: 0 points with 50% Chance, 10 points with 50% Chance [A/B]

Outcome:

Decision [] was randomly selected from the list of decisions you just made. In that decision, you had the choice between [] points and the risky gamble of 10 points with 50% chance and 0 points with 50% chance.

You chose the [first option / risky gamble]. Your score for this round is [] points.

[Demographic questions]

Pay Intro:

Now, the computer will randomly choose 1 out of the 3 games you played, and count your point earnings in that game for bonus payment.

We will pay you

\$0.10 PER POINT

Payment:

The computer has randomly chosen the following game for bonus payment:

Game: []

- [] points = Your score in the selected game = **\$[] in cash**

You will be paid your point earnings as a bonus through Mechanical Turk after all participants have completed the HIT

Bye:

Thank you for your participation! Do you have any comments for us?
[Open Response]

APPENDIX A4

Experimental Instructions for Chapter 4 (Experiment 5, 6, 7)

[For all participants]

You will read about a type of situation in international relations that had occurred in the past and might occur again in the future. In the next screen, we will describe the situation and ask for your opinion.

[For all participants]

Two countries – Country X and Country Z – have staked their claims on an important piece of territory. Both countries believe that the territory is rightfully theirs. The following facts are known:

[Participants are randomly assigned to ONE of the treatments below.]

[Experiment 5]:

- Z has announced that it will move into the territory.
- X has responded by sending a message to Z. In the message, X has threatened to fight a war with Z if Z moves into the territory.
- **[Treatment 5A/5B]** At the same time, X has mobilized its military. It is clear that military mobilization at this time is [very costly / not very costly] to X.
- * **[For those in Treatments 5C/5D/5E/5F]** At the same time, X has mobilized its military. It is clear that military mobilization at this time is [very costly / not very costly] to X.
- * **[For those in Treatments 5C/5D/5E/5F]** At the same time, it is clear that fighting a war at this time will [be / not be] very costly to X.

If Z moves into the territory, do you think that it is likely or unlikely that X will fulfill its threat?

[* These two sentences are presented in random order]

[Assignment Matrix for Treatments 5C/5D/5E/5F]

		<i>Cost of Making Threat (Sunken Cost)</i>	
		High Cost	Low Cost
<i>Cost of Fulfilling Threat</i>	High Cost	Treatment 5C	Treatment 5E
	Low Cost	Treatment 5D	Treatment 5F

[Experiment 6]:

- Z has announced that it will move into the territory.
- X has responded by sending a message to Z. In the message, X has threatened to fight a war with Z if Z moves into the territory.
- **[Treatment 6A/6B]** At the same time, it is clear that fighting a war at this time will [be / not be] very costly to X.

If Z moves into the territory, do you think that it is likely or unlikely that X will fulfill its threat?

[Experiment 7]:

- Z has announced that it will move into the territory.
- X has responded by sending a message to Z. In the message, X has threatened to fight a war with Z if Z moves into the territory.
- **[Treatment 7A/7B]** At the same time, X has publicly announced its threat of war. It is clear that it will [be very costly / not be very costly] to X if X reneges on its threat.

If Z moves into the territory, do you think that it is likely or unlikely that X will fulfill its threat?

[For all participants]

X has threatened to fight a war with Z if Z moves into the territory.

If Z moves into the territory, do you think that it is likely or unlikely that X will fulfill its threat? [Likely / Unlikely / Neither likely nor unlikely]

[If likely]: Do you think that it is very likely, or only somewhat likely, that X will fulfill its threat?

[If unlikely]: Do you think that it is very unlikely, or only somewhat unlikely, that X will fulfill its threat?

[If neither]: Do you lean toward believing that X will fulfill its threat, lean toward disbelieving, or don't you lean either way?

[All]: Can you tell us why you believe or disbelieve Country X's threat?

APPENDIX A5

Experimental Instructions for Chapter 4 (Experiment 8, 9, 10)

[For all participants]

The following questions are about international relations.

There will be a slight delay before you can move to the next screen. At the end of the pause, an arrow will appear at the bottom of the screen. Once the arrow appears, you may move to the next screen by clicking on the arrow.

[For all participants]

You will read about a type of situation in international relations that had occurred in the past and might occur again in the future.

In the next screen, we will describe the situation and ask for your opinion.

[For all participants]

Two countries – Country X and Country Y – have staked their claims on an important piece of territory. Both countries believe that the territory is rightfully theirs. The following facts are known:

- X had sent its troops into the territory.

[For Treatment 1A/1B/1C] *

(Experiments 9-10)

- Subsequently, X had continued to maintain its troops on the territory.
- Later, X had included the defense of the territory as one of its “core security interests” in a public announcement of its national defense strategy.
- **[For Treatment 1B/1C, Add:]** It is known that fighting a war at this time will [be very / not be very] costly to X.

* The first two bullet points are presented in random order.

[For Treatment 2A/2B/2C] *

(Experiments 9-10)

- Subsequently, X had withdrawn its troops from the territory.
- Later, X did not include the defense of the territory as one of its “core security interests” in a public announcement of its national defense strategy.
- **[For Treatment 2B/2C, Add:]** It is known that fighting a war at this time will [be very / not be very] costly to X.

[For Treatment 3A/3B] *

(Experiment 9)

- Subsequently, X had [withdrawn / continued to maintain] its troops [from / on] the territory.
- Later, X [had included / did not include] the defense of the territory as one of its “core security interests” in a public announcement of its national defense strategy.

[Response Screens For Treatments 1-3]

Now, X has threatened to fight a war with Y if Y’s troops enter the territory.

If Y enters the territory, do you think that it is likely or unlikely that X will fulfill its threat? [Likely / Unlikely / Neither likely nor unlikely]

[If likely]: Do you think that it is very likely, or only somewhat likely, that X will fulfill its threat?

[If unlikely]: Do you think that it is very unlikely, or only somewhat unlikely, that X will fulfill its threat?

[If neither]: Do you lean toward believing that X will fulfill its threat, lean toward disbelieving, or don't you lean either way?

[All]: Can you tell us why you believe or disbelieve Country X’s threat?

[For Treatment 4A/4B]

(Experiment 8)

- In a previous incident, X had also put its troops on the territory. In response, Y had sent its troops into the territory. [Y / X] had a strong military, so it [did not take / took] a major effort for Y to push X out from the territory previously.

[For Treatment 5A/5B/5C]

(Experiment 8)

- Y lodged a diplomatic protest against this intrusion.
- Later, Y warned X that “he who plays with fire will eventually be consumed by fire.”
- Subsequently, Y lodged “the strongest and most serious protest” against the intrusion and warned X that if it continued “to spread the flames of war, it must bear full responsibility for the resulting casualties on both sides and all other consequences that may ensue.”
- **[For Treatment 5B/5C, Add:]** It is known that fighting a war at this time will [be very / not be very] costly to Y.

[For Treatment 6A/6B/6C]

(Experiment 8)

- Y lodged “the strongest and most serious protest” against the intrusion and warned X that if it continued “to spread the flames of war, it must bear full responsibility for the resulting casualties on both sides and all other consequences that may ensue.”
- Later, Y increased its troops at the border between both countries.
- Subsequently, Y built military installations around the disputed territory.
- **[For Treatment 6B/6C, Add:]** It is known that fighting a war at this time will [be very / not be very] costly to Y

[Response Screens For Treatments 4-6]

Now, Y has threatened to fight a war with X if X does not withdraw from the territory.

If X does not withdraw, do you think that it is likely or unlikely that Y will fulfill its threat? [Likely / Unlikely / Neither likely nor unlikely]

[If likely]: Do you think that it is very likely, or only somewhat likely, that Y will fulfill its threat?

[If unlikely]: Do you think that it is very unlikely, or only somewhat unlikely, that Y will fulfill its threat?

[If neither]: Do you lean toward believing that Y will fulfill its threat, lean toward disbelieving, or don't you lean either way?

[All]: Can you tell us why you believe or disbelieve Country Y's threat?

APPENDIX B1

Table B1. Logit Estimates of Determinants for the Decision for War: Rounds 11-15

	(1)	(2)	(3)	(4)	(5)	(6)
Enforcement	-3.459*** (0.661)	-3.254*** (0.626)	-3.256*** (0.676)	-4.199*** (0.756)	-3.778*** (0.732)	-3.832*** (0.822)
Initial offer		-0.141 (0.266)	-0.142 (0.261)		0.058 (0.249)	0.049 (0.245)
Risk preference			-0.003 (0.230)			-0.114 (0.254)
Constant	0.975 (0.302)**	1.331 (1.313)	1.350 (1.755)	0.663 (0.702)	0.332 (1.470)	0.994 (1.997)
Round and session fixed-effects	No	No	No	Yes	Yes	Yes
Prob>Chi ²	0.000	0.000	0.000	0.000	0.001	0.000
Log-likelihood	-64.016	-58.767	-58.766	-56.110	-52.826	-52.664
Pseudo-R ²	0.366	0.322	0.322	0.445	0.391	0.393
Observations	151	137	137	151	137	137

Notes: *** $p \leq 0.001$; ** $p \leq 0.01$; * $p \leq 0.05$. Player B can decide for war in these rounds without seeing Player A's initial offer. In parentheses are robust standard errors corrected for clustering at the subject level. Round and session dummies are used to control for round and session fixed effects.

APPENDIX B2

Table B2. Logit Estimates of Determinants for War: Rounds 1-5

	(1)	(2)	(3)	(4)	(5)	(6)
Inadvertent enforcement	-2.374*** (0.460)	-2.846*** (0.520)	-2.990*** (0.538)	-2.471*** (0.438)	-3.124*** (0.528)	-3.280*** (0.562)
Initial offer		-0.477*** (0.182)	-0.508*** (0.189)		-0.589*** (0.197)	-0.608*** (0.208)
Risk preference			-0.116 (0.250)			-0.00136 (0.284)
Constant	1.081*** (0.283)	3.525*** (1.031)	4.128*** (1.318)	1.618** (0.641)	4.687*** (1.373)	5.052*** (1.614)
Round and session fixed-effects	No	No	No	Yes	Yes	Yes
Observations	140	140	135	140	140	135
Pseudo-R ²	0.214	0.262	0.280	0.230	0.291	0.313

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. Variables are defined in Section 1.3 in Chapter 4. In parentheses are robust standard errors corrected for clustering at the subject level. Round and session dummies are used to control for round and session fixed effects.

Table B3. Logit Estimates of Determinants for War Decision in Stage 1: Rounds 1-5

	(1)	(2)	(3)	(4)	(5)	(6)
Inadvertent enforcement	-2.808*** (0.634)	-3.200*** (0.688)	-3.309*** (0.673)	-3.045*** (0.592)	-3.884*** (0.778)	-3.956*** (0.751)
Initial offer		-0.404** (0.172)	-0.425** (0.175)		-0.641*** (0.218)	-0.635*** (0.223)
Risk preference			-0.178 (0.227)			-0.0982 (0.305)
Constant	0.693** (0.277)	2.729*** (0.983)	3.425*** (1.209)	0.971 (0.622)	4.236*** (1.425)	4.643*** (1.577)
Round and session fixed-effects	No	No	No	Yes	Yes	Yes
Observations	140	140	135	140	140	135
Pseudo-R ²	0.261	0.296	0.311	0.299	0.359	0.374

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. Variables are defined in Section 1.3 in Chapter 4. In parentheses are robust standard errors corrected for clustering at the subject level.

Table B4. Logit Estimates of Determinants for War in the Control Group

	(1)	(2)	(3)	(4)
Inadvertent enforcement	-1.758*** (0.416)	-1.869*** (0.493)	-1.764*** (0.436)	-1.904*** (0.515)
Initial offer		-0.126 (0.193)		-0.156 (0.205)
Risk preference			0.232 (0.235)	0.235 (0.231)
Constant	1.688*** (0.647)	2.412* (1.335)	1.048 (0.917)	1.936 (1.497)
Session fixed-effects	Yes	Yes	Yes	Yes
Observations	118	118	111	111
Pseudo-R ²	0.131	0.134	0.138	0.143

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. In parentheses are robust standard errors corrected for clustering at the subject level. Variables are defined in Section 1.3 in Chapter 4. The control group played Rounds 1-5 without enforcement before switching to asymmetric inadvertent enforcement in Rounds 6-8.

Table B5. Logit Estimates of Determinants for War in the Treatment Group

	(1)	(2)	(3)	(4)
Symmetric-enforcement dummy	0.233 (0.548)	0.211 (0.617)	0.253 (0.535)	0.212 (0.599)
Initial offer		-0.646*** (0.190)		-0.645*** (0.192)
Risk preference			-0.0889 (0.364)	-0.00738 (0.382)
Constant	-2.141*** (0.594)	0.0399 (0.856)	-1.871 (1.210)	0.0613 (1.405)
Session fixed-effects	Yes	Yes	Yes	Yes
Observations	106	106	106	106
Pseudo-R ²	0.0349	0.140	0.0355	0.140

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. In parentheses are robust standard errors corrected for clustering at the subject level. Variables are defined in Section 1.3 in Chapter 4. The treatment group played Rounds 1-5 with inadvertent enforcement of the symmetric variant before switching to the asymmetric variant in Rounds 6-8.

Table B6. OLS Estimates of Determinants for Final Offer Size: Treatment Group

	(1)	(2)	(3)	(4)	(5)	(6)
Symmetric-enforcement dummy	-1.167*** (0.285)	-1.170*** (0.280)	-1.010*** (0.320)	-1.188*** (0.278)	-1.197*** (0.275)	-1.047*** (0.300)
Risk preference		-0.00894 (0.227)	0.0194 (0.233)		-0.0295 (0.211)	-0.0418 (0.237)
Initial offer			0.273 (0.170)			0.279* (0.156)
Constant	3.190*** (0.202)	3.217*** (0.746)	1.932* (1.097)	3.767*** (0.501)	3.853*** (0.821)	2.564** (1.007)
Session fixed-effects	No	No	No	Yes	Yes	Yes
Observations	102	102	102	102	102	102
R ²	0.156	0.156	0.194	0.206	0.207	0.242

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. Variables are defined in Section 1.3 in Chapter 4. The treatment group played Rounds 1-5 with symmetric inadvertent enforcement before switching to asymmetric inadvertent enforcement in Rounds 6-8. In parentheses are robust standard errors corrected for clustering at the subject level.

Table B7. Logit Estimates of Determinants for War: Experiment 2

	(1)	(2)	(3)	(4)	(5)	(6)
Endogenous enforcement	-2.547*** (0.516)	-2.565*** (0.519)	-2.491*** (0.514)	-3.359*** (0.627)	-3.641*** (0.758)	-3.597*** (0.762)
Initial offer		-0.0891 (0.196)	-0.0693 (0.198)		-0.500* (0.267)	-0.484* (0.270)
Risk preference			-0.165 (0.245)			0.0125 (0.284)
Constant	1.186*** (0.372)	1.620 (1.089)	1.979 (1.401)	1.007 (0.665)	3.438** (1.567)	3.338* (1.709)
Round and session fixed-effects	No	No	No	Yes	Yes	Yes
Observations	96	96	95	96	96	95
Pseudo-R ²	0.242	0.243	0.242	0.325	0.347	0.345

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. Variables are defined in Section 2.3 in Chapter 4. In parentheses are robust standard errors corrected for clustering at the subject level.

Table B8. Logit Estimates of Determinants for War Decision in Stage 1: Experiment 2

	(1)	(2)	(3)	(4)	(5)	(6)
Endogenous enforcement	-2.827*** (0.553)	-2.856*** (0.547)	-2.803*** (0.559)	-3.666*** (0.757)	-3.962*** (0.827)	-3.993*** (0.860)
Initial offer		-0.130 (0.235)	-0.111 (0.236)		-0.479 (0.306)	-0.483 (0.318)
Risk preference			-0.0643 (0.243)			0.138 (0.296)
Constant	0.857** (0.344)	1.492 (1.250)	1.555 (1.462)	0.600 (0.691)	2.932* (1.742)	2.590 (1.805)
Round and session fixed-effects	No	No	No	Yes	Yes	Yes
Observations	96	96	95	96	96	95
Pseudo-R ²	0.278	0.280	0.275	0.356	0.377	0.374

Notes: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. Variables are defined in Section 2.3 in Chapter 4. In parentheses are robust standard errors corrected for clustering at the subject level.