

When Hegemons Arise: Explaining Balance of Power Failure in World History

VOLUME I of II

A Dissertation

Presented to

The Faculty of the Graduate School of Arts and Sciences
Brandeis University

Department of Politics
Robert J. Art, Advisor

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By

Eliane Tschaen Barbieri

February 2011

UMI Number: 3428182

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3428182

Copyright 2011 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Signature Page

The original signature page of this form is on file in the Graduate School of Arts and Science.

This dissertation, directed and approved by **Eliane Tschaen Barbieri**'s Committee, has been accepted and approved by the Faculty of Brandeis University in partial fulfillment of the requirements for the degree of:

DOCTOR OF PHILOSOPHY

Adam B. Jaffe, Dean of Arts and Sciences
Brandeis University

Dissertation Committee:

Robert J. Art, Department of Politics, Brandeis University

Kerry A. Chase, Department of Politics, Brandeis University

Robert S. Ross, Department of Political Science, Boston College

© Eliane Tschaen Barbieri 2010
All rights reserved.

Acknowledgements

I would like to express my sincerest gratitude to all those who have contributed to the completion of this dissertation.

I am particularly indebted to my advisor, Professor Robert J. Art, for his continuous support and invaluable guidance throughout my graduate career. Professor Art's dedication to the field, the profession, and his students has inspired me to work hard and to push my limits every day. He was always available whenever I needed advice, and discussing my work with him invariably helped me gain faith in my ability to complete this thesis and succeed in my graduate studies.

Professors Robert S. Ross and Kerry A. Chase generously offered their time and expertise to help me better my thesis. I am grateful for their insightful feedback and kind encouragement. My heart-felt thanks also go to Professor Jytte Klausen for taking me under her wing and helping me broaden my intellectual horizon.

I would like to convey my appreciation to the faculty and staff of the Department of Politics at Brandeis University, who provided helpful intellectual and technical support. I must acknowledge as well my fellow graduate students in the Department, who gave useful comments on successive iterations of this thesis and whose friendship brightened my graduate school experience.

On a personal level, I am grateful to my parents for encouraging me to continue my studies and being confident in my ability to succeed in my endeavors. Lastly, I would not have completed this dissertation without the unwavering support of my husband, whose infinite patience, kindness, and dedication have been my pillar through every step of the research and writing.

Abstract

When Hegemons Arise: Explaining Balance of Power Failure in World History provides a first theoretical systematization of balance of power failure. It seeks to explain why states sometimes fail to stop hegemons from arising, contrary to the prediction of balance of power theory. While throughout history, most potential hegemons were stopped, large empires of hegemonic stature have occasionally emerged. This thesis offers a conceptual framework to analyze such exceptions. It identifies twenty-one structural and unit-level factors that can cause balance of power to fail and tests them in four in-depth case studies—the Mongol conquests, the rise of Rome, the unification of China under the Qin Dynasty, and the expansion of the United States into the Western Hemisphere—to determine which factors play a primary, secondary, and marginal role.

The thesis concludes that three factors are critical in explaining why rising hegemons occasionally succeed: the occurrence of acute distrust and a short-term gains focus among the potential balancers that prevent cooperative balancing; the rising hegemon's development of innovative ways to extract and use military resources that the potential balancers are unable to emulate or counteract; and the rising hegemon's provision of unique state-building measures that afford it more efficient control over the system than its competitors. The findings highlight that contrary to conventional wisdom, communication hurdles and bandwagoning play only secondary roles in allowing the success of an aspiring hegemon. They also substantiate the fact that the emerging hegemon plays a determinant role in the outcome, even though the balance of power scholarship emphasizes almost exclusively the role of the balancers, and they stress that the unit-level processing of structural constraints contributes significantly to producing systemic results like balance of power or hegemony.

Table of Contents

VOLUME I

Acknowledgements	iv
Abstract	v
List of Tables, Figures, Maps and Illustrations	xi
Chapter [1] Introduction	1
<u>1. Questions Addressed</u>	7
<u>2. The Argument</u>	10
<u>3. The Evidence</u>	13
<u>4. Structure of the Project</u>	14
Chapter [2] A Conceptual Framework for Balance of Power Failure	18
<u>1. The Limits of Balance of Power Theory and Why Failures Matter</u>	22
<i>a. The Workings of Balance of Power</i>	22
1. What is Balance of Power Theory	22
2. Additional Assumptions for Balance of Power	24
What Do States Balance Against?	25
What Types of States Balance?	27
What Are the Characteristics of Balancers?	29
How Do States Balance?	34
<i>b. The Importance of Balance of Power Failure</i>	37
1. There Are No Structural Balances of Power	40
2. Balance of Power is not the Most Common Outcome in International Relations	43
3. Balance of Power Failure as a Discrepancy within the Theory	47
<u>2. Balance of Power Failure Mechanisms</u>	51
<i>a. What is Balance of Power Failure?</i>	51
<i>b. What Explains Balance of Power Failure?</i>	54
1. Communication Problems	59
2. Collective Inaction	62
3. Laggard Balancing	67
4. Bandwagoning	69
5. Rising Hegemon's Military Achievements/Innovation	72
6. Rising Hegemon's Unique Non-Military Achievements/Skills	74
<i>c. Evidence and Case Selection</i>	78
Chapter [3] The Mongol Conquests: The Potential Balancers	83

Timeframe and Boundary of Study	84
Brief Chronology of the Mongol Expansion	87
1. <u>Communication Problems</u>	95
<i>a. Physical Difficulty</i>	95
<i>b. Misperception</i>	102
<i>c. Deliberate Deception</i>	105
1. Spies	105
2. Divide-and-Conquer	108
3. Terror and Intimidation	112
2. <u>Collective Inaction</u>	119
<i>a. Communication Problems</i>	120
<i>b. Lack of Trust</i>	121
1. Mongolia	122
2. China	123
3. Middle East	125
4. Russia and Europe	131
<i>c. Lack of Sufficient Interest and Buckpassing</i>	135
3. <u>Laggard Balancing</u>	136
<i>a. Communication Problems</i>	137
<i>b. Lack of Trust</i>	138
1. Mongolia	139
2. China	139
3. Middle East	140
4. Russia and Europe	141
4. <u>Bandwagoning</u>	146
<i>a. Fear</i>	146
<i>b. Profit</i>	148
1. East Asia	148
2. Southeast Asia	151
3. Middle East	152
4. Russia and Europe	153
Chapter [4] The Mongol Conquests: The Rising Hegemon	158
1. <u>Military Organization</u>	158
<i>a. Advantage of the Steppes: Logistical Innovations and Adaptability to Foreign Theatres</i>	159
<i>b. Learning and Foreign Input</i>	170
1. Technical Experts	171
2. High-Ranking Leaders	172
3. Scholars and Craftsmen	174

<i>c. Decimal Army Organization</i>	176
<i>d. Weaponry</i>	184
<i>e. Strategies and Tactics</i>	190
1. Strategic Planning	190
2. Tactical Innovation and the Influence of Hunting	191
3. Communication System	193
4. Quantitative Improvements	195
5. Key Tactical Innovations	199
6. Innovative Siege Techniques	209
2. <u>Unique Non-Military Skills and Innovations</u>	216
<i>a. Economic Advantage</i>	217
1. Fair Trading Practices	219
2. Uniform Customs and Tax system	219
3. Infrastructure Expansion	220
4. Universal Monetary Instruments	223
5. Central Support of Local Communities	225
6. Foreign Influence	227
<i>b. Political, Administrative, and Social Organization</i>	229
1. Semi-Democratic Central Leadership	229
2. Local Autonomy with Mongol Oversight	232
3. Universal Legal Rules	238
4. Limited Taxation	242
5. Improved Administrative Practices	244
6. Census and Record-Keeping	247
7. Promotion of Cosmopolitanism and Cultural Diversity	248
8. Religious Tolerance	252
9. Egalitarian Social Structure	259
10. Gender Equality	261
Conclusion	265
Chapter [5] The Rise of Rome: The Potential Balancers	271
Timeframe, Boundary of Study, and Chronological Considerations	273
<u>1. Communication Problems</u>	282
<u>2. Collective Inaction</u>	291
<i>a. Communication Problems</i>	291
<i>b. Lack of Trust</i>	293
<u>3. Laggard Balancing</u>	301
<u>4. Bandwagoning</u>	318

VOLUME II

Chapter [6] The Rise of Rome: The Rising Hegemon	330
<u>1. Military Skills and Innovation</u>	331
<i>a. Army Organization</i>	331
1. Tactical Organization: From the Phalanx to the Maniples	332
2. Structural Organization: From Militia to Professional Military	336
3. Ethnic Organization: Legionaries and Auxiliaries	344
4. Sustaining the Organization: Rome's Unparalleled Military Discipline	349
<i>b. Roman Strategies and Tactics</i>	368
1. Superior Strategy and Grand-Strategy	368
2. Superior Tactics	374
<i>c. Proficient Technological Adaptation</i>	379
<u>2. Non-Military Skills and Innovation</u>	386
<i>a. Geography</i>	386
<i>b. Economic Development</i>	389
<i>c. Stable Yet Flexible Political Rule</i>	399
<i>d. Social Inclusiveness</i>	409
Conclusion	415
Chapter [7] The Qin Unification of China	418
Boundary of Inquiry and Timeline	418
<u>1. The Balancers' Side: Collective Action Disasters and Vulnerability to Deception</u>	425
<i>a. Communication Problems</i>	425
1. Physical Communication Problems	425
2. Misperception	428
3. Deliberate Deception	429
<i>b. Collective Inaction</i>	431
<i>c. Laggard Balancing</i>	437
<i>d. Bandwagoning</i>	445
<u>2. The Rising Hegemon's Side: Innovative Military and State Institutions</u>	450
<i>a. Military Innovation</i>	451
1. Military Organization	451
2. Army Discipline	457
3. Military Strategy, Tactics, and Technology	459
<i>b. Non-Military Skills and Innovation</i>	464
1. Geography and Geopolitical Assets	466

2. Economic Advantage	468
3. Political Organization	477
4. Social Control and Organization	486
Conclusion	491
Chapter [8] U.S. Expansion into the Western Hemisphere	493
Boundary of Inquiry	497
Timeline of U.S. Expansion	503
<u>1. The Balancers' Side: Absence of Collective Action and Bandwagoning</u>	513
<i>a. Communication Problems</i>	513
1. Physical Communication Problems	514
2. Misperception	516
3. Deliberate Deception	519
<i>b. Collective Inaction</i>	520
1. Lack of Trust	523
2. Lack of Sufficient Interest	530
3. Buckpassing	533
<i>c. Bandwagoning</i>	538
<u>2. The Rising Hegemon's Side: Unmatched Military and Economic Growth</u>	550
<i>a. Military Achievements</i>	550
<i>b. Non-Military Achievements and Characteristics</i>	557
Conclusion	565
Chapter [9] Conclusion	567
<u>1. Findings</u>	568
<i>a. What the Cases Tell Us</i>	568
<i>b. Research Probes</i>	587
1. Spain under Philip II	588
2. France under Louis XIV	593
3. Napoleonic France	601
<i>c. Research Agenda: the Rarity of Balance of Power Failure</i>	611
<u>2. The Impact of Modernity</u>	614
<u>3. Balance of Power Failure and Unipolarity</u>	621
Bibliography	636

List of Tables, Figures, Maps and Illustrations

Table 2.1: Structural Outcome (Neorealist Approach)	56
Table 2.2: Structural Outcome (Multi-Level Approach)	57
Figure 2.1: Balance of Power Failure Mechanisms	58
Figure 3.1: Genealogy of the Mongol Khans	86
Map 3.1: The World before the Mongol Empire, ca. 1200	89
Map 3.2: The World after the Mongol Empire, ca. 1300	94
Map 5.1: The Roman Empire at its Height, 125 A.D.	281
Map 7.1: China ca. 278 B.C.	424
Map 7.2: China ca. 210 B.C.	424
Illustration 8.1: "Ten Thousand Miles from Tip to Tip"	497
Illustration 8.2: "Columbia's Easter Bonnet"	555
Table 9.1: Variable Scores in Four Cases of Balance of Power Failure	569
Table 9.2: Individual Factors in Declining Order of Causal Impact	586

Introduction

How do great powers react when they face a rival that may surpass them? Even the mightiest states in the world are concerned about potential competitors that could endanger their position. In its first National Security Strategy issuance of the new millennium, the U.S. government declared that although "today, the United States enjoys a position of unparalleled military strength and great economic and political influence... it is time to reaffirm the essential role of American military strength." The United States "must build and maintain [its] defenses beyond challenge" and "must ... dissuade future military competition."¹ Like the United States, every great power seeks to keep competitors in check and maintain its position in the global configuration of power. In fact, what great power would not want to prevent threatening rivals from arising?

When a great power faces a rising great power, it seems only logical that it would do everything in its power to stop, or at least seriously slow down, the challenger. Not reacting would be dangerous in a world of anarchy, where each state must provide for its own security and where the support of other states is not guaranteed. As the power of the competitor grows, the power of other great powers correspondingly decreases and they become more vulnerable to the competitor. Because it is impossible to accurately foresee the intentions of other actors in international relations, no great power can afford to be in a position of vulnerability. The

¹ The White House, The National Security Strategy of the United States of America, Washington, D.C. (September 2002), 29.

contender could be peaceful and cooperative, but it could also become hostile and seek expansion at another great power's expense. The stronger the contender becomes, the more difficult it is for other powers to ensure their security. In fact, if the contender is left alone to freely develop its capabilities, it may become a hegemon—a state with power so overwhelming that other great powers would no longer be able to defend themselves if it became hostile, but would be at the mercy of the competitor. In essence, their basic survival as states would be at stake. Any great power thus logically wants to maintain an edge over potential challengers and keep them in check so they do not become hegemons. Ignoring a competitor would not only be risky, and likely costly and dangerous, but also potentially fatal. If great powers want to ensure their security and survival, then, they need to keep the configuration of power stable and to their advantage, and in order to do so, they must act against any power that threatens to become hegemonic.

Since antiquity, great powers have been doing just that. Rising great powers have recurrently been met with renewed strength, military and economic resistance, and even war, when necessary. Contenders for hegemonic power and position have been persistently opposed. Sparta and its allies decided to go to war against Athens in 431 B.C. in the Peloponnesian War to check the growing power of Athens. According to Thucydides' account of the conflict, the Spartans reasoned that if they waited and did not act against Athens, their rival might soon be strong enough to take over all of Greece and overcome Sparta, thus becoming a hegemon in the region. To avoid eventual defeat, the Spartans had to "crush... [the] enemy in its infancy, [rather than] wait until it ha[d] doubled its strength."² Similarly, during the Punic Wars in the 2nd century B.C., several powers joined Carthage in its efforts to rein in the growth of the emerging Roman Empire—another oft mentioned example. King Hiero of Syracuse joined the fight against

² Thucydides, *History of the Peloponnesian War*, transl. Rex Warner (New York: Penguin Books, 1954), I. 69.

Rome despite having previously signed an alliance treaty with the Romans, according to Polybius, because he was "convinced that it was in his own interest for securing ... his ... dominions." Polybius lauds Hiero's decision, remarking that Hiero "reasoned very wisely and sensibly" and recognized the danger of rising Rome and the necessity to stop it before it became hegemonic.³

Instances of great powers meeting their challengers with force continue through modern times and are too numerous to list. Attempts at hegemony, particularly in the European theatre, have repeatedly been put down, and have been amply discussed and analyzed. Most notably, rulers such as Charlemagne and successive Holy Roman Emperors, Philip II of Spain, Louis XIV, Napoleon, Wilhelm II, and Hitler all tried to surpass their fellow great powers and take over the European continent. In each case neighbors near and far reacted, separately or in concert, and the would-be hegemon failed to become a hegemon.

Today as well, as the U.S. National Security Strategy document shows, great powers continue to take steps to slow contenders for hegemony and keep them in check, even though open conflict has become infrequent. The United States is reinforcing its capabilities to stay ahead of a rising China, for example, and China and other aspiring great powers are doing the same vis-à-vis the United States as they worry about America's growing power and influence. In response to a recent Pentagon report on China that highlighted China's increasingly strong military capabilities and its 17.8% increase in military spending in 2007, U.S. Defense Secretary Robert Gates expressed his uneasiness about the growth in Chinese power: "It would be nice to hear first-hand from the Chinese... we wish there were greater transparency, that they would talk

³ Polybius, *The Histories*, transl. W.R. Paton (Cambridge, MA: Loeb Classical Library, 2010), I.83.

more about what their intentions are."⁴ Although China's avowed goal is peaceful rise, it is competing with the United States for power and a hegemonic role in Asia. As "the world is experiencing a profound transformation ... China should be alert to potential risks," PRC president Hu Jintao recently affirmed at the 80th anniversary of the founding of the People's Liberation Army. Therefore, Hu continued, China will keep fortifying its "economic growth [as] the basis for enhancing defense capability, which is, in turn, an important indicator of overall national strength."⁵

Thus, keeping rising hegemony at bay, whether by simply strengthening one's power to deny further relative power to the rival, or by forcefully opposing the would-be hegemon and going to war if necessary, is not only a logical, but also common behavior in international relations. In fact, stopping aspiring hegemony is such standard conduct among states that scholars consider it a law of international relations. Studying such conduct has become a primary focus of the realist branch of international relations. These scholars refer to the tendency to act in all possible ways to prevent the rise of a hegemon as the theory of balance of power. When states face an emerging hegemon that threatens to out-power them, they balance its power. They enhance their power compared to the rival's by strengthening their capabilities and/or banding together to restore the preexisting distribution of power. Because states must always balance power if they want to preserve their security, there is an automatic and recurrent propensity to balance power in international relations. Whenever the world is off-balance, i.e., when one or several great powers set out to out-power and dominate all the others, other states naturally act to restore an equilibrium.

⁴ "U.S. Fears Grow over China Military," *BBC News* (25 May 2007), available at <http://news.bbc.co.uk/2/hi/6691691.stm> (accessed Aug. 10, 2010).

⁵ "President Hu: PLA Budget to Rise with the Economy," *China Daily* (2 August 2007), available at http://www.chinadaily.com.cn/china/2007-08/02/content_5447330.htm (accessed Aug. 10, 2010).

Scholars have recognized balance of power as a quintessential pivot of international relations for centuries and perhaps even millenniums, since at least Thucydides and Polybius, making it one of the oldest theories of international relations.⁶ Explaining King Hiero's decision to join Carthage, Polybius mentions that as the power best able to stop rising Rome, "Carthage should be preserved, and ... the stronger power [Rome] should not be able to attain its ultimate object entirely without effort." Other powers should thus have helped Carthage restore the balance of power.⁷ Even before Polybius and Thucydides, Chinese general Sun Tzu, writing 2,500 years ago in what is perhaps the oldest known military treatise, remarked that "... the men of Wu and the men of Yueh are enemies; yet if they are crossing a river in the same boat and are caught by a storm, they will come to each other's assistance just as the left hand helps the right." In other words, when a common, hegemonic threat arises, states should forget their immediate quarrels and join hands to balance its power.⁸

While the core assumption of balance of power—that states will balance power to prevent any other state from becoming a hegemon—has thus been around for ages, scholars have continued to analyze the theory and its manifestation up to this day. The most significant refinement to balance of power theory came about in 1979 with Kenneth Waltz's *Theory of International Relations*, which systematizes the concept of balance of power from a neorealist perspective by formalizing the theory's structural hypotheses.⁹ The debate has since tended to shift away from grand-theoretical inquiry and focus increasingly on more detailed questions,

⁶ Greek historians Polybius and Thucydides wrote in the 2nd and 4th/5th century B.C., respectively. Some historians speculate that Mauryan politician Chanakya (a.k.a. Kautilya and/or Vishnugupta), who wrote the Sanskrit political treatise *Arthśāstra* in the 3rd/4th century B.C., may also have been among the first scholars to write about balance of power. But because translations remain interpretative, historiographer Herbert Butterfield warns, it is impossible to be sure. See Herbert Butterfield, "Balance of Power," in Philip P. Wiener, ed., *Dictionary of the History of Ideas: Studies of Selected Pivotal Ideas* (New York: Charles Scribner's Sons, 1973).

⁷ Polybius, I.83.

⁸ Sun Tzu, *The Art of War*, transl. Lionel Giles (El Paso, TX: El Paso Norte Press, 2009), XI.30.

⁹ Kenneth Waltz, *Theory of International Relations* (Boston: McGraw Hill, 1979).

such as how and when states balance. Because balance of power has been so extensively and thoroughly examined, it is now considered "the best known, and perhaps the best, theory in international politics."¹⁰

The theory provides, in a nutshell, that when states face a rising hegemon, balancing its power is not only the most logical thing to do and the most frequent reaction they have had historically to preserve their security, but it is also the soundest behavior. Yet, even though great powers always strive to check would-be hegemons and have done so recurrently throughout history and to this date, they do not always succeed. Sometimes the aspiring hegemon continues to grow until it becomes too powerful to stop. When facing a rising hegemon, some great powers balance but do not manage to check the rival, while others even fail to balance at all. During the Peloponnesian War, for instance, Sparta's balancing actions proved successful and led to Athens' surrender in 404 B.C. Carthage and Syracuse, however, failed miserably in their efforts to stop the expanding Roman Republic, which not only conquered both, but also grew to become one of the greatest empires of all time, dominating a vast portion of Europe, the Middle East, and North Africa.

Since failing to check an emerging hegemon can thus put a state's survival in peril, as the fate of Carthage and Syracuse shows, the occurrence of balance of power failures is puzzling. A number of great empires have emerged throughout history, taking over their immediate and distant neighbors, recurrently using force to expand beyond their borders. A few hegemons have risen to maturity over large portions of the globe, sometimes even spreading over several continents, like the Romans and the Mongols. Most have ascended to power on a regional scale, controlling or taking over a whole region, as did the Qin Dynasty in China, or the United States

¹⁰ Robert Jervis, *Systems Effect: Complexity in Political and Social Life* (Princeton: Princeton University Press, 1997), 131.

in the Western Hemisphere. Their growth into hegemons did not materialize overnight; instead, it was generally a slow and foreseeable process. Yet, contrary to Waltz's predictions, it was not stopped. If balance of power is such a potent, natural, and vital force in international relations, why does it sometimes fail and allows hegemonic bids to succeed? The aim of this dissertation is to explain this puzzle by offering a systematic framework to understand balance of power failures.

1. Questions Addressed

This thesis asks two new, significant questions that promise to bring original contributions to the international relations scholarship.

First, why do balance of power failures occur? Under what conditions does the balance of power mechanism fail? The primary aim of this project is to provide the first systematic explanation of balance of power failures. Because balance of power and hegemony are system-level outcomes, I initially set out to seek an explanation for the success of hegemons within the structural framework in which Waltz's neorealist theory operates. I reasoned that since balance of power was the most common outcome historically, and hegemony occurred only occasionally, there must be a structural explanation for this anomaly that neorealists might have overlooked. I soon realized, however, that a purely structural approach like neorealism cannot account for balance of power failure. If we take unit motivations for granted, as a structural approach does, and assume that states always put their security first, then by definition there is no room for hegemony because security-driven states will always, at least in aggregate, act to prevent any peer whose power threatens to become overwhelming. Yet, hegemony has occurred.

Since systemic effects derive from aggregated unit-level behavior, if the systemic outcome defies predictions, then something must have happened at the unit level that the structural assumptions do not account for. The unit level must be richer and more complex than neorealism presumes. When hegemony succeeds, for instance, states must have somehow gotten their preferences mixed up or failed to correctly assess the aspiring hegemon's threat to their security. The occurrence of hegemony does not mean that structural factors do not matter, but rather highlights that non-structural factors combine with structural factors to yield a less determinate outcome than neorealism predicts. Neoclassical realism provides the answer to the thesis's first question, therefore, by allowing for both systemic and unit-level factors in explaining outcomes. This thesis thus aims to locate which factors, both at the systemic and unit levels, are most important in explaining why balance of power failed in a set of critical historical cases of hegemony and to derive from those observations the causal path to balance of power failure.

Explaining why balance of power sometimes fails has great intrinsic value because it is a phenomenon of international relations that remains largely unexplained. While the scholarship on balance of power provides unending lists and analyses of successful balancing, it does not account for failures and the incidence of hegemons. Balance of power success is often simply taken for granted, and most scholars overlook cases where it does not work and hegemons take power. In reality, no systematic study of balance of power failure has so far been undertaken, which constitutes a serious gap in the literature. A few scholars have begun to look at the issue, but their analyses remain incomplete. A thorough investigation of balance of power failures is indispensable not only to close that gap, but also to complete and refine balance of power theory itself. Hegemons are a blatant exception to the theory of balance of power; they contradict the

outcome it predicts. The existence of balance of power failures therefore represents a major empirical breach in the theory that weakens its explanatory power. A scientifically driven theory must account for such key exceptions that falsify its hypotheses. In reality, balance of power theory has never been thoroughly tested, and Waltz has suggested that examining outcomes that falsify the theory would constitute a strong test.¹¹ To complete our existing understanding of balance of power, therefore, we must explain why, given the central role of balancing behavior in international relations, some great powers balance against their opponents unsuccessfully or fail to balance against them at all before it is too late. Although the goal of this project is mostly to explain the occurrence of hegemony, and not to test balance of power theory, the systematic analysis of balance of power failure that it provides in effect represents a first step in testing balance of power theory itself. This dissertation, by identifying the conditions under which balance of power fails, also by necessity lays out the conditions for it to succeed.

A second and closely related question is, why does balance of power failure occur so infrequently? Great powers have relentlessly attempted to reach hegemony, but only a handful have succeeded throughout several millenniums of international history. In the overwhelming majority of cases, aspiring hegemons fail and the balance of power is restored. Why did hegemony succeed in those few cases but not in others? This thesis will focus on answering the first question and providing an explanation for failures, because that has not yet been thoroughly investigated and requires considerable empirical research. Although this thesis will not provide a definite answer to the second question, it offers a few suggestions as to what might explain the infrequency of hegemonic success, and proposes an agenda for future research on that subject.

¹¹ Waltz, 123-4.

2. The Argument

My purpose is to develop a theoretical framework that maps out the causal mechanism leading to balance of power failure. What causes this outcome? What factors explain why states may fail to prevent an aspiring hegemon from becoming an actual hegemon? I identify a total of twenty-one factors that operate both on the systemic and unit levels and appear most important in enabling hegemons to arise. All rely on international relations phenomena previously discussed in the international relations literature, but that have not yet been integrated into one model and causally associated with balance of power failure.

The factors focus on two sets of actors: the potential balancers, and the rising hegemon. Although potential balancers necessarily play a critical role in determining whether hegemons fail or succeed, since by their intervention they can stop an aspiring hegemon and restore the balance of power, they only constitute one side of the equation of balance of power. The achievements, or under-achievements, of the rising hegemon also matter in determining whether it succeeds or not. Because my aim is not solely to explain why great powers balance or fail to balance in the face of a rising hegemon, but rather to find out what accounts for the structural occurrence of hegemonic systems instead of the predicted balance of power systems, the factors I identify include both balancer-driven and hegemon-driven factors. The role of the rising hegemon in the balance of power equation is a facet that is notably absent from both the neorealist theory of balance of power and the balance of power literature more generally, because it pertains to the ability of the rising hegemon to extract and use resources efficiently, which essentially depends on unit-level choices.

The twenty-one factors of balance of power failure are grouped into six main clusters of explanation:

First, dysfunctional communication (1) can lead potential balancers to lack the necessary information to stop a rising hegemon and thus fail to act. This is a cognitive, unit-level variable and the only variable for which both the potential balancers and/or the aspiring hegemon may be responsible. There are three types of communication problems: physical problems (1.1), due to distance or poor communication technology, that leave the potential balancers unaware of the rise of the hegemon or the possibility of joining others to balance collectively; perceptual problems (1.2), when a state misperceives the systemic pressures, such as the rising hegemon's threat or its fellow balancers' balancing actions; and deceptive problems (1.3), when a potential balancer is deliberately misled, either by the rising hegemon or by others, about the rising hegemon's threat or the possibility of balancing. Communication hurdles can operate either directly by undermining a potential balancer's individual decision to balance, or indirectly by hampering collective balancing possibilities.

The next two explanations, collective inaction (2) and laggard balancing (3), are rooted in problems of cooperation, and concern the potential balancers. When facing a rising hegemon, great powers can balance individually, but because a rising hegemon is often uniquely powerful, pooling their efforts and balancing collectively may yield better results. However, because cooperation is inherently difficult in anarchy, states may make decisions that fall short of successful balancing. Either they are immobilized by collective inaction (2), and fail to cooperate and form a balancing coalition altogether; or they engage in laggard balancing (3), managing to join forces and balance, but too late, too slowly, or not efficiently enough to stop the rising hegemon. There are four main types of cooperation problems that lead to sub-optimal balancing: indirect communication problems (2.1 and 3.1), when balancers suffer from communication problems that specifically render them incapable of coordinating their balancing efforts; trust

issues and a focus on short-term gains (2.2 and 3.2), when potential balancers are more intent on improving their position at the expense of other potential balancers than on the longer term threat of the rising hegemon and as a result suspicion between potential balancing partners is acute; lack of sufficient interest (2.3 and 3.3), when potential balancers hesitate to participate in a common balancing effort because the rising hegemon is far from them or not powerful enough to be an immediate threat to them; and buck-passing (2.4 and 3.4), when some potential balancers attempt to free-ride and pass the cost of balancing onto other potential balancers. Collective inaction and laggard balancing occur because cooperation is inherently difficult in anarchy. Therefore, they constitute unit-level deficiencies but are in reality derived from the prime structural characteristic of international relations, anarchy, and can thus be considered to be structurally-driven factors.

A fourth explanation, also concerning the potential balancers, is that some states, instead of balancing the rising hegemon, decide to bandwagon with it, thus inflating the rising hegemon's relative power (4). Some potential balancers may bandwagon out of fear (4.1), calculating that the rising hegemon may treat them better if they help it. Others bandwagon out of profit (4.2), reasoning that by collaborating they could extract advantages out of the rising hegemon once it becomes all-powerful. Yet others may bandwagon for both reasons simultaneously (4.3). Neorealists generally argue that it is mostly weaker states that bandwagon because they have few or no alternatives, while great powers tend to chose balancing over bandwagoning because the risks involved with bandwagoning undermine their security. Even collaborators are at the rising hegemon's mercy once it has achieved hegemony. But bandwagoning could be more prevalent than neorealists acknowledge and involve more than just the weakest states in cases when rising hegemons succeed.

The last two explanations look at the unit-level achievements of the rising hegemon. It may extract and use its resources in such superior ways that potential balancers are unable to emulate and/or counteract. The rising hegemon's military achievements matter greatly, of course (5): it may manage to develop particularly innovative weapons technology (5.1), or unique strategies or tactics that give it an edge (5.2), or even a novel, rationalized military organization that no other great power has mastered (5.3). In addition to unique military assets, the rising hegemon may also gain a crucial advantage by developing certain non-military skills and characteristics that other powers are lacking (6), such as a valuable geographic position (6.1), highly profitable economic assets (6.2), or uniquely effective political and administrative (6.3) or social (6.4) structures that may reinforce its relative power position and facilitate efforts to expand its control.

We should not expect any of these factors to supersede all the others and singlehandedly explain the outcome. Rather, several variables will undoubtedly work simultaneously to cause balance of power failure. The goal of this thesis is to assess which factors play a primary role in allowing aspiring hegemons to succeed, which factors are secondary, and which factors are merely marginal. Empirical research will allow us to do this and rank the factors in order of importance.

3. The Evidence

My evidence is based on case studies, using the method of structured, focused comparison familiar to qualitative analysis, which will test the twenty-one variables laid out above. In addition, the conclusion will assess a set of short research probes or control cases of balance of power success, to confirm that the variables identified as primary causes of hegemony

are also the most important variables responsible for successful balance of power, and thus avoid the kind of selection bias that has vitiated previous studies of balance of power.

Cases of balance of power failure are not only limited in number, but have also never occurred at the global scale. Hegemony is mainly a regional phenomenon, save perhaps one exception. The Mongols are one case of what I call historic global hegemony, where one state dominated not the entire world as we know it now, but the entire world as it was known then. The Mongols controlled most of the world as they knew it, from East Asia to the Middle East to Europe. They did not know America, Australia, and most of Africa, so in light of history they were a global hegemon. The absence of a true global hegemon is not a problem, though. One can simply reduce the size of the system used as a reference point and study cases of regional hegemony to observe the phenomenon's causal patterns. There are several historical cases where one state dominated an entire region: the Roman Empire, the Qin Dynasty in China, and the United States in late 19th/early 20th century, are the most notable ones.

4. Structure of the Project

Following this introduction, the second chapter forms the theoretical core of the dissertation. It situates this project within the international relations literature, defines the notion of balance of power failure, and lays out the twenty-one possible causal mechanisms behind it. The next six chapters consist of detailed, empirical tests of the twenty-one factors, with the first two case studies divided along the lines of balancer-driven and hegemon-driven factors: the Mongol conquests (chapters 3 and 4), the rise of the Roman Empire (chapters 5 and 6), the unification of China under the Qin (chapter 7), and U.S. hegemony over the Western Hemisphere in the late 19th and early 20th century (chapter 8). A final, ninth chapter, concludes by analyzing

the results across cases, evaluating the evolution of balance of power failure over time and speculating on its infrequency, and offering a brief application of the thesis's framework under the current unique constraints of unipolarity.

The Mongol Empire is the largest case of hegemony the world has witnessed, and perhaps the only quasi-global hegemon that controlled more than a region. They were a global hegemon in their time, before the many future territorial discoveries that completed our world map but of which they were unaware. The Mongols constitute the most fascinating case of expansion in history. In a time span of barely two hundred years, they were able to control almost all of the world as it was then known, building an empire that ranged from Korea to Western Europe, and encompassing Russia, the Middle East, the Persian Gulf and Southeast Asia. The Mongols evoke images of incredible military prowess and merciless barbarism that sparked fear so great it is present to this day in the European and Asian collective psyche, through stories and expressions that have transcended the centuries; yet, the Mongols were only dispersed nomadic horsemen, often inferior in number and weaponry to the great royal armies and feudal mercenaries of the time. How is it possible that they succeeded in controlling virtually the entire civilized world?

Ancient China provides a great opportunity to examine balance of power failure at the regional level. Because of the constant change in the number of states and distribution of power in the region, and the quasi-continuous warfare that went along with it, balances of power in ancient China constantly formed and re-formed, providing several examples for analyzing balancing behavior. Although several kingdoms have united China throughout history, the rise of the Qin during the Warring States period is a remarkable case in that it is the only time one kingdom of ancient China became so overwhelmingly powerful and controlled such an extensive

amount of land that it not only united all of China but also reached well beyond. While other great Chinese kingdoms may have unified the various Chinese fiefs, the Qin expanded to control the entire East Asian continent. Yet, although they were late-comers to the Chinese struggle for supremacy compared to other, better-established kingdoms, the Qin were surprisingly never checked and had little difficulty acquiring a hegemonic grasp over the region.

The rise of Rome, just like the rise of the Qin, constitutes another case of regional hegemony. In fact, the rise of both empires coincides, since the Qin completed their unification of China in 221 B.C. and Rome began expanding its influence beyond the Italian peninsula around 241 B.C. Both conquered the entire continent surrounding them, but neither controlled areas beyond their extended regions, as evidenced by the fact that the two regional hegemonies coexisted. Unlike the Mongols, they are not cases of historical global hegemony, where the hegemon controls the entire world of its time, because both the Chinese and the Romans coexisted and were aware of each others' presence. In the first century A.D., a king from the Han dynasty, which succeeded to the Qin dynasty, sent an emissary to obtain allies against his barbarian enemies and open the silk trade route to the West. The emissary was instructed to get in touch with the Roman Emperor, but his mission eventually failed before he was able to reach Rome.¹² The Roman Empire rose slowly, from a small, weak Republic, to one of the most historically prominent and best-documented regional hegemonies, expanding all around the Mediterranean Sea into Europe, Africa, and the Middle East. While some of its neighbors clearly understood the danger it posed as they witnessed its rise, their balancing actions all inexplicably failed, and all shared Carthage and Syracuse's dismal fate.

The rise of the United States and its domination of the Western Hemisphere in the late 19th and early 20th century is the only example of a great power achieving regional hegemony in

¹² Dun J. Li, *The Ageless Chinese: A History* (New York: Charles Scribner's Sons, 1978), 111.

the modern world. Throughout the late 1800s, the United States succeeded in driving the European great powers out of their traditional areas of influence in North, South, and Central America and the Pacific and pursued a policy of influence and domination over most countries of the American continent, with virtually no opposition. Little by little, the United States, using tools ranging from economic aid and financial investment to military coercion and intervention, took advantage of the growing conflicts that occupied the European powers on their own continent, and consolidated its grip over the Western Hemisphere, easily becoming a regional hegemon.¹³

¹³ Christopher Layne, "Kant or Cant: The Myth of the Democratic Peace," *International Security* 19.2 (Autumn 1994), 22-28; Martha Finnemore, *The Purpose of Intervention: Changing Beliefs about the Use of Force* (Cornell: Cornell University Press, 2004), 33

A Conceptual Framework for Balance of Power Failure

This chapter offers a review of balance of power theory and argues that our understanding of the theory will remain incomplete until we look more closely into instances where it does not work. It then examines the notion of balance of power failure and maps out twenty-one causal mechanisms that lead to such failures.

Although the concept of 'balance of power' has been prominent in the international relations vocabulary for centuries, it has often been a source of misunderstanding because policy-makers, journalists, historians, and political scientists give it many different, and often contradictory, meanings. As Martin Wight explains, "the notion of balance of power is notoriously full of confusions, so that it is impossible to make any statement about the 'law' or principle of the balance of power that will command general acceptance."¹⁴

There are several reasons for this lack of clarity. The word 'balance' itself is ambiguous— for some it stands for a perfect equality of forces, while for others it implies the ability to maintain or restore that equality, and thus may presupposes one party's control over the rest. In addition, balance of power can have both a normative and a descriptive connotation. In some narratives it is viewed as a fact or tendency of international politics observed in history, while in others it entails a prescription for foreign policy behavior. To complicate matters even further,

¹⁴ Martin Wight, "The Balance of Power," in Herbert Butterfield and Martin Wight, eds., *Diplomatic Investigations: Essays in the Theory of International Politics* (Cambridge: Harvard University Press, 1996), 149-151.

sometimes the descriptive and normative approaches are combined. Finally, the balance of power is difficult to estimate, and determining what the exact balance is necessarily implies some subjectivity. "Clearly, the meaning of the term is obscured by the varying intentions of writers employing it," as Ernst Haas concludes.¹⁵

Sorting out the multiple definitions of balance of power is not an easy task. Several prominent scholars offer classifications of the multiple meanings generally associated with the notion – including Wight, Haas, Hans Morgenthau, and Inis Claude. But each winds up with a different number of definitions and all admit that their categories overlap.¹⁶ To avoid this, we can eliminate several definitions that are inadequate or are mistakenly equated with balance of power and in reality describe something different. The term balance of power is frequently used to point out the general distribution of power, when saying for example that 'the balance of power has changed.' This definition is too broad to be useful. Because power shifts across states and time, the balance of power could be just about any configuration of power.

The most obvious use of balance of power is to depict a situation of equipoise, or equilibrium of power between states or blocs of states at a given point, for example when saying 'there was a balance of power in 17th century Europe.' This definition is problematic too. It offers a static view of balance of power that fails to take into account the constant movement of power and alliances. An exact equilibrium is temporary at best, and highly unstable. A static view also exacerbates the problem of subjectivity in measuring the balance, since an equilibrium at any given point, as opposed to a trend overtime, will more likely be perceived differently by the parties at stake.

¹⁵ Ernst Haas, "The Balance of Power: Prescription, Concept, or Propaganda," *World Politics* 5.4 (July 1953), 444.

¹⁶ Wight 151; Haas 446-7; Hans J. Morgenthau, *Politics Among Nations: The Struggle for Power and Peace* (New York: Alfred A. Knopf, 2nd Ed., 1956), 155; Inis Claude, *Power and International Relations* (New York: Random House, 1962), 12-22.

Another common use of balance of power is to suggest one state's superiority or even quest for hegemony, as in 'the balance of power is in our favor.' This definition associates balance with positive sum power, often invoking the image of a bank account balance. It is not satisfying because it concerns disparity in power, the exact opposite of a balance. "If the phrase comes to mean predominance ... the very idea and language of the balance of power ... tends, so to speak, to defeat its own original purpose," Wight recognizes. This definition also gives the balance a propagandistic undertone by encouraging states to twist the 'balance' to their advantage. It can even provide justification for their expansionist policies, conducted in the name of the balance. As Morgenthau explains, "a nation seeking empire has often claimed that all it wanted was equilibrium."¹⁷

Then, balance of power is sometimes used to describe the existence of a balancer, an unaffiliated third party that stands between two roughly even states or groups of states and has the ability to determine the winning side. According to this definition, the third party 'holds the balance of power.' This sense is too restrictive, however. Balancers are not so frequent, and they can only act to maintain the balance as long as they remain uncommitted. Moreover, the neutrality of a balancer is ambiguous, because its unique position gives it a special advantage to determine international outcomes. Thus, it may favor one side for reasons other than preserving a balance.

Finally, some narratives use the term to portray power politics in general—the constant shifting of power distribution and alignments as states use power strategically to determine their position on the international scene. This definition often identifies balance of power with the succession of war and peace and describes the state system as a whole, not specifically balance of power. "It would be absolutely correct to consider the balance of power as a refinement of a

¹⁷ Claude, 14; Wight 168; Morgenthau, 2nd Ed., 176, 192.

general system of power politics," as Haas points out, but it does not encompass the whole system.¹⁸

After eliminating these misleading definitions, we are left with only one meaning, which constitutes the essence of balance of power and which I will henceforth use: the tendency of international systems to revert to equilibrium. This definition avoids all the mishaps of the previous definitions because it is a theoretical approach to balance of power rather than a descriptive or prescriptive one. It enunciates a causal claim (or a 'law of history') about the outcome of international politics. As Herbert Butterfield illustrates:

... the whole order in Europe was a kind of terrestrial counterpart of the Newtonian system of astronomy. All the various bodies, the greater and the lesser powers, were poised against one another, each exercising a kind of gravitational pull on all the rest – and the pull of each would be proportionate to its mass ... When one of these bodies increased its mass, therefore – when, for some reason, France for example had an undue accession of strength – the rest could recover an equilibrium only by regrouping itself, like sets of ballet dancers, making a necessary rectification in the distances, and producing new combinations. Otherwise, the overgrown power would swallow up the little ones near at hand, and become greater still – just as the moon would fall into the earth if there were no counteracting forces to offset the effect of gravity.¹⁹

Thus, balance of power means that whenever the distribution of power becomes so imbalanced in favor of one state that it can threaten the existence of other states, the others will naturally counteract through a balancing effort to restore a more even power distribution. In other words, through balance of power, the international system "tends toward its preservation by avoiding the

¹⁸ Haas 476.

¹⁹ Herbert Butterfield, "The Balance of Power," in Herbert Butterfield and Martin Wight, eds., *Diplomatic Investigations: Essays in the Theory of International Politics* (Cambridge: Harvard University Press, 1996), 132.

hegemony of a single member."²⁰ It is as a theory that the notion of balance of power is most significant, because it moves beyond a mere observation or policy goal and becomes a universal tool of analysis, providing explanations and enabling predictions.

1. The Limits of Balance of Power Theory and Why Failures Matter

While the literature provides considerable detail about the workings of balance of power theory, the theory's strength is limited by insufficient, and almost non-existent, attention given to its exception, balance of power failure.

a. The Workings of Balance of Power

Scholars have spent much time and ink exploring the different facets of balance of power theory, debating the conditions for the theory to work and the various ways of establishing balanced power.

1. What is Balance of Power Theory?

Balance of power theory is rooted in structural realist, or neorealist, thinking. It assumes that states are unitary actors who seek, above all, survival and security. If one state, or one group of states, grows stronger relative to the others, power will be imbalanced and the latter states' relative security will decrease. The weaker states will take actions necessary to improve their security by increasing their power relative to the stronger state and thereby restoring a balance of power. The balance results from the structural constraints inherent to the world system: because the world is anarchic, conflict may arise and states may be harmed. States consequently have an

²⁰ Haas 453.

interest in preserving themselves and avoiding situations of power imbalances that put them at risk. As Waltz writes, “fear of such unwanted consequences [stemming from anarchy] stimulates states to behave in ways that tend toward the creation of balances of power.” The state system thus constantly strives toward power equilibrium. Unbalanced power becomes checked and the outcome of international relations is always a balance of power. As Morgenthau explains, “whenever the equilibrium is disturbed ... the system shows a tendency to reestablish ... an equilibrium.” Because power is in constant motion, balances of power recurrently form and reform. Balance of power creates a “precarious stability” in international relations, which is “always in danger of being disturbed and therefore always in need of being restored.” It is the only way some stability is possible given “the [anarchic] condition of the power pattern.”²¹

States assess the distribution of power and its evolution by evaluating the different characteristics of power. The “balance of power is a function of tangible assets,” John Mearsheimer points out. States pay attention to other states’ growth of not only military capabilities, but also latent power, which includes socio-economic and demographic factors and is a crucial determinant of military power. They react against other states’ increases of power because those can become dangerous and threatening. The “formidable capabilities” of a powerful state generate fear among other states for their security and survival.²² The increased power of the stronger state correspondingly diminishes their ability to defend their independence. States therefore seek to check any state that develops stronger capabilities relative to them.²³

²¹ Kenneth Waltz, *Theory of International Politics* (Boston: McGraw Hill, 1979), 117-118, 128; Hans Morgenthau, *Politics Among Nations: The Struggle for Power and Peace* (New York: McGraw Hill, Brief Ed., 7th Reprint, 1993), 184-5, 188-9; T.V. Paul, “The Enduring Axioms of Balance of Power Theory and their Contemporary Relevance,” in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004), 4-5.

²² John Mearsheimer, *The Tragedy of Great Power Politics* (New York: W.W. Norton, 2001), 55-56, 345.

²³ Morgenthau, Brief Ed., 188-89.

This version of balance of power theory, first defined by Kenneth Waltz, constitutes the most eloquent and parsimonious version of the theory. It is based on just three assumptions: that states are unitary actors who seek survival and operate in an environment of anarchy. As Waltz explains:

A balance of power theory, properly stated, begins with [an] assumption... about states: they are unitary actors who, at a minimum, seek their own preservation and, at a maximum, drive for universal domination... We then add the condition for [the theory's] operation: that two or more states coexist in a self-help system, one with no superior agent to come to the aid of states... The theory, then, is built up, from the assumed motivations of states and the actions that correspond to them. It describes the constraints that arise from the system that those actions produce, and it indicates the expected outcome: namely, the formation of balances of power.²⁴

While most realists agree that Waltz's basic structural assumptions are necessary, some point out that they are not always sufficient to account for the constant restoration of balance of power, and that additional conditions may apply.

2. Additional Assumptions for Balance of Power

Scholars from various realist schools have sought to add to Waltz's neorealist theory by offering additional conditions that may explain with more detail why balances of power constantly reoccur. Those conditions generally center around four questions: what prompts states' balancing actions; who will be the actors most likely to balance; what unit-level characteristics promote a balance of power; and how do states balance overwhelming power.

²⁴ Kenneth Waltz, "Theory of International Relations," in Fred Greenstein and Nelson Polsby, eds. *Handbook of Political Science Volume 8: International Politics* (Reading, MA: Addison Wesley Publishing Co., 1975), 36-37.

What Do States Balance Against?

First, some neorealists have sought to clarify what makes states balance and introduced the idea of a balance of threat. Stephen Walt argues that what triggers the movement toward a balance of power is not directly a gap in relative power, as Waltz's account suggests, but rather the threat posed by the rising power, measured by combining its offensive military capabilities, overall strength (including economic and technological assets and population), distance and expansionist intentions. States respond to threat rather than power, Walt argues, and states that develop stronger capabilities are not opposed unless they engage in actions other states perceive as threatening. Thus, even if one state possesses overwhelming power, it will not be balanced as long as it behaves in a benevolent, non-threatening way toward others. Great powers may not necessarily balance in the face of superior power alone. Walt's approach is enlightening in that it stresses that non-structural factors such as a state's economic strength or its leaders' perceptions intervene in determining balancing decisions and may therefore affect the aggregate balance of power outcome.²⁵

Scholars in other schools of thought have made similar claims. Inis Claude, for example, writes that it is a strong state's "attitude toward external power" and its character and morals, not just "the quantitative measurement of the power itself," that determine how dangerous it appears to others, and thus how likely others are to balance against it.²⁶ Morton Kaplan, as well as Jack Levy and William Thompson offer another variant of the balance of threat argument when they argue that growing naval powers are less prone to balancing than growing land powers, because their frequent "lack... [of] territorial ambitions," preference for economic influence, and focus on

²⁵ Stephen M. Walt, "Alliance Formation and the Balance of Power," *International Security* 9.4 (Spring 1985), 8-12; Stephen M. Walt, *Taming American Power. The Global Response to U.S. Primacy* (New York: W.W. Norton, 2005), 73; Paul 8.

²⁶ Claude 64-65.

navies rather than large armies "increase the tolerance of other national actors" for their power.²⁷ Brooks and Wohlforth similarly believe that it is the imminence of danger, rather than the disparity in power, that triggers balancing movements. They hold that opposition to a dominant state can only be labeled a balance of power effort when it is a response to an immediate security threat and when the opponent's power advantage is so great that it might reach hegemonic status in the very near future. Brooks and Wohlforth maintain that to employ the terms 'balancing' and 'balance of power' for any less dangerous situation is to engage in conceptual stretching and excessively "broaden" the theory.²⁸

The notion of balance of threat is useful but it does not dramatically alter the theory because threat remains fundamentally a function of military power. Walt himself gives power a prominent place in his definition of threat. The threatening behavior of a powerful state may act as a trigger for balance of power, but the absence of immediate threat should not prevent balance of power from occurring because the gap in relative power is by itself a threat. If a state becomes more powerful, it might not be immediately threatening but it constitutes a latent threat that may materialize at any time. As Mearsheimer concludes, a state's increased "offensive military capability is always a tangible threat to [other states'] survival," regardless of its actions, intentions, or character.²⁹ In other words, too much power is its own threat. Thus, balance of power is an automatic adjustment to bring back the stability of a given system whenever one state accumulates a disproportionate concentration of military power.³⁰ A powerful state that

²⁷ Morton Kaplan, *System and Process in International Politics* (New York: John Wiley & Sons, Inc., 1957), 34; Jack S. Levy and William R. Thompson, "Balancing at Sea: Do States Ally against the Leading Global Power?", paper presented at the MIT Security Studies Program seminar, Feb. 24, 2010, 8-10, 20-22.

²⁸ Stephen G. Brooks and William C. Wohlforth, "Hard Times for Soft Balancing," *International Security* 30.1 (Summer 2005), 186, 190.

²⁹ Mearsheimer 45.

³⁰ For a useful discussion of the threshold of military superiority that triggers a balancing reaction, see Jack S. Levy and William R. Thompson, "Hegemonic Threats and Great-Power Balancing in Europe, 1495-1999," *Security Studies* 14.1 (Jan.-March 1995). Levy and Thompson argue that "European great powers have demonstrated a strong

suddenly becomes threatening when it had previously been benevolent will only precipitate the reaction leading to a balance of power; but the reaction would have occurred anyway at some point due to the gap in military power.

What Types of States Balance?

Another set of assumptions realists sometimes seek to add to Waltz's theory concerns what actors are involved in the balancing process. For most neorealists, including Waltz, the actors undertaking the balancing are the system's great powers. Great powers are those states that "have the largest impact on what happens in international politics." Although there have been some disagreements as to exactly how powerful a state must be to qualify as a great power, the status of great power generally implies that a state possesses enough capabilities to contend in a military conflict with the most powerful state in the system, with reasonable chances of weakening it.³¹ Smaller, or secondary, states are not powerful enough to individually influence system-wide outcomes. Therefore, according to Waltz, "a general theory of international politics [like balance of power] ... applies to lesser states ... [only] insofar as their interactions are insulated from the intervention of the great powers of a system." Secondary powers can thus balance, but only amongst each other in their own league and not against a rising hegemon, and their actions will not influence what happens at the level of the system.³²

But although small powers cannot stop a rising hegemon on their own, they may not be as insignificant on a larger scale as Waltz asserts. Small powers, like great powers, are primarily security-seekers, and thus will do what they must to guarantee their survival. However, since

propensity to balance when one state had acquired a third or more of the total military capabilities in the system, but not at lower concentrations of power, and that higher concentrations of power usually lead to larger balancing coalitions" (4, 30).

³¹ Mearsheimer 5.

³² Waltz, *Theory*, 73; Walt, "Alliance Formation," 17.

they have limited capabilities to defend themselves, unlike great powers, ensuring survival often involves finding and collaborating with a great power protector, whether such great power is on the side of the balancers or is a rising hegemon. While small powers cannot hope to affect the system on their own, nothing prevents them from contributing to the outcome, albeit to a lesser extent than great powers, by throwing their collective weight behind one side.³³ They can support a great power balancer directly, or engage in peripheral conflicts, for example, which can raise costs for the rising hegemon and divert its resources from the main balancing effort. Especially if a large number of secondary powers join the balancing side, their aggregate effort may contribute, even if not decisively, to increasing the relative weight of that side.

Unlike great powers, for which balancing is the only option to ensure security, small states' preference for joining either side is indeterminate because their security requires a protector. They will join whichever great power they deem will provide greatest protection. As Dan Reiter explains, "when facing a systemic threat, the affected minor powers could enter alliances to either balance against or bandwagon with the threat."³⁴ Some scholars point out that secondary powers, unlike great powers, are vulnerable to a multitude of security threats, including threats from small-power neighbors, and therefore might be more tempted to ignore the rising hegemon's longer-term threat and bandwagon with it if it offers to help against more immediate local threats.³⁵ Others point out that smaller powers are invariably a rising hegemon's first victims because they are easier to subdue. As a result, "if they are free to choose, [small powers] flock to the weaker side [i.e., the balancers' side] because it is the stronger side that

³³ A number of realist scholars acknowledge that secondary powers are not indifferent to a system's power configuration and do not restrict themselves to interacting with peer small powers. See, for example, Levy and Thompson, "Hegemonic Threats," 8, 29.

³⁴ Dan Reiter, "Learning, Realism, and Alliances: The Weight of the Shadow of the Past," *World Politics* 46.4 (July 1994), 502.

³⁵ Stephen M. Walt, *The Origins of Alliances* (Ithaca, NY: Cornell University Press, 1990), 31-37, 164-5, 171.

threatens them."³⁶ Although their individual impact may remain marginal and they may not balance as systematically as great powers, secondary powers' participation in balancing efforts should not be ignored.

A related question concerns the number of great powers involved in balancing. Some realists argue that there is a minimum number of participants required for balancing to work. Claude points out, for example, that "there should be a broad consensus" on balancing and restraining power if the system is to be preserved.³⁷ Along the same lines, Kaplan writes that if there are not enough great powers that can balance in a system, successful balancing is improbable and instability will follow. A well-functioning balance of power system "must [have] ... at least five and preferably more... essential actors."³⁸ However, as Waltz's second assumption indicates, a balance of power system does not require more than two actors. Although Kaplan contends that bipolarity is not compatible with balance of power, the Cold War's numerous instances of balancing behavior between the superpowers proves otherwise. And while one can discuss the merits of bipolarity or multipolarity for the stability of a system, one great power balancer is by definition sufficient to stop a rising hegemon.

What Are the Characteristics of Balancers?

A third set of additional conditions some realists list as necessary for balance of power to work concerns the domestic characteristics of the balancers. For example, a frequent claim is that states must share similar political systems, political philosophies, and rules of behavior to be able to establish a stable balance of power. Morgenthau highlights the importance of "intellectual and

³⁶ Waltz, *Theory*, 127.

³⁷ Claude 91.

³⁸ Kaplan 22, 24, 26, 34.

moral unity," fostered by shared ideals and close trade relations, in enabling balances of power.³⁹ Organski likewise claims that balancing coalitions are more likely to form around shared ideological affiliations than in their absence.⁴⁰ Edward Vose Gulick makes similar, yet more muted arguments. He acknowledges that the "homogeneity" of the states in the system can be "a necessary precondition for a workable equilibrium," though he recognizes that it is not as critical an assumption as the anarchic state system. "The absence of homogeneity," he explains, "could be described more accurately as crippling to the balance of power, not destructive of it." Gulick also maintains that democracies, because they have a slow decision-making process, are not the most effective candidates for balancing, which often requires quick reaction. In reality, however, successful balancing has frequently occurred when such internal conditions were not present. Although shared political, ideological, and moral characteristics may be conducive to garnering a balancing coalition, strategic alliances also recurrently bring together domestically diverse actors. As Gulick acknowledges, "morality ha[s] little or no place in the balance system, when the "higher interest [i.e., security and survival] [is] at stake."⁴¹

But while those particular state characteristics might not intervene in generating balances of power, the argument that domestic characteristics might matter unearths a deeper issue regarding the theory. Do other unit-level factors influence how great powers react in the face of overwhelming power, and therefore perhaps play a role in the establishment balances of power? As a structural approach, the neorealist account of balance of power focuses exclusively on the characteristics of the international system as a whole and makes no claims about the characteristics of the units that compose it. Waltz and fellow neorealists look at the relative

³⁹ Morgenthau, 2nd Ed., 194-195, 199.

⁴⁰ A.F.K. Organski, *World Politics* (New York: Alfred A. Knopf, 1988), 289.

⁴¹ Edward Vose Gulick, *Europe's Classical Balance of Power: A Case History of the Theory and Practice of One of the Great Concepts of European Statecraft* (Wesport, CT: Greenwood Press, 1955), 20, 24, 68, 70.

power and arrangement of states in the system but not at the composition and individual behavior of each state. What happens inside each unit is not the concern of a structural approach, and neorealism therefore simply assumes that state action is dictated by the system's key structural feature, anarchy, which leads states to be motivated by security concerns above anything else. For neorealists, balance of power is a theory about structural outcomes; it explains the aggregate result of the interaction of units in anarchy—the recurrent formation of balances of power. It tells us how the state system will look given the environmental constraints that states face. It does not explain how each unit will look or what each unit will do, or even what role each unit will play in generating the aggregate balance. As Waltz puts it, "the theory [of balance of power] does not tell us why state X made a certain move last Tuesday. To expect it to do so, however, would be like expecting the theory of universal gravitation to explain the wayward path of a falling leaf." In other words, Waltz concludes, "a theory at one level of generality cannot answer questions about matters at a different level of generality."⁴²

But the existence of balance of power failures poses a serious problem for such a structural approach. The structural approach does not and cannot account for the occurrence of hegemons. If state units act in aggregate as Waltz assumes, always prioritizing their security, it would not allow for the rise of a hegemon. The historical occurrence of hegemons points out that something must happen the unit level that influences the establishment or lack of a balance of power. When a hegemon succeeds, something at the state level must have gone wrong. Some or all of the units must have behaved in contradiction to Waltz's assumptions, favoring other interests over their long-term security, for instance, or misreading the growth of a peer, even if mistakenly. Balance of power theory in fact cannot entirely ignore unit-level behavior, as balances do not simply emerge from the blue. In order for a balance to be established, some

⁴² Waltz, "Theory," 41.

states must inevitably act. Although balance of power is a structural outcome, it is ultimately the units that act, not the structure. It is the aggregate behavior of the units that produces a structural outcome; therefore, how units individually behave, and why they behave certain ways, matters in determining the outcome. Neorealists recognize that balance of power tells us something about state behavior and is thus also a partial theory of foreign policy. Colin Elman points to Waltz's claim that states' rational choice in the face of a rising hegemon is to balance. This constitutes a prediction of states' foreign policies, Elman argues, and qualifies balance of power as a theory of foreign policy. Waltz also briefly recognizes this, writing that "balance of power theory is a theory about the results produced by the uncoordinated action of individual states" and that "it ... provides some clues to the expected reactions of states."⁴³

In order to fully understand the cyclical reestablishment of balance of power, as well as account for the occasional emergence of hegemons, therefore, it is necessary not only to assess, as neorealism does, the structural constraints that bring about balances, but also to look inside the units that aggregately generate balances of power and identify the unit-level characteristics conducive to the establishment of such balances. As long as units are considered black boxes, balance of power theory will explain the general historical trend toward the recurrence of balances of power, but might not be specific enough to account for exceptions. That is why this thesis opts for a neoclassical realist approach, which offers a multi-level analysis, rather than a neorealist, purely structural approach. Neoclassical realism combines structural factors, such as anarchy and the distribution of capabilities, and unit-level factors, including both domestic and cognitive factors, to explain international outcomes. It is in reality the interaction of both levels that generates those outcomes.

⁴³ Colin Elman, "Horses for Courses: Why *Not* Neorealist Theories of Foreign Policy?," *Security Studies* 6.1 (Autumn 1996), 23-25, 30-31; Waltz, "Theory," 41.

The neoclassical realist scholarship underlines that "systemic pressure must [first] be translated through intervening unit-level variables such as decision-makers' perceptions and state structure" before it influences state action.⁴⁴ Because balance of power necessarily works through the foreign policies of some states, the unit-level processing of structural pressures may play a crucial role in the establishment of a balance. Jeffrey Taliaferro, Steven Lobell, and Norrin Ripsman further explain that structural pressures "filter... through the medium of state structure and affect... how top officials assess likely threats, identify... viable strategies in response to those threats, and ultimately extract... and mobilize... the societal resources necessary to implement and sustain those strategies." Thus, "unit-level variables constrain or facilitate the ability of ... states ... to respond to systemic imperatives." Because there is great variation and diversity in unit-level characteristics, as Randall Schweller argues, "states often react differently to similar systemic pressures and opportunities." While recognizing the systemic driving force of anarchy, then, neoclassical realism provides an additional layer of precision by incorporating the "complexities of statecraft" and foreign policy into the analysis, which neorealism bypasses.⁴⁵ Waltz sums up the neorealist approach by declaring that "balance of power politics prevail whenever two, and only two, requirements are met: that the order be anarchic and that it be populated by units wishing to survive."⁴⁶ Neoclassical realism adds a third, critical requirement: that the units adequately perceive and process those systemic constraints.

⁴⁴ Gideon Rose, "Neoclassical Realism and Theories of Foreign Policy," *World Politics* 51.1 (Oct. 1998), 144-77.

⁴⁵ Steven E. Lobell, Norrin M. Ripsman, and Jeffrey W. Taliaferro, eds., *Neoclassical Realism, the State, and Foreign Policy* (Cambridge University Press, 2009), 4-6, Ch. 1, Ch. 7; Randall L. Schweller, *Unanswered Threats: Political Constraints on the Balance of Power* (Princeton: Princeton University Press, 2006), 6.

⁴⁶ Waltz, *Theory*, 118-121.

How Do States Balance?

A fourth issue the literature has sought to clarify concerns the methods states use to balance power. No matter what method states choose, it must work in one of two ways that alter the system's distribution of power: decrease the stronger state's power, or increase their own power. As Robert Art explains, balancing is a reaction triggered when states feel threatened by another state's possession of superior power and reason that more resources will counter the threat.⁴⁷

The most straightforward balancing option is for a state or group of states to use hard assets such as military force and economic power to alter the relative power distribution. Hard balancing can be undertaken by a state internally, by increasing its military and potential capabilities to outpace the rising hegemon, i.e., through an arms race, or externally in a collective effort by an alliance of states that pool their military assets together to counter the emerging hegemon, forming a system of two opposing scales.⁴⁸ Alliances can be problematic, however, because they are often formed in extreme situations like war, rendering them ephemeral and difficult to sustain. In addition, building alliances is a viable option only in a multipolar setting, where there are several great powers available to join the resisting side. Internal balancing, on the other hand, requires vast capabilities and long-term investments that are only available to great powers. Secondary powers can only opt to join great powers and contribute their resources to external balancing efforts, because they are weaker and unable to make a system-wide impact on their own.⁴⁹

⁴⁷ Robert Art, "Europe Hedges Its Security Bets," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004), 179-86.

⁴⁸ Morgenthau, Brief Ed., 197.

⁴⁹ Waltz, *Theory*, 127.

Balancers have other options besides engaging in hard, military balancing. States can also engage in soft balancing, a form of “tacit balancing” short of formal alliances and all-out arms buildup. According to Robert Pape, soft balancing consists of “making a superior state’s military forces harder to use without directly confronting that state’s power with one’s own forces.” It entails the use of non-military tools, but it still constitutes balancing because it “aims” to change the balance and weaken the superior state by raising the cost of its expansion. Examples of soft balancing include “territorial denial, entangling diplomacy, economic strengthening, [and] signaling of resolve to take part to a balancing coalition.” Art also identifies as soft-balancing strategies such as informal alignment or the use of voting and veto power in international organizations. Soft balancing is an especially adequate strategy when the growing power's threat to a state's physical security is not immediate, or when the rising hegemon has acquired a large gap in power and balancers may be wary to provoke a military showdown. It does not directly alter the distribution of power, unlike hard balancing, and is used not to thwart the opponent’s growth but to delay it and gain time. In this sense, soft balancing can be a first step that eventually leads to hard balancing.⁵⁰

In addition to soft balancing, some disproportionately weak states or non-state actors may consider developing tools of asymmetric balancing, which are used to artificially enhance their weight against a rising power. Because such actors are too weak to compete with the leading power using hard or soft balancing tools, as great powers and some secondary powers may do, they use tools such as the acquisition of nuclear, biological and chemical weapons, the promotion of terrorism, or asymmetric warfare to compensate for their weakness. Even if obtaining nuclear weapons, for example, does not drastically affect the relative power structure, it still affects the

⁵⁰ Robert Pape, “Soft Balancing Against the United States,” *International Security* 30.1 (Summer 2005), 9-10, 36; Art, “Europe,” 179-86; Robert J. Art, “Correspondence: Striking the Balance,” *International Security* 30.3 (Winter 2005/06), 183-185.

rising hegemon's growth, at least by significantly raising the costs of expansion and potentially denying it influence and access to certain areas. Asymmetric strategies work for some secondary powers, too, and even perhaps for some great powers when the emerging hegemon has acquired a vast lead. For instance, developing survivable nuclear capabilities may seriously limit the aspiring hegemon's ability to control and influence entire regions, thus diminishing its power, as evidenced during the Cold War. Conventional power still matters, Mearsheimer concludes, but weaker actors increasingly seek to use tools such as nuclear weapons and unconventional warfare in attempts to asymmetrically balance against leading powers. As long as the effect of state action remains to handicap the rise of a great power that accumulates a lopsided share of power, that action constitutes balancing regardless of the means employed—hard, soft, or asymmetric.⁵¹

How do states know that the balance has been reestablished? What is the threshold in terms of power for a balance to be restored? Mearsheimer explains that the equilibrium is reached when there are “no *marked* differences in power between the leading states” in the system. He adds that “power does not need to be distributed equally among all the major states in a balanced system, although it can be,” as long as there is no large gap.⁵² If we follow Jack Levy and William Thompson's assessment that concentrations of power become acute and render hegemony “feasible” when they reach a third of the total military capabilities of a given system, we can assume that power distributions reach a balance when no one great power amasses more than one third of a system's military capabilities.⁵³

Thus, since Waltz formalized the theory of balance of power, the literature has focused on developing the operational details of the theory, debating in particular what conditions may be

⁵¹ Paul 1-4; Mearsheimer 128-133.

⁵² Mearsheimer 337.

⁵³ Levy and Thompson, “Hegemonic Threats,” 6, fn. 17, 20.

necessary for balances to recur as the theory predicts—who balances, what is balanced and how to balance. These debates have gone to great lengths in the past few decades, and many scholars have begun to assume that the literature has sufficiently explored the universe of balance of power theory. Many are under the impression that because there has been such a “theoretical fertility ... exhibited by realism in the last twenty years or so,” the subject of balance of power has been exhausted and additional research efforts are futile.⁵⁴ However, several crucial questions concerning the theory remain unexplored.

Perhaps most importantly, what happens when power is not balanced as the theory claims? Why does balance of power sometimes fail? The occasional emergence of hegemons must be examined if we want to have a complete picture of balance of power. Yet, so far, scholars have mostly evaded the question of the theory's possible exceptions. Levy and Thompson stress that some of the most critical issues that remain on the realist agenda are to "understand exactly how frequently hegemonies emerge in multistate system and ... identify the conditions under which they emerge and the causal mechanisms that drive the process."⁵⁵ These are the two questions that this thesis aims to explore and answer.

b. The Importance of Balance of Power Failure

Studies of balance of power almost always overlook the issue of balance of power failure. When confronting the theory with historical evidence, existing studies most often cite examples from the 18th and 19th century heyday of balance of power politics in the European theatre, which present ideal conditions favorable to balance of power, and provide an easy validation of the

⁵⁴ John A. Vasquez, "The Realist Paradigm and Degenerative versus Progressive Research Programs: An Appraisal of Neotraditional Research on Waltz's Balancing Proposition," in John A. Vasquez and Colin Elman, eds., *Realism and the Balance of Power: A New Debate* (Upper Saddle River: Prentice Hall, 2003) , 25.

⁵⁵ Levy and Thompson, "Hegemonic Threats," 33.

theory.⁵⁶ Claude, for example, recognizes that "reviewers of the record" of balance of power usually "single out ... the... period when the balance system functioned most successfully."⁵⁷ They also single out one system, Europe. "The multistate system in Europe has been fairly stable in that a sustained European hegemony has not formed since Rome, but the same is not true of all other multistate systems, where system-wide empires have occasionally emerged," like Asia or America.⁵⁸ More rigorous testing efforts would have highlighted that in several historical instances, balance of power did not work, and would have prompted a search for the causes of this 'malfunction' of the theory.

Testing remains one of the weaknesses of balance of power theory unanimously recognized by both critics and balance of power theorists themselves. As Paul, Fortman, and Wirtz rightly underline, there have been very few empirical studies undertaken to prove the theory correct. They and others even point out that the theory is empirically weak because it has many historical anomalies, hinting that avoiding serious tests has led us to neglect the question of balance of power failure.⁵⁹ This weakness is problematic, especially since testing is a key component of any theory. As Waltz notes, when we find theories, "we will of course want to know how good the explanations they offer may be."⁶⁰

This thesis seeks to close the empirical gap of balance of power by providing a first systematic analysis of its exceptions, balance of power failures. It specifically investigates two key questions that the literature has not yet answered. First, which factors explain why balance of power sometimes fail? This chapter explores twenty-one multi-level factors, grouped in six

⁵⁶ Waltz, "Theory," 43.

⁵⁷ Claude 67, 70.

⁵⁸ Levy and Thompson, "Hegemonic Threats," 32.

⁵⁹ Michael Fortman, T.V. Paul, and James J. Wirtz, "Balance of Power at the Turn of the New Century," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004), 363-373; Paul 4-11; Vasquez 31-33, 37-39.

⁶⁰ Waltz, *Theory*, 13, 124-5.

categories, that appear to play a causal role in the successful emergence of hegemons, and the following chapters proceed to identify which of those factors are most important by testing each factor in a number of empirical cases. Second, why is balance of power failure such an infrequent phenomenon? Why were certain factors prominent in the few historical cases of hegemony the world has experienced, but not prominent enough to produce balance of power failures in the other, more numerous cases of attempted hegemony that resulted in the restoration of a balance? This thesis asks both questions but focuses primarily on answering the first because it is the starting point to understanding the occasional emergence of hegemons and it requires extensive empirical research. While comprehensive testing of the theory is beyond the scope of the thesis, determining why balances sometimes fail to be established is also a first step in that direction. Developing a conceptual framework for balance of power failure helps refine our understanding of balance of power mechanisms themselves, because determining why and when balance of power does not work also necessarily sheds light on why and when it does.

The few scholars that have touched upon the question of hegemony—from afar or near—fit into one of three categories. A first group rejects balance of power altogether, claiming balances rarely occur in the first place. A second group acknowledges the existence of balances of power, but argues that it is not the most common outcome of international interaction. Instead, they insist, hegemony is more frequent. This second group, which also includes some realists, illustrates that even though most of the literature on balance of power is realist, "not all realist theories posit balanced outcomes and balancing behavior."⁶¹ A third group embraces the theory and its prediction of balanced outcomes, but stresses that sometimes balances fail inexplicably and seeks to understand why. However, their propositions remain incomplete. This thesis belongs to the third group, which recognizes that historically, the outcome of international

⁶¹ Levy and Thompson, "Hegemonic Threats," 6, fn. 16.

politics has overwhelmingly been a balance, but notes that hegemony has occasionally occurred and that such an exception to the theory's prediction must be accounted for.

1. There Are No Structural Balances of Power

A handful of scholars have examined balance of power in the light of historical events and have concluded that balance of power theory is a chimera—empirically, balances frequently fail; in reality, states simply do not balance as the theory would predict. Thus, these scholars approach balance of power failure by discarding Waltz's conclusions. Their findings, however, are not convincing.

Historian Paul Schroeder examines the theory's claim in the light of recent European history. Looking at the post-Westphalian era from 1648 to 1945, the period in which the authority of balance of power theory is “widely accepted even by non-realists” and thus not a very hard test, Schroeder surprisingly finds that balancing against a potential hegemon was “relatively rare.” States, in fact, reacted to the threat of increased power in a variety of unforeseeable ways, including by hiding and bandwagoning, rather than by balancing.⁶² Richard Rosecrance and Arthur Stein, who take a liberal approach, similarly find that balance of power is empirically disproved because states engage in different behavior when facing threat, not necessarily balancing. In the examples they study, states engage in a wide variety of actions, and only very few “get it just right” and balance.⁶³

Michael Doyle, who also embraces the liberal tradition, argues that the logic of balance of power is frequently overridden by other considerations that complicate the picture. As a result,

⁶² Paul Schroeder, “Historical Reality v. Neo-Realist Theory,” *International Security* 19.1 (Summer 1994), 111, 118.

⁶³ Richard Rosecrance, “Overextension, Vulnerability, and Conflict: the 'Goldilocks Problem' in International Strategy,” *International Security* 19.4 (Spring 1995), 145; Richard Rosecrance and Arthur Stein, eds., *The Domestic Bases of Grand Strategy* (Ithaca: Cornell University Press, 1993), 17-21.

we cannot conclude that balance of power constitutes the most potent explanation for international alignments. Doyle believes there are "striking departures from power balancing [in] significant episodes of international history." Although it may often look like states balanced, their actions were based on a different logic, he explains. In the sixteenth century states were prompted by religious rivalry "as much as, if not more than, ... measurable capacity, or power." In the nineteenth century it was internal revolution, colonial competition for "trade and prestige," and a quest for "national dignity" that led states to join a certain side. In the twentieth century those concerns were replaced by "ideological divides among liberalism, fascism, and communism," which became the main cause of alignment and "overcomplicated" the balance of power logic. Thus, in certain historical periods states may not have actually balanced, though the outcome may have accidentally turned out to look like a balance.⁶⁴

These scholars' conclusion that balances rarely if ever occur is in reality distorted by the narrow timeframe they choose. They look exclusively at the short term behavior of states, rather than at the eventual outcome. In other words, they ask whether or not individual states balance, and not whether in the end hegemonic bids succeed or balances are reestablished as a result of those states' actions. Their focus is solely on foreign policy, not on the impact of foreign policy on structural outcomes, as evidenced by the fact that although they argue that states rarely balance, they do not claim that hegemony triumphs either. It is not a surprise, then, that they find that many states do not balance. For a balance to be established, some states obviously need to balance, but there is no requirement that every state balance as soon as the threat of a rising hegemon is perceived, or that every state balance at all, or that states' only goal is to maintain balanced power in the short term. Even if some states may hide or bandwagon in the historical

⁶⁴ Michael Doyle, *Ways of War and Peace* (New York: W.W. Norton & Company, 1997), 172-174, 462; Michael Doyle, "Balancing Power Classically: An Alternative to Collective Security?" in George Downs, ed., *Collective Security Beyond the Cold War* (Ann Arbor: University of Michigan Press, 1994), 147-148.

cases that Schroeder and Rosecrance mention, and even if states may sometimes be spurred by things other than power, as Doyle mentions, this does not necessarily result in a balance of power failure. It is likely that the balance will eventually be reestablished, either because other states balance or because the states that first hid, bandwagoned or had other goals eventually join the balancing coalition when the threat becomes too great.

During the Napoleonic wars, for example, which Schroeder mentions as a malfunction of the balance of power, some states did bandwagon and hide, especially the smaller states weary of the French threat. Yet, in the end, France was defeated by England, Russia, and Prussia. The most convincing example is probably World War II. Schroeder contends that the events leading to World War II disprove the logic of balance of power. It is true that while Nazi power was expanding dangerously in the 1930s, France and Britain hid and avoided action by passing each other the buck, and the Soviet Union joined the rising hegemon's side in an alliance with Hitler. Yet, Nazi Germany did not overtake Europe. As the threat became unbearable, France, England, and later the United States entered into a balancing alliance, joined by the Soviet Union, which abandoned its bandwagoning bid to restore the balance of power. The historical events of the post-Wesphalian era are thus in no way examples of balance of power failure, as Schroeder and Rosecrance claim, or insufficient balancing, as Doyle suggests; instead, they are strong examples of balancing success. During that period, no hegemon successfully emerged. Their conclusions that balances simply do not occur the way the theory predicts are mistaken because they focus solely on the short term behavior and purpose of states, which may be erratic, and not on the long term combined result of their actions.

2. Balance of Power Is Not the Most Common Outcome in International Relations

The second group of scholars is less categorical. Instead of casting doubt on the empirical soundness of balance of power, they recognize that balance of power has explanatory value in international relations and is supported by historical evidence. However, they reject the neorealist hypothesis that anarchical state systems naturally and universally strive for balanced power and the prediction that the outcome of international interaction is always a balance. They admit that balance of power is sometimes the best theory to explain international outcomes, but not all the time. Instead, the English School, which blends constructivist and realist outlooks on international relations, and the power transition school, a sub-group of realism, assert that in reality, hegemony, and not balance, is the most common outcome of international politics.

The English School envisions the international system as a society of states whose organization at any given point can be placed along a spectrum that oscillates between two extremes: sovereign, independent states (a system dominated by anarchy) on the one side, and hegemony (from loose authority to determine the rules of the system to imperial, direct territorial control of the system), on the other. Looking at history, English School theorists have concluded that although occasionally there are balancing movements that prevent hegemony, hegemony, not balance, is the most frequent outcome. Adam Watson sees a general tendency toward hegemony as early as ancient times and argues that the European-centered system of the 16th-20th centuries also exhibited hegemonic trends. Watson admits, however, that some systems have a deep "commitment to an anti-hegemonial principle, the insistence that independent states must in the last resort be free to act as they see fit, ... that hegemony may be legitimately resisted by force" even at the cost of war—so in the midst of a history mostly characterized by hegemony, balances have sometimes occurred. But "the gravitational pull toward hegemony ... stands out so

clearly from the evidence that the question arises why studies of state systems ... underestimate or even ignore it."⁶⁵

In fact, the English School views successful balances as an exception because it uses very different definitions than most realists. For English School theorists, in particular, the assumption of anarchy is much more malleable than for realists. Adam Watson adopts a constructivist perspective where anarchy does not necessarily imply self-help and mistrust of other states, and thus where balancing is not necessary for survival when facing a great power that accumulates a large concentration of power. Anarchy can be cooperative, he argues, so dependence upon another state, even a hegemon, does not carry the dire consequences predicted by realists. Dependence on a hegemon can even be beneficial, according to Watson, because the "hegemonial authority order[s] the system in such a way that all its members see a balance of advantage in accepting the hegemony." As a result, states do not always reject and oppose hegemony, as realists claim, but instead frequently embrace it. The English School also favors a much more broadly defined notion of hegemony. Far from being a world where states may be erased from the map at the will of the hegemon, hegemony is simply a relatively harmonious system where one state or one group of states can "to some extent" influence the others' international interaction.⁶⁶ Thus, English school theorists reach the conclusion that balance of power failures are frequent because they start from different, looser definitions of anarchy and hegemony. While they may recognize that balances sometimes occur, their work draws on constructivist assumptions at odds with the realist framework of balance of power, and is thus too singular to serve as a reference point for this study.

⁶⁵ Adam Watson, *The Evolution of International Society: A Comparative Historical Analysis* (New York: Routledge, 1992), 313-314.

⁶⁶ *Ibid.*, 13-16.

The power transition school reaches a conclusion similar to that of the English School, but using realist assumptions and definitions. Robert Gilpin was the first scholar to develop a theory of power transition, which he called hegemonic war theory. According to the theory, because of the anarchic nature of the state system and the subsequent goal of states to secure power, the world is characterized by hegemonic cycles. States "always seek to increase their wealth and power" until one state gets the upper hand and is able to exert control over the others. The power of one hegemon decreases when a challenger arises, and a hegemonic war ensues that allows the challenger to become the new hegemon. Instead of a constant anarchic system thriving toward balance, the international system is "a hierarchical ordering of states with a dominant or hegemonic power" that brings stability.⁶⁷

A.F.K. Organski complements Gilpin's account and similarly "takes exception" with balance of power theory's core prediction that the system will tend toward a balance. "This is not the case, at least since the industrial revolution," Organski argues. Instead, Europe has recurrently exemplified "combination[s] of overwhelming power," such as England, France and Russia in the 19th century, or the Western bloc during the Cold War. While there were some short-lived balances, they were usually rapidly eclipsed by "a growing giant." In fact, "if we look at the whole sweep of international history for the past 150 years, we find that balances of power are the exception, not the rule," Organski concludes.⁶⁸

However, much of the historical evidence provided by power transition theorists is not convincing. The hegemonic wars that Gilpin mentions as proof of the hegemonic cycles of history did not in fact serve to establish a new hegemon but helped avoid one. Gilpin himself

⁶⁷ Robert Gilpin, "The Theory of Hegemonic War," *Journal of Interdisciplinary History* 18.4 (Spring 1988): 593-596.

⁶⁸ Organski, *World Politics*, 290-292; see also A.F.K. Organiski and Jacek Kugler, *The War Ledger* (Chicago: University of Chicago Press, 1980), 14-17; Stephen Rock, *Why Peace Breaks Out: Great Power Rapprochement in Historical Perspective* (Chapel Hill: University of North Carolina Press, 1989), 6.

acknowledges that it is impossible to "forecast ... what the consequences [of a hegemonic war] will be." For example, he recognizes that the Treaty of Vienna that ended the Napoleonic Wars (which he calls a hegemonic war) resulted in "an equilibrium ... on the continent of Europe," not a new hegemony. Britain defeated the hegemonic ambitions of Napoleonic France and restored a balance over the continent, but the presence of other sizeable great powers on the continent and beyond prevented Britain from becoming a hegemon itself. What Gilpin calls a hegemonic war was in fact a tool used by Britain and European balancers to restore the balance of power.

Perhaps, as Jack Levy suggests, part of the problem lies in measurement. Because power transition theorists use mostly GDP and not military capability as a measure of power, they may persistently overestimate the power of some states that they term "hegemons," which are only great powers, and likely underestimate the power of other great powers that act as their restraints. For example, Organski claims that between 1815 and 1914 the dominant role was held by "a combination of overwhelming power" including England, France, Russia, and eventually the United States.⁶⁹ But because power transition looks mainly at GDP to calculate national strength, it "cannot capture Germany's power and influence in the European system" after 1871, which made imperial Germany the rising hegemon and was largely based "on the size, efficiency and leadership of the German army."⁷⁰ In the end, power transition theorists' claim that international systems tend more toward hegemony than balances, just like that of the English School, rests on loose definitions and imprecise evidence. More importantly, they focus on building an alternate

⁶⁹ Organski, *World Politics*, 291. This claim is problematic without even considering measurement issues because England, France and Russia hardly constituted a cohesive power unit throughout the century. Instead, they frequently experienced shaky relations and even open conflict. They eventually allied in the Triple Entente, a loose alignment not formalized until 1907, though, and at the time Russia had become a minor partner due to its domestic turmoil.

⁷⁰ Jack S. Levy, "Power Transition Theory and the Rise of China," in Robert Ross and Zhu Feng, eds., *China's Ascent: Power, Security, and the Future of International Politics* (Ithaca, NY: Cornell University Press, 2008), 18-19; 21-22.

theory that seeks to explain peace and war rather than balance of power and its shortcomings, and therefore do not offer a systematic assessment of the reasons balances may fail.

3. Balance of Power Failure as a Discrepancy within the Theory

Finally, a few scholars have come closer to the question I pose. They recognize that balance of power theory has strong explanatory power, evidenced by the fact that balances are by far the most common outcome in anarchy, but that balances sometimes fail without explanation. Their purpose, like mine, is to account for such puzzling failures while remaining within the larger realist framework.

In *Unanswered Threats*, Schweller is the first to formally identify the concept of balance of power failure. The previously mentioned groups of scholars just vaguely referred to the notion by saying balances did not materialize as the theory claimed. Schweller calls it “underbalancing,” which he defines as “instances ... where threatened countries have failed to recognize a clear and present danger, or, more typically, have simply not reacted to it or, more typically still, have responded in paltry and imprudent ways.” However, a first problem is that he does not focus on the outcome of balance of power failure itself but rather describes instances of late balancing, such as the delayed European responses to Wilhelmine Germany. Late balancing, as this thesis shows, may be one reason why hegemony succeed but it does not necessarily lead to their success, and it is certainly not the sole possible cause of their success. In the case of Wilhelmine Germany, for instance, even though European states underbalanced during Bismarck's tenure, they later recovered and successfully stopped imperial Germany from

becoming the European hegemon. Schweller's account of "underbalancing" is thus useful, but incomplete to explain balance of power failure.⁷¹

Still, Schweller's work is relevant for this study because it looks beyond the purely structural determinants of balance of power and offers a neo-classical realist approach consistent with this thesis. It discusses a number of domestic-political factors that may "thwart balancing behavior" and lead to underbalancing: elite disagreement about the threat and the response, elite division and fragmentation of the political system, lack of social cohesion among the population, and lack of legitimacy and weakness of the government. Nonetheless, Schweller contends that balancing failures are "primarily determined ... by the domestic political process" and stresses that "the theory offered [in *Unanswered Threats*] ... centers on the behavior and choices of policymaking elites." While he acknowledges that structural factors and other unit-level factors also influence states' balancing behavior and ultimately affect the establishment of a balance, his theory focuses on one particular aspect of the domestic political process. Schweller's account therefore helps us understand the individual decision-making process of states in reaction to threats, but remains incomplete in explaining balance of power failures in general because it does not specify the causal role of other possible explanations both at the structural and unit levels, such as geography or cognitive threat processing. In addition, Schweller's four domestic political factors operate by "frustrate[ing] the state's ability to extract resources for balancing purposes" and therefore "increase the risks and costs of internal balancing behavior," but it remains unclear how they impact external balancing efforts and alliance formation, the more common form of balancing. As Schweller concedes, his study is limited to assessing specific domestic-political

⁷¹ Schweller, *Unanswered Threats*, 1-2.

aspects of underbalancing and therefore cannot claim to provide a comprehensive analysis of balance of power failure.⁷²

Victoria Tin-bor Hui looks at ancient China and is puzzled by the fact that while Europe and the ancient Western system seemed to match the balance of power model perfectly, China and the East Asian system were simultaneously producing hegemons. How can we explain such opposite outcomes, especially from two regions that shared relatively comparable patterns of domestic organization and international interaction at the time, Hui asks? It would be "one-sided" to "presume that attempts at domination are necessarily checked by countervailing mechanisms," she writes. A main obstacle is that balancing efforts can be ruined by the collective action problems of the balancers. "Resistance to domination is a form of public good that entails costs to participants" and "those who pursue short-term self-interest will free-ride on others' contribution." Hui's contribution is particularly helpful because she is the first scholar to recognize that aspiring hegemons may play a critical role in their own success. She argues that one reason why balance of power may fail is that balancing efforts can be defeated by a counter-strategy of the rising hegemon. As Hui explains, "domination-seekers can overcome resistance and costs by pursuing counterbalancing strategies, ruthless stratagems, and self-strengthening reforms." While this thesis agrees with Hui that the success or failure of balance of power does not solely rest in the hands of the balancers but may also have a lot to do with the specific actions of the rising hegemon, it seeks to elaborate on Hui's discovery and pinpoint the specific impact of the actions of aspiring hegemons on balance of power failure. Overall, Hui transcends Schweller's analysis by offering multi-level explanations, combining structural and unit-level mechanisms, rather than solely domestic-political explanations. Yet although her explanations are useful and insightful, she focuses solely on a handful of factors and therefore falls short of a

⁷² Ibid., 9, 11-13, 16.

systematic examination of balance of power failures. In fact, she acknowledges that such is not her goal.⁷³

In an edited volume to which Hui contributed, William Wohlforth, Stuart Kaufman, and Richard Little push the analysis of balance of power failure still further. Like Hui, Wohlforth, Kaufman and Little find that "balance of power does not survive contact with non-European evidence," and the pre-modern world also presents instances where it fails. They identify a number of key factors that may enable the emergence of hegemons: technological advances can allow a rising hegemon to expand more efficiently and expansion in turn can generate economic returns that facilitate further expansion; promoting shared identities and cultural norms can help a rising hegemon attract followers; effective government and administration can boost the rise of hemegons while poor or rigid government can hurt balancing efforts; and collective action problems, as well as "uncertainty about the identity and severity of the hegemonic threat," can be exploited by the rising hegemon and may hinder balancing efforts. While they offer a broader and likely more complete account of why balances may fail, Wohlforth, Kaufman, and Little unfortunately neglect to spend time detailing the factors they identify and the processes through which they contribute to hegemony or impede balancing. As they concede, "we do not attempt to offer a complete theory that explains variation in the balancing propensity of systems."⁷⁴

The few studies of hegemony that have been undertaken thus offer helpful propositions about what may cause balance of power to fail, but they fall short of providing a systematic account of failures. Balance of power theory must go further toward building a general framework that focuses not only on a few possible explanations but comprises all possible causes

⁷³ Victoria Tin-bor Hui, *War and State Formation in Ancient China and Early Modern Europe* (New York: Cambridge University Press, 2005), 2, 12-13, 15-16, 23, 26.

⁷⁴ William Wohlforth, Stuart Kaufman, and Richard Little, "Introduction: Balance and Hierarchy in International Systems," in William Wohlforth, Stuart Kaufman, and Richard Little, eds., *The Balance of Power in World History* (New York: Palgrave MacMillan, 2007), 1, 4, 7-8, 12-19.

of hegemonic success. This is necessary in order to empirically assess which factors are key and supply a definite answer on why hegemons have occasionally arisen in a balance of power world. The next section attempts to devise such a framework.

2. Balance of Power Failure Mechanisms

To assemble a comprehensive picture of the causal mechanisms leading to balance of power failure, we must first clearly define the outcome of inquiry, and then identify the various explanatory variables that play a causal role in that outcome.

a. What Is Balance of Power Failure?

The purpose of this inquiry is to determine what causes the structural outcome of balance of power failure. How do we know that a system has achieved a balance of power failure? By definition, a system-wide failure of balance of power is measured by the emergence of a hegemon. In the language of balance of power theory, hegemony means balance of power failure. When an emerging hegemon reaches hegemonic status, then balance of power has failed. When an aspiring hegemon falls short of achieving hegemony and instead returns to being just a great power among others, then the balance of power has been restored. Asking what factors cause balance of power failures thus comes down to asking what factors enable hegemony to be established. The outcome of interest, or dependent variable, is therefore hegemony.

It is critical, then, to identify at what exact point a great power becomes a hegemon and the systemic power configuration shifts from a balance of power world to a hegemonic world. According to Robert Pape, balance of power worlds denote a system in which states can, in one way or another, balance the rise of the leading state and limit its actions, thus preserving their

security. Balance of power worlds range from multipolar systems, where there are several great powers that can independently or in group take on the leading state, to unipolar systems, where the leading state is so powerful that it requires the conjunction of all other powers to control it. Unipolar systems lie at the threshold of hegemony because the great gap in power between the leading state and the other states makes it very difficult to balance the leading state. But even in a unipolar system, balancing remains possible.

Once the system is hegemonic, however, balancing the leading state is impossible, even if all other states ally against it, because the disparity in power has become too great. As Pape explains, "the key boundary is between a balance of power system and a hegemonic one. This line separates a world in which the second-ranked powers can still act to preserve their security independent of the leading state from a world in which they cannot." There are different degrees of hegemony depending on the relative capabilities of the hegemon; but in all cases of hegemony, all other states face a life-threatening situation because they are not able to control the hegemon's power and restrict its actions. A hegemon is...

... capable of many harmful actions, such as rewriting the rules of international conduct to its long-term advantage, exploiting world economic resources for relative gain, imposing imperial rule on second-ranked powers, and even conquering any state in the system.⁷⁵

Therefore, the difference between a leading state that has reached hegemony and one that has not yet reached hegemony lies in the ability of the remaining powers to control the leading state and preserve their own security against it. Once the leading power is able to act with impunity and no other power in the world, not even with the help of large-scale alliances, is able to bar it from doing as it pleases, the leading power has become a hegemon. The stronger a rising hegemon

⁷⁵ Pape 13.

becomes, the more resources are required for the remaining states to restore the balance of power.

How can we measure with precision when a state has reached the stage of hegemony, where balancing is no longer possible and the balance of power has failed? Power status in international relations is reflected by the control and influence one state can exert over others.⁷⁶ A state is considered a great power when it has the “ability to contend in war ... with any other country in the system,” including the most powerful one.⁷⁷ A great power becomes a hegemon when it has the ability to contend against *all* other powers in the system.

Hegemons exercise control over the system in different ways. Some hegemons, particularly in pre-modern times, simply invaded all other states, allowing for easy measurement of hegemony. When a rising hegemon uses conquest to expand, it reaches hegemony when it has taken over a majority of countries in the system. Measuring the capacity to control may appear more difficult when an aspiring hegemon grows in power but does not invade the system, for example when it exerts control by expanding its political and economic influence. Military invasion is not a requirement to become a hegemon, and a number of alternate, subtler means to acquire control over the system exist for aspiring hegemons that prefer to grow without resorting to conquest. They include tools such as economic and financial pressure, or limited military coercion based on the hegemon's unsurpassed military capabilities, for example, or even nuclear coercion, to lead other states to behave as the rising hegemon wishes. Because rising hegemons may not always use conquest on their way to hegemony, the most accurate way to measure hegemony is to assess the rising hegemon's *ability* to defeat any and all other states of the system combined, even if it does not actually do so. Assessing such ability may include "estimat[ing] ...

⁷⁶ Joseph Nye, *Understanding International Conflicts* (New York: Pearson Longman, 5th edition, 2005), 59.

⁷⁷ Robert Ross, “Bipolarity and Balancing in East Asia,” in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power. Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004), 268, 272.

the relative war-fighting capabilities of the contenders,”⁷⁸ as well as examining the rising hegemon's propensity to obtain the behavior it seeks from other states in the system. In cases where conquest is a rising hegemon's primary tool of expansion, the extent of conquest may be used as a shortcut.

b. What Explains Balance of Power Failure?

Now that the outcome we seek to explain, hegemony, has been clarified, we must look at what makes it possible. What causes the configuration of power in a given system to become hegemonic as opposed to balanced? This thesis lists twenty-one factors, grouped by affinity in six categories, that may contribute to the successful establishment of a hegemony. They combine structural and unit-level factors, in line with the thesis's neoclassical realist approach and effort to provide a comprehensive explanation for balance of power failure. The structural factors reflect the anarchical constraints of the system that operate on states and condition their actions. The unit-level factors highlight the domestic and cognitive dispositions of states that affect their ability to process those environmental constraints and to extract the resources they need to affect the international distribution of power.

The twenty-one factors highlight both the role of the potential balancers and that of the emerging hegemon. The actions of the potential balancers—how many balance and how many do not, and whether the most powerful potential balancers balance or not—necessarily have a direct impact on the eventual establishment of a hegemony, so some variables must explain why they may not balance. But while the ultimate power configuration—hegemony or balance of power—thus depends partly on the success or failure of balancing movements undertaken to stop

⁷⁸ Ibid., 272.

a rising hegemon, it also depends on the achievements of the rising hegemon itself. The rising hegemon can impede the balancers' actions, or it may have exceptional skills and develop innovative and efficient ways to extract and use resources that the balancers are unable to match or counteract. Therefore, other variables must account for the unique skills of the rising hegemon. The inclusion of rising hegemon-related variables is crucial because we seek to explain a structural power configuration, hegemony or balance of power, and not just the individual state action of balancing or not balancing. Whether and why individual states may fail to balance is part of the explanation but the structural equation remains incomplete if the role of the rising hegemon is omitted. Hegemony is thus not solely about failed balancing, but also about the skillfulness of the rising hegemon.

Victoria Tin-Bor Hui was the first scholar to stress that the rising hegemon's own actions may play a critical role in enabling the establishment of a hegemony, by undermining the balancers and engaging in dramatic self-strengthening that can outpace the balancers.⁷⁹ In fact, the literature overwhelmingly credits the structural outcome—hegemony or balance of power—to the balancers alone. The neorealist account of balance of power theory, for instance, looks exclusively at the balancers' side to explain why balances are reestablished. When an aspiring hegemon arises, the theory assumes that if it is left alone, it will become an actual hegemon (Table 2.1). That is because the neorealist explanation focuses exclusively on structural factors, and therefore does not assess the domestic characteristics of the rising hegemon. In reality, the rising hegemon's ability to extract and use resources efficiently, which are unit-level attributes, matter greatly in determining whether or not it will succeed in its hegemonic bid. A potential hegemon might not succeed in becoming a hegemon if it does not possess considerably superior

⁷⁹ Hui 101.

skills, at least militarily and economically, and if it is unable to sustain its growth, whether other states actively try to stop it or not.

Table 2.1: Structural Outcome (Neorealist Approach)

<i>Determining Factor (Independent Variable):</i> BALANCING MOVEMENT	<i>Outcome (Dependent Variable):</i> CONFIGURATION OF POWER
Strong	Balance of Power
Weak/Absent	Hegemony

By investigating the unit level, this thesis thus provides another important contribution to the balance of power analysis and addresses the role of the rising hegemon, the second side of the balance of power equation that has traditionally been neglected. Because it acknowledges two sets of actors involved in causing hegemonic or balanced outcomes, the potential balancers and the rising hegemon, this approach highlights that every strong balancing movement does not automatically result in balance of power and that every weak balancing movement does not necessarily lead to hegemony (Table 2.2). If a rising hegemon has uniquely superior skills, even strong balancing attempts may not be enough to stop it. Symmetrically, a lack of sufficiently superior skills may be a reason for many balance of power successes, which are often credited solely to strong balancing movements. Whether balance of power or hegemony ultimately triumphs, then, may depend on the balance of skills and capabilities between the rising hegemon and the balancers.

Table 2.2: Structural Outcome (Multi-Level Approach)

<i>Determining Factors (Independent Variables):</i>		<i>Outcome (Dependent Variable):</i>
BALANCING MOVEMENT	HEGEMONIC SKILLS	CONFIGURATION OF POWER
Strong	Weak/Average	Balance of Power
Strong	Strong	Indeterminate/Depends on Balance of Skills
Weak/Absent	Weak/Average	Indeterminate/Depends on Balance of Skills
Weak/Absent	Strong	Hegemony

The twenty-one causal mechanisms of balance of power failure this thesis identifies consequently operate either by weakening the balancing movement, or by reinforcing the rising hegemon's assets (Figure 2.1). Each factor highlights the responsibility of one of the two sides.

Figure 2.1: Balance of Power Failure Mechanisms

DV = Dependent Variable (outcome of interest)

IV = Independent Variable (explanatory factor)

Italics = Side responsible for Variable

IV1] Communication Problems

[IV1.1] Physical Difficulty (*potential balancers*)

[IV1.2] Misperception (*potential balancers*)

[IV1.3] Deliberate Deception (*rising hegemon*)

Factors related specifically to the potential balancers' side:

[IV2] Collective Inaction →

→ [IV2.1] Communication Problems

[IV2.2] Lack of Trust

[IV2.3] Lack of Sufficient Interest

[IV2.4] Buckpassing

[IV3] Laggard Balancing →

→ [IV3.1] Communication Problems

[IV3.2] Lack of Trust

[IV3.3] Lack of Sufficient Interest

[IV3.4] Buckpassing

[IV4] Bandwagoning →

[IV4.1] Fear

[IV4.2] Profit

[IV4.3] Combination

**[DV] Balance
of Power
Failure/
Hegemony**

Factors related specifically to the rising hegemon's side:

[IV5] Military Innovation →

[IV5.1] Weaponry

[IV5.2] Strategy

[IV5.3] Organization

[IV6] Unique Non-Military Skills →

[IV6.1] Geography/Geopolitics

[IV6.2] Economic Advantage

[IV6.3] Political Organization

[IV6.4] Social Advantage

1. Communication Problems [IV1]

Balance of power can fail because of the potential balancers' communication problems. If potential balancers lack information, or possess incorrect information, about the rising hegemon's potential, they will be unable to react and balance properly. Imperfect information may cause potential balancers to remain unaware of the growth and threat of the rising hegemon, or to mistakenly gauge the danger as less imminent than it actually is. Those states may then choose not to balance or to delay balancing, when they would have chosen to balance had they possessed the right information. Imperfect information stems from poor communication, which can occur for three different reasons.

First, communication may be hampered by physical hurdles [IV1.1]. Distance and levels of technology play a crucial role in enabling information to circulate. If the rising hegemon is far and communication technology is primitive, information flows will be slow and potential balancers may not realize the rising hegemon's threat in time to balance. Only immediate neighbors may be aware of the danger, and may be unable to warn more powerful, more distant potential balancers in a timely fashion. Likewise, if a potential balancer is isolated, it may be unaware of the existence of a collective balancing effort by other balancers and may thus fail to join it. The absence of diplomatic channels of communication or the lack of regular diplomatic exchanges may present another physical hurdle for balancers, leading them to rely on ad-hoc or non-official messengers to convey information about the rising hegemon that may as a result be inaccurate and undervalue its real threat. One can suspect that physical communication difficulty is a factor that may have played a particularly critical role in hindering balancing in pre-modern cases, when information traveled via horse messengers and boats and could take months or even years to reach the opposite side of a continent, but is becoming increasingly less relevant due to

dramatic advances in communications technology and transportation. Although it is virtually impossible in the modern world that states would fail to balance a rising hegemon because they had not heard about its powerful capabilities, this factor cannot be discarded if we want to preserve an historically accurate account hegemony, because it may have played a considerable role in preventing successful balancing in the past.

But communication problems are not entirely an issue of the past. While physical hindrances to gathering information have become increasingly rare, perceptual barriers to the acquisition of information remain widespread [IV1.2]. Robert Jervis and psychological theorists of international relations point out that man often wrongly interprets information. Because of an unconscious tendency to cling to our own pre-conceived notions of how the world works, we often fail to process information that others send us accurately, leading us to misperceive their intentions.⁸⁰ The human psychology is hardwired to distort incoming information, and state leaders may therefore fail to correctly interpret warning signs about the rise of a new hegemon, or mistakenly view the rising hegemon's intentions as benign, until it is too late to balance. Potential balancers may as a result neglect offers to cooperate in balancing, missing an opportunity to check the rising hegemon. Misperception can thus be an important factor that leads to balance of power failure.

While misperception is usually subconscious and not deliberate, poor communication can also be calculated. A potential balancer may fail to balance because it is intentionally misled about the threat of the rising hegemon [IV1.3]. Potential balancers are responsible for their physical difficulties to gather information and their possible misperception of incoming information, but the rising hegemon can also tamper with their efforts to acquire correct

⁸⁰ See Robert Jervis, *Perception and Misperception in International Politics*" (Princeton, NJ: Princeton University Press, 1976).

information by deceiving them about its rise and threat. States frequently use deception to influence outcomes. A rising hegemon can purposely play down its offensive posture and power advantage in order to give potential balancers a false sense of security. It thus shields its rise until it has achieved an absolute lead. Arnd Plagge demonstrates that a rising power has an incentive to fool its rivals and hide its growing power until the gap in power between them has become significant. Cloaking its power will allow it to discourage preemptive attacks by its rivals until it is too late and the rising power has already outpaced them.⁸¹ Besides disguising its real power, the rising hegemon's deceptive efforts can also involve bullying or intimidating potential balancers, especially weaker great powers or secondary powers that may wish to join the balancing side, so they do not dare to balance and remain neutral or even accommodate the rising hegemon instead.⁸² Deceptive communication can therefore lead to balance of power failure.

Communication problems can cause balance of power failure both directly or indirectly. Physical, perceptual, and deliberate hindrances to communication can operate to trigger balance of power failures directly when they affect states' individual decisions to balance, as the three factors above [IV1.1; IV1.2; IV1.3] highlight. They can affect a state's decision to balance internally and increase its own capabilities, or they can affect a state's decision to balance externally and join a balancing coalition. If a number of potential balancers are affected by the communication problems and therefore fail to balance, in aggregate the balancing response to the rising power will be weak and hegemony may ensue. But communication hurdles can also operate indirectly by affecting the effectiveness of balancing coalitions rather than states'

⁸¹ Arnd Plagge, "Patterns of Deception: Why and How Rising States Cloak their Power," paper presented at Yale University's International Security Studies Colloquium in International History and Security (Sept. 29, 2009), 6-8, 36-7.

⁸² This scenario is different from bandwagoning [IV4] because bandwagoning consists in voluntarily joining the rising hegemon's side, not being bullied into joining it, even if the voluntary decision might be motivated by the potential balancer's calculation that it may eventually be hurt by the rising hegemon's power.

individual decisions to balance. Coordinating balancing efforts is much harder when facing physical, perceptual, or deceptive communication problems. Such problems can easily derail a common balancing effort. The indirect effects of communication problems on the workings of alliances are appraised as separate factors. They may result either in collective inaction (no cooperative balancing, codified as [IV2.1], detailed below) or laggard balancing (some cooperative balancing, but too late or not sufficiently, codified as [IV3.1], detailed below). Collective inaction and laggard balancing are, in turn, direct causes of external balance of power failure. Communication problems are thus simultaneously a direct cause of internal and external balance of power failure by affecting states' own decisions to balance [IV1.1; IV1.2; IV1.3] and an indirect cause of external balance of power failure by hindering the workings of balancing alliances [IV2.1; IV3.1].

2. Collective Inaction [IV2] ⁸³

Probably the most obvious cause of balance of power failure resides in collective inaction, also commonly called collective action failure. Collective inaction has to do with the potential balancers; it occurs when they are unable to balance together and no external balancing therefore takes place. In many cases internal balancing may not be an option, because no individual states are willing, or powerful enough, to take the risk of warding off the aspiring hegemon by themselves. External balancing, or the creation of a balancing coalition, is then the only way to stop the rising hegemon. If potential balancers powers are incapable of

⁸³ While some scholars have mentioned the factors in IVs 2, 3, and 4 individually, no study has so far systematically combined and articulated them. Randall Schweller has argued that states bandwagon more frequently than realists usually admit when facing growing power. See Randall Schweller, "Bandwagoning for Profit: Bringing the Revisionist State Back In," *International Security* 19.1 (Spring 1994): 75. Mearsheimer has argued, on the other hand, that great powers always prefer to pass the buck than to balance whenever the option exists, but that this is dangerous and can lead to hegemony if the buck catcher fails to balance. See Mearsheimer, 160-161, 270.

collaborating, however, and no collective balancing occurs, the only chance of curtailing the rising hegemon and restoring the balance of power vanishes. When there is no internal balancing, collective inaction results in the absence of a balancing movement, which can directly lead to balance of power failure. The occurrence of hegemony is then entirely in the hands of the rising hegemon. Collective inaction is thus a situation where states are incapable of allying and cooperating to balance a rising hegemon, although collaboration would be in their best interest. Why would potential balancers fail to cooperate when their survival is at stake?

First, as mentioned above, communication problems can cause collective inaction, and thus, indirectly lead to balance of power failure [IV2.1]. Because of distance and poor communication technology, potential balancers may not be aware that they share similar fates and may not know of other potential balancers' willingness to cooperate. As a result, they may remain isolated instead of forming a coalition against the rising hegemon. The absence of diplomatic channels and regular official exchanges may also prevent them from joining forces. In addition, a potential ally may misperceive the intentions of its proposed partner, wrongly assuming that the partner does not plan to get fully involved and share the costs, for example, and, as a result, refuse to form the balancing coalition. Potential balancers can also be deliberately misled by the rising hegemon about a balancing coalition and driven to reject cooperation. For example, the rising hegemon can play the divide-and-conquer card and try to sabotage an opposing coalition by unobtrusively shedding doubts on the coalition's viability, raising questions about its members' reliability and their capabilities, or evoking the possibility that some might leave the coalition and abandon their partners. It is also possible that members of the coalition themselves could deceive other members into believing that they have fewer means than they really have, for instance, in order to contribute less to the common effort and

save resources, leading to a breakdown of the coalition.⁸⁴ As these scenarios show, communication deficiencies can lead to collective inaction in the face of a rising hegemon.

Cooperation among potential balancers can also fail due to issues of trust [IV2.2]. The anarchical context of international relations impedes cooperation among states, even when they share a similar goal, such as stopping a rising hegemon. As Robert Jervis explains, "although actors may know that they seek a common goal, they may not be able to reach it."⁸⁵ States fear being cheated by their allies, who might extract relative gains out of the collaboration and become stronger and threatening in the future.⁸⁶ Because states are concerned about others' defection and fear being left with the sucker's payoff if the others skip, they often wind up not cooperating at all, despite that cooperating against the rising hegemon is vital. Trust concerns result in collective inaction. Distrust-related collective action failure can occur with regard to any type of international cooperation, but state suspicion about other states' intentions is even greater when a state's essential security is at stake as is the case when balancing a rising hegemon. Getting the sucker's payoff means balancing the stronger enemy by oneself and likely taking a bad defeat. The prospect of going to war together in an effort to balance a rising hegemon is wholly a different matter than economic cooperation, for example. Trust issues become more salient when the stakes rise, and as a consequence, collective action is much harder to achieve. Thus, when confronted with a rising hegemon, states will hesitate to collaborate with neighbors and other states with whom they have a history of conflict, focusing on short term danger and opportunities while ignoring the more menacing, long term threat posed by the rising hegemon.

⁸⁴ This scenario is different from buckpassing [IV2.4; IV3.4] because buckpassing does not involve deceiving one's partner and tricking it into believing one will participate when it is not the case; buckpassing simply refers to letting someone else do the job.

⁸⁵ Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics* 30.2 (January 1978), 167-70.

⁸⁶ Joseph Grieco, Robert Powell, and Duncan Snidal, "The Relative Gains Problem for International Cooperation," *The American Political Science Review* 87.3 (Sept. 1993), 729.

Lack of trust-induced collective inaction is a factor that operates at the unit-level since it is about states' decisions to cooperate. It is not unit-driven, however, but is instead a system-effect observed at the unit level. Trust issues and immediate gains concerns derive from structural anarchy, which breeds mistrust and "encourages behavior that leaves all concerned [i.e., the potential balancers] worse off than they could be." It is the "lack of an international sovereign," the structural feature of international politics, that allows for cooperation to "bring disaster" and stimulates states' suspicions about the intentions of others.⁸⁷ States should be able to overcome the mistrust inherent to anarchy and balance jointly when their security and survival is at stake, as is the case when they face a rising hegemon, but the existence of cases of hegemony suggests that they may not always do so and that the trust issues may prevail over the urge to balance.

If neither trust nor communication is an issue, states can still end up with collective inaction if they lack a sufficient shared interest in a joint balancing endeavor [IV2.3]. When a number of potential balancers face a rising hegemon, they do not necessarily have the same interest in balancing it. Some states may perceive that the rising hegemon's threat does not affect all remaining powers equally, and that the threat is more imminent for some states than for others. While in reality, once a hegemon exists, each state's survival is threatened in the same way, states may *perceive* the extent of that threat differently. Threat perception depends significantly on power differentials.⁸⁸ The most powerful potential balancers may feel that their survival is less in jeopardy because they possess more capabilities to defend themselves and may therefore hesitate to join a balancing coalition, for example. But if the major great powers are indecisive, a coalition might not be formed at all because it would lack means. Threat perception

⁸⁷ Jervis, "Cooperation," 167.

⁸⁸ Walt, "Alliance Formation," 8-10.

also depends on more tangible assets such as geography. As Henry Kissinger demonstrates, differences in geographic positions can lead to opposite foreign policies. A continental power, because of its confined position, has an acute threat perception and may therefore be more likely to join a balancing coalition. In contrast, there is a certain "consciousness of safety conferred by an insular position." Being more shielded from direct attack an insular power will probably perceive the threat of the rising hegemon as more distant, and may shy away from joining a balancing coalition.⁸⁹ Not all potential balancers thus share the same interests in responding to the threat of a rising hegemon. Those that are more distant or more powerful, for example, may decide that they do not need to join in a balancing movement.

Finally, collective inaction can occur because potential balancers pass each other the buck [IV2.4]. Since it is both costly and risky to balance against a rising hegemon, states are often tempted to let others do the job if possible. If there is only one great power in the system, passing the buck is not an option for it because its participation is required for balancing to succeed, but buckpassing can occur in any system with more than one great power. Buckpassers choose to not balance, just like in [IV2.3], but their motivation for not balancing is not a lack of perceived interest. Instead, buckpassers are conscious of the imminent danger posed to them by the rising hegemon and believe immediate steps need to be taken to stop it. But they think that other states, perhaps stronger great powers or the ones ready for a possible war, will balance and save them the effort. They understand that balancing is a public good and that once it is done, all potential balancers benefit from the result the same way, whether they participated in the effort or not. If only a few states pass the buck and there is a sufficient number of potential balancers remaining to take on the rising hegemon, the balance could still succeed. But if all potential balancers pass

⁸⁹ Henry Kissinger, *A World Restored: Metternich, Castlereagh and the Problems of Peace 1812-1822* (London: Victor Gollancz, 1973), 285.

the buck amongst themselves, and especially if all great power balancers do, collective inaction will ensue, and the hegemon will be able to rise unchecked.

3. Laggard Balancing [IV3]

While the potential balancers end up not cooperating at all in a case of collective inaction, they do cooperate in a case of laggard balancing, albeit too late, too slowly, or too inefficiently to allow the balancing to succeed. The two outcomes are drastically different. In the case of collective inaction, there is no cooperation; in the case of laggard balancing, there is some cooperation but it is insufficient. The latter case captures the idea that potential balancers are sometimes sluggish in their reactions, which allows the rising hegemon time to rise peacefully and widen its power advantage over them. Once they start to balance adequately, the rising hegemon may already be too far ahead to be stopped. Laggard balancing and collective inaction occur for the same reasons because both describe dysfunctional cooperation.

Just as communication problems can prevent collective balancing from occurring altogether, they can also slow a balancing coalition or render it ineffective [IV3.1]. Communication is a crucial element of any joint military campaign, for instance, which requires the arrangement of each participant's role and the assignment of leadership responsibilities ahead of time, and if it is deficient may undermine the whole effort. Distance and poor communication and transportation technology delay the transmission of information among allies and can thus slow a collective balancing movement or impede its effectiveness. The absence of formal diplomatic channels may also prevent the potential balancers from coordinating their efforts well. Similarly, misperception of an ally's intentions or deliberate deception, whether by another balancer or the rising hegemon, can severely strain the effectiveness of a balancing coalition. If it

does not destroy the balancing coalition and result in collective inaction, it will at least delay the balancing, possibly enough to allow hegemony.

Trust issues can also hinder a balancing coalition and reduce its efficiency [IV3.2]. Because there is no guarantee of commitments in anarchy, mistrust is endemic to international interaction and allies in a balancing coalition are constantly suspicious of their partners and their future behavior. Balancers worry that they their coalition partners may abandon them and let them bear the brunt of the costs, or even that they may renege their commitments and turn against them, worsening their security outlook. As a result they may not commit wholeheartedly to the balancing coalition and may refuse to take risks necessary to make it work. Since balancers are on the weaker side to begin with, slow and overly cautious commitments to the balancing effort will most likely be insufficient to stop a rising hegemon.

The same can happen if states perceive their interests in balancing the rising hegemon to be uneven [IV3.3]. States that fail to capture the urgency of the threat of a rising hegemon may partake in a balancing coalition but not commit all the resources they can to hedge their bets. But by sitting on the sidelines while others do the work or participating only partially, they weaken the balancing coalition and endanger its success.

Likewise, buckpassing can lead to laggard balancing [IV3.4]. If a few but not all states pass the buck, there will be some balancing instead of collective inaction, but it may not be sufficient to stop the rising hegemon. When a few potential balancers free ride and let others balance, the balancers will not benefit from their resources, and the strength of the balancing coalition will be affected. Without those added resources, the balancing effort will be less efficient, and might be insufficient to prevent hegemony.

But unlike collective inaction, laggard balancing does not always lead to balance of power failure. When there is collective inaction and internal balancing is not an option, balance of power can only fail, because no one acts to check the rising hegemon. When potential balancers engage in laggard balancing, however, balance of power may be preserved if they recover and start balancing with full force before the rising hegemon reaches hegemony and balancing becomes impossible. The effort required for balancing to be successful after a phase of laggard balancing, however, is significantly greater because the rising hegemon has accrued power in the meantime. Of course, if the balancers continue to balance laggardly and do not improve the pace and quality of their effort, success in restoring the balance is unlikely. Throughout the 1930s, France and Britain were laggard balancers and passed the buck amongst each other, but eventually in 1939 they went to war to defeat Germany. They succeeded because they had already individually started rearming, and because they eventually benefited from the help of an extra-regional great power, the United States. In other circumstances it may be too late, once laggard balancers start balancing adequately, to catch up with and stop the rising hegemon. The success of late balancing will depend on the relative capabilities amassed by the aspiring hegemon during the potential balancers' hesitation period. The greater the gap in power between the aspiring hegemon and the next strongest powers, the harder it will be for those powers, after having been laggards, to balance successfully.

4. Bandwagoning [IV4]

Bandwagoning with the rising hegemon is a counterproductive strategy that may lead to balance of power failure if undertaken by a sufficient number of potential balancers. Bandwagoning is the opposite of balancing. While in the case of collective inaction and laggard

balancing, states seek to balance the rising hegemon but are unable or unwilling to work together in order to do so, in the case of bandwagoning, states seek not to oppose the threatening side but to join it for various reasons. Bandwagoning occurs when a potential balancer voluntarily allies with the aspiring hegemon instead of opposing it. Bandwagoners who join the stronger side in effect add their power to the rising hegemon's instead of contributing it to the weaker, balancing movement. If the most powerful potential balancer bandwagons, or if several major potential balancers bandwagon simultaneously, the hegemon can rise virtually unchecked. Even if only one or a few potential balancers bandwagon while the others balance, the defection of those few states to the stronger side can easily doom the balancers by adding considerable weight and resources to the already stronger rising hegemon, resulting in balance of power failure.

There are two reasons why states might bandwagon, even if it threatens the balancing effort and reinforces the rising hegemon's growth. First, they can choose to bandwagon out of fear [IV4.1]. The bandwagoner is well aware of the danger posed by the rising hegemon and evaluates the threat correctly. It knows that as the strongest power, the rising hegemon will likely prevail. The bandwagoner therefore reasons that it is better to ally with the future hegemon not only to avoid defeat, but also in the hope that the future hegemon will later spare its ally and act benevolently towards it in return. As Walt states, "bandwagoning is essentially a form of appeasement: ... by bandwagoning, threatened states hope to convince a threatening power that they are in fact loyal supporters, so that it will leave them alone."⁹⁰ Bandwagoners think that aligning themselves with the rising hegemon is the only way to preserve themselves. Secondary powers especially may also be in a situation where other potential balancers pose a greater immediate threat to them than the rising hegemon. In that case, their fear of those balancers may lead them to join the rising hegemon rather than risk allying with the balancers.

⁹⁰ Walt, *Taming*, 183; Walt, "Alliance Formation," 7-8.

Other states choose to bandwagon for profit [IV4.2]. They similarly understand that the rising power constitutes the stronger side that is likely to prevail, but join it in order to share the spoils of victory. The rising hegemon might help them acquire territorial gains, for example. Profit-driven bandwagoners follow the creed that “nothing succeeds like success,” as Randall Schweller explains. Of course, a state’s motivation for bandwagoning could be a combination of both fear and profit [IV4.3]. It may reason that siding with the rising hegemon would bring it both security and gains shared with the victor. Alliance choices, as Schweller concludes, can be motivated “by appetite as well as fear.”⁹¹

Bandwagoning is a highly dangerous gamble, however, because once a great power has reached hegemony, it is unstoppable, and the bandwagoner cannot be assured that the hegemon will not eventually turn against it (and it is even less assured that the hegemon will consent to share the spoils of victory). If the rising power achieves hegemony, the bandwagoner, as well as all other states, will lose the ability to act independently, and will therefore be unable to defend itself. Due to the great long-term risk of bandwagoning, the realist scholarship generally agrees that bandwagoning is a rare behavior.⁹² In addition, just like laggard balancing, its occurrence does not always lead to balance of power failure. Because balance of power is a systemic outcome, even if a few states bandwagon with an aspiring hegemon, it will not necessarily prevent the balance of power from eventually being restored, because bandwagoning may be only temporary or pursued by a few states.⁹³ Just before World War II, for example, although the

⁹¹ Randall Schweller, "Bandwagoning," 73.

⁹² Waltz, *Theory*, 126.

⁹³ Balancers may even engage in defensive "wedge" strategies to prevent other potential balancers from bandwagoning, break up a bandwagoner's commitment to the rising hegemon, or attract a bandwagoner into a balancing coalition. See Timothy W. Crawford, "Wedge Strategy, Balancing, and the Deviant Case of Spain, 1940-1941," *Security Studies* 17.1 (January 2008), 1-38.

Soviet Union bandwagoned with Nazi Germany, the balance was eventually restored and Germany was defeated as the Soviet Union switched to the balancing coalition.

But bandwagoning, while perhaps not always sufficient to prevent the balancing of the rising hegemon, may be underestimated by realist scholars. In fact, some realist scholars point out that bandwagoning is more common than the literature traditionally assessed.⁹⁴ If it is sustained, bandwagoning takes away capabilities from the balancing side to enhance the strength of the aspiring hegemon, especially when one or several great powers team up with the already stronger rising hegemon. It may thus seriously weaken balancing efforts and tip the balance in favor of the threatening side, contributing to the rising hegemon's success. Extended or large-scale bandwagoning, in particular, is sure to doom balancing efforts.

5. Rising Hegemon's Military Achievements/Innovation [IV5]

The rising hegemon can have a direct impact on its success if it develops the ability to generate and use military resources more efficiently than the potential balancers, and the potential balancers are unable to match or counteract those achievements. The rising hegemon's superior military skills will then help increase its advantage in military power relative to the balancers and enable it to attain hegemonic status. The rising hegemon's military power will leap ahead of the potential balancers if it develops an innovative monopoly in the military arena, particularly if the potential balancers are unable to adapt to the innovation, by either copying it, developing a counter-innovation to neutralize it, or creating a substitute of similar value. Such monopolies can stem either from a brand new invention or less drastically from the rising hegemon's blending of existing knowledge from various sources or from its own improvements upon an existing device or technique.

⁹⁴ Schweller, "Bandwagoning," 79.

There are three areas of the military where the rising hegemon can exercise superior skills and gain efficiency at the expense of its competitors. The first, and most obvious, is technology and weaponry [IV5.1]. The creation of a new weapon or type of weapon, like siege engines or battleships, for instance, can give the rising hegemon a critical edge. The discovery of new resources or a new technology, such as iron smelting, gun powder, or nuclear fission, can similarly affect the distribution of military capabilities.⁹⁵ Military power is not solely based on technology and material superiority, however. Strategies and tactics, which are essentially the ways in which states use existing material resources and technologies, can also give one actor a significant advance [IV5.2]. Stephen Biddle argues that force employment, the combination of a state's non-material military assets such as strategies and tactics, doctrine, morale, and leadership, is the "most influential" aspect of military capability and can largely determine a state's ability to exercise territorial control in war. Material factors alone do not accurately predict war outcomes. Nevertheless, "international relations theorists mostly ignore force employment. Many simply assume that states will use materiel 'optimally,' hence the materiel itself is the only important variable."⁹⁶ Therefore, if a rising hegemon develops new, particularly efficient ways to use its material resources, both at the strategic level in campaign planning and at the tactical level in specific operations, it may gain a decisive advantage even if the resources themselves may not be revolutionary. This is especially the case if potential balancers are unable to imitate or find ways to offset those advances in strategy and tactics. Finally, a rising hegemon's rationalization of its military organization can dramatically increase its military instruments' efficiency and as a result give it a significant edge over the potential balancers

⁹⁵ For a historical analysis of key innovations in weapons and weapons systems, see William H. McNeill, *The Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000* (Chicago: University of Chicago Press, 1982).

⁹⁶ Stephen Biddle, *Military Power: Explaining Victory and Defeat in Modern Battle* (Princeton, NJ: Princeton University Press, 2004), 17-19, 190-91.

[IV5.3]. Organizational innovations may include changes in areas as diverse as the hierarchy and chain of command, the arrangements within combat units, logistics and supplies, and combat support services like transportation, facilities and medical assistance. If the balancers are unable to identify and duplicate organizational restructuring that dramatically improves the efficiency of the rising hegemon's armed forces, those innovations may enable the rising hegemon to permanently surpass them.

Fortunately for potential balancers, learning is endemic to the anarchic international system. According to Waltz, because military innovation offers a boost in relative power to the innovator, and a serious security threat to whomever does not innovate, states naturally copy other's achievements.⁹⁷ Key military innovations such as Napoleon's levee en masse and the Prussian general staff system were quickly adopted throughout Europe and beyond. However, not all military innovations are easily copied or countered, and some military secrets are well guarded. A number of domestic issues may inhibit adaptation, just as they may hinder innovation in the first place, such as bureaucratic inertia, centralization, or specialization.⁹⁸ Any military improvement can be countered in time, but time may be the crucial issue. It may take a while for potential balancers to adapt to the rising hegemon's achievements, and that time can be sufficient for the rising hegemon to overcome them and reach hegemony.

6. Rising Hegemon's Unique Non-Military Achievements/Skills [IV6]

Superior military capabilities are necessary to become a hegemon, but the rising hegemon may also benefit from additional, non-military achievements and attributes that contribute to its success and that balancers are unable to match. In order to rule over the system, in fact, it is quite

⁹⁷ Waltz, *Theory*, 104.

⁹⁸ Stephen Peter Rosen, *Winning the Next War: Innovation and the Modern Military* (Ithaca, NY: Cornell University Press, 1991), 3-6, 105, 128-9.

likely that the rising hegemon will have or develop special, unique traits that sets it apart. While its military power may be sufficient to give the rising hegemon the formal territorial control, or at least the ability to acquire such control, that it needs to reach hegemony, other skills and characteristics may facilitate and even critically reinforce that control. The rising hegemon must also be able to maintain the control established by its military supremacy and fuel and sustain its leadership to succeed as a hegemon, and military instruments may be inadequate for that purpose. In fact, the organization of peace to sustain expansion and transform military control into political rule requires a different set of skills than just military superiority. However, whether those special abilities and characteristics make a difference in the rising hegemon's path to hegemony depends on the potential balancers' ability to match them or adapt to them and neutralize the advantage they procure to the rising hegemon.

The non-military advantages that can contribute to an aspiring hegemon's success can be manifold, and can be regrouped in a several categories. First, geographic advantages can help tilt the balance in favor of the rising hegemon [IV6.1]. The rising hegemon's position in the system can affect its ascent. A central position will grant it easy access to all parts of the system and may thus enhance its ability to exercise control. However, a peripheral location may also be advantageous because it isolates the rising hegemon from the system's main areas of activity and conflict and may thus shield its early rise until it has already developed a significant advance over competitors. For example, the rising hegemon could be protected by high mountains or an ocean, rendering it less accessible and less visible and discouraging balancers. Christopher Chase-Dunn and other world systems theorists argue that a peripheral or semi-peripheral position also encourages innovation because actors on the outskirts of a system are more willing to invest in novelty and are greater risk-takers than the more static, long-established societies that

constitute the core of a system.⁹⁹ Moreover, states with extensive sea access traditionally identified as maritime powers may have an advantage over land powers, particularly in regions laced with waterways. Not only their naval capabilities allow them to project power around the region more efficiently than land powers, but their expansion generally appears less threatening and is therefore more tolerated than that of land-based powers because maritime strength is often associated with trade rather than territorial encroachment.¹⁰⁰ In addition to position, land composition is another geographic asset that can affect the rising hegemon's ascent. States are endowed with contrasting natural resources, such as soil fertility, irrigation opportunities, climate, and raw materials, that will impact their economic diversity and growth and therefore the reservoir of power they can rely on to fuel their rise.

In fact, a rising hegemon's innovative economic choices constitute the second non-military advantage that may play a large role in its eventual success [IV6.2]. Because economic assets stimulate military growth and thus overall power increases, the rising hegemon may gain a decisive edge if it is able to generate substantial economic growth that the potential balancers cannot match. The rising hegemon's economic policies can be innovative in many ways, for example, by developing new instruments to spur domestic production, acquiring an exceptionally prominent commercial position, securing its supply lanes and becoming immune to economic pressure from the balancers, or pioneering infrastructural improvements that reinforce its economic posture. Third, the rising hegemon's advance can be bolstered by its innovative state-building abilities [IV6.3]. It may devise political, diplomatic and administrative structures that are particularly efficient in establishing control and ruling over the large populations that come

⁹⁹ Christopher Chase-Dunn et. al., "Semi-Peripheral Development and Empire Upsweeps Since the Bronze Age," University of California-Riverside Institute for Research on World-Systems Working Paper #56, presented at the annual meeting of the International Studies Association, New Orleans, LA (Feb. 20, 2010), 2, 5-7.

¹⁰⁰ Levy and Thompson, "Balancing at Sea," 8-10.

under its sway as it expands. For instance, the emerging hegemon may maximize its influence over the system but also retain that influence by identifying and implementing the right combination of centralized supervision and local diversity and autonomy. It may benefit from a uniquely skillful leadership that allows it to build a cohesive empire. A rising hegemon might even copy, improve, and surpass a rival's trademark features, developing better governmental structures or enticing stronger citizenship support, for example. Finally, the rising hegemon's social policies may play a critical role [IV6.4]. The rising hegemon may offer social reforms and opportunities in sharp contrast to its rivals' and therefore not only forestall the revolt of populations brought under its control, but also attract the allegiance of populations and actors it still seeks to control. Such policies may include, among other things, the treatment of minorities and local religious or cultural traditions.

The geographic, economic, political, and social advantages the rising hegemon secures depend on its own qualities and talents to extract and use its resources more efficiently than the potential balancers. Those non-military factors may directly influence the rising hegemon's success if the balancers are not able to catch up and overcome the lead they grant the rising hegemon. While each of the twenty-one factors encompasses a clear and distinct path to balance of power failure, it is unlikely that one variable will supersede all others in the causal relationship because balance of power is a complex mechanism. Generally, we can presume that it will take a conjunction of several factors to trigger a balance of power failure. The independent variables operate at least in part simultaneously. The cases studies will provide an approximation of how much each factor weighs in and will thus permit a ranking of the factors in order of overall causal importance.

c. Evidence and Case Selection

Because balance of power failures occur infrequently, they are best suited for small-n analysis, i.e., case study. “Case studies allow a researcher to achieve high levels of conceptual validity, and to identify and measure the indicators that best represent the theoretical concepts the researcher intends to measure” and are especially appropriate for concepts that are observable in small numbers and “are notoriously difficult to measure, ... such as power ... [or] state strength,” according to Alexander George and Andrew Bennett. Case studies are thus the perfect tool to address the causal complexity of balance of power failure, which requires attention to “the operation of causal mechanisms in individual cases [with a level of] detail” impossible to achieve in large-scale, statistical studies that necessarily omit some variables.¹⁰¹ In addition, case studies permit the capture of the richness and diversity of historical material, and are particularly necessary here because most instances of balance of power failure occurred in pre-modern times and occasionally involve material that is incomplete, requires interpretation, has not been systematically catalogued, or is otherwise inadequate for statistical manipulation. Primary sources, in general the writings of statesmen, historians, and other analysts contemporary to each case, thus provide the bulk of the material for case studies of balance of power failure, reinforced by secondary sources that enable the application of a new intellectual perspective to historical cases. In methodological terms, this thesis's approach is closest to structured, focused comparison, a method based on comparing case outcomes to extract causal variables (“comparison”), asking similar questions in each case study to standardize the results (“structured”), and dealing with specific aspects of the cases in question (“focused”).¹⁰²

¹⁰¹ Alexander George and Andrew Bennett, *Case Studies and Theory Development in the Social Sciences* (Cambridge: BCSIA Harvard University, 2005), 19, 21.

¹⁰² *Ibid.*, 67.

Although the sparseness of cases of successful hegemony was a main criteria leading to the choice of a qualitative approach, it might also provoke a methodological difficulty because the number of variables exceeds the number of cases. It is frequent in macro-political analysis that the number of factors exceeds the number of cases, however. One solution is to develop more encompassing factors, which reduces the number of variables while retaining the overall causal complexity of the object of analysis. This study does so by presenting balance of power as an equation between balancer-driven and hegemon-driven behavior, and thus by identifying balancer- and hegemon-based variables, which are simultaneously responsible for failure and success.¹⁰³ Moreover, the integrity of the findings is not compromised by the larger ratio of variables to cases because the variables are not mutually exclusive. The twenty-one factors are not alternate explanations for one outcome, which would make the outcome over-determined. Instead, it is the variables' combination that generates the outcome, so it is assumed that more than one factor will be present in each case and that the potential combinations are exponential. It is not the individual combinations that are valuable, however, but the recurrence of the variables within each combination. This study is concerned with assessing the factors' complex simultaneous causal effects and identifying their general ranking order rather than just registering their presence or absence in each case.

In addition to analyzing cases of balance of power failure, this thesis will briefly examine a number of short control cases in the conclusion, cases of successful balance of power that will act as a reference. Such research probes are important for methodological purposes, to lend greater confidence in the results of the case studies. In a scientific experiment, control material is used as a standard of comparison. It is exposed to the same conditions as the material that is

¹⁰³ Charles C. Ragin, Dirk Berg-Schlosser, and Gisèle de Meur, "Political Methodology: Qualitative Methods," in Robert E. Goodin and Hans-Dieter Klingemann, eds., *A New Handbook of Political Science* (Oxford: Oxford University Press, 1998), 760-1.

submitted to the experiment, except for the variables being tested, and thus testifies to the change produced by the tested variables. Control cases play a similar role in this inquiry. In addition, they are necessary to avoid selection bias. As Gary King, Robert Keohane, and Sidney Verba emphasize, “in qualitative research, the decision as to which observations [i.e., cases] to select is crucial ... to produce determinate and reliable results.” While “random selection is not generally appropriate in small-n research... abandoning randomness opens the door to many sources of bias.” The most “obvious and basic rule” to avoid selection bias is that “the selection should allow for the possibility of ... variation on the dependent variable.”¹⁰⁴ In order for our dependent variable, hegemony, to take on both a positive and a negative value, we must examine not only cases where hegemony is established, but also cases where the rising hegemon fails and balance of power restored. Using the analyses of a few successful cases of balance of power as control material and comparing them to cases with similar conditions involving a hegemonic rise can help confirm that the causal variables that determine failure of balancing are also those that allow for its success when they register the opposite score.

One of the reasons only few international relations scholars have mentioned, let alone studied, balance of power failure is likely the fact that cases of hegemony have been historically infrequent. The rarity of hegemony makes case selection straightforward. There has never been a case of universal hegemony, where one great power has reached hegemony at the world level and dominated every continent and state. The largest hegemon the world ever experienced was the Mongols, who managed to control a majority of the world that they knew and are therefore particularly worth investigating. In fact, if one redefines the world system in the light of history, one could consider the Mongols to be a case of quasi-universal hegemony. Because the global

¹⁰⁴ Gary King, Robert Keohane, and Sidney Verba, *Designing Social Inquiry—Scientific Inference in Qualitative Research* (Princeton: Princeton University Press, 1994), 128-129.

system did not include as many regions and continents in pre-modern times as it does now, it appeared considerably smaller to pre-modern actors than it is today. For the Greeks, the Romans, the Persians, the Mongols, and other pre-modern peoples, the known world did not include the American continent, most of Africa, Australia, or Oceania. For ancient civilizations, becoming a hegemon meant controlling all of their known world and not the entire world as we now know it. By these standards, the Mongols qualify as a quasi-global hegemon. The study of a pre-modern case like the Mongols begs the question, do balance of power mechanisms apply in ancient times, when actors in the system were not the modern states to which the balance of power literature generally refers? Balance of power mechanisms in reality operate just the same in the absence of modern states. Actors in the pre-modern world, whether they were kingdoms, city-states, feudal entities, or tribal and nomadic groups, responded to the same security impetus that drives modern nation-states to seek to prevent excessive concentration of power in the hands of one of their peers.

The absence of global hegemonies in historic records does not mean that the world lacks examples of hegemony. There are a few cases of regional hegemony that provide adequate material for assessing the causal mechanisms of hegemony. In fact, the logic of balance of power does not operate solely at the global level, but also in smaller areas, where hegemony similarly reflects a failure of the balance of power. The outcome remains the same, except that the scale of the system has been reduced. Regions and continents constitute smaller-scale systems, with their own balance of power, which can be more or less dependent on the global balance. Overall, except perhaps for the Mongols, hegemony has so far been a regional phenomenon. This thesis will therefore examine, in addition to the Mongols, three of the largest regional hegemonies to date: the ascendance of the Roman Empire, the rise of the Qin Dynasty during the Warring State

period in ancient China, and the expansion of the United States into the Western Hemisphere in the late 19th and early 20th centuries. Each case presents critical features that justify its inclusion: the Roman Empire formed the second-largest hegemony aside from the Mongols; the Qin Dynasty exemplifies a purely continental hegemony; and the expansion of the United States is the only instance of regional hegemony in modern history. Because even regional hegemony has occurred infrequently, the cases chosen in this thesis reflect the quasi-totality of the universe of cases. The conclusion will look beyond cases of hegemony by briefly assessing the current U.S. global unipolarity, which is not hegemony and does therefore not warrant a full case study, but must nevertheless be probed because it constitutes the last step before hegemony.

The next six chapters will examine each case of hegemony in turn, in decreasing order of size and closing with the United States, the most recent case.

The Mongol Conquests I The Mongols' Potential Balancers

In a little over a century (1179-1294) the Mongols went from being a small tribe of illiterate, technologically backward, nomadic herders and hunters with an army of less than 100,000 men, to being the largest empire of all time, encompassing 11 to 12 million adjacent square miles (about the size of Africa) of the most thickly populated areas of the planet, from the frozen steppes of Siberia down to the jungles of Vietnam and Burma, from the Korean peninsula to the plains of Hungary and the outskirts of Vienna, via the Balkans, Russia, Persia, and Tibet. In other words, if we control for the vast territory that still lay undiscovered at the time, the Mongols' reach extended so far that they constituted a global hegemon, the only such case history has provided us. Starting from Mongolia, Genghis Khan and his descendents "[went] on to create ... an empire that spanned much of the known world."¹⁰⁵ However, according to balance of power theory, this should never have happened. "On every level and from any perspective, the scale and scope of Genghis Khan's accomplishments challenge the limits of imagination and tax the resources of scholarly explanation," anthropologist and Mongol-expert Jack Weatherford concludes.¹⁰⁶ So, we must ask, what went wrong? Why were the mechanisms of balance of power so utterly inefficient in checking the Mongol advance?

While balance of power failed because of a combination of factors, a few variables

¹⁰⁵ Stephen Turnbull, *Genghis Khan and the Mongol Conquests 1190-1400* (Oxford: Osprey Publishing, 2003), 90.

¹⁰⁶ Weatherford, Jack, *Genghis Khan and the Making of the Modern World* (New York: Crown Publishers, 2004), xviii.

clearly played a crucial role: the Mongols expertly manipulated their opponents with deceptive tactics (IV1), other powers faced tremendous collective action hurdles in their attempts to ally against the rising Mongols (IV2), while the Mongols displayed unusual learning skills and innovative organizational abilities, both on and off the battlefield, that made up for their initial backwardness (IV5; IV6). Taking advantage of the weak and confused state of their opponents, the Mongols crafted and consolidated an empire that was unprecedented not only because of its massive size, but also because of its modern and proficient political, socio-economic, and military achievements that enabled them to transcend the issues that kept the other powers aground. Chapters 3 and 4 identify and trace with precision the factors that led to the Mongols' success, a sweeping hegemony that many scholars of the Mongols still consider "marvelous" and "not fully explicable."¹⁰⁷ Chapter 3 focuses on the deficiencies of the potential balancers, who succumbed to the Mongols' psychological warfare and were ultimately unable to overcome their deep-seated trust issues and preference for immediate gains to form a viable balancing coalition. Chapter 4 highlights the Mongols' unique accomplishments, both in the military sphere where they pioneered a new, highly efficient military structure inspired from their nomadic roots, and in the non-military sphere, with sweeping economic, political, and social empire-building efforts that transformed their military victories into successful political rule.

Timeframe and Boundary of Study¹⁰⁸

The starting point of the Mongol's ascent to hegemony takes place in 1179 when Temujin, the future Genghis Khan and first Mongol emperor who was then the isolated leader of

¹⁰⁷ J.J. Saunders, *The History of the Mongol Conquests* (London: Routledge & Paul Keegan, 1971), 178.

¹⁰⁸ A brief remark on transliteration: Because there are numerous spelling variations in the translation of Mongol and other proper names into Latin script, I followed Jack Weatherford's example and chose "the renderings that are easiest for the English speaker to read, understand, and pronounce." See Weatherford, 287-288.

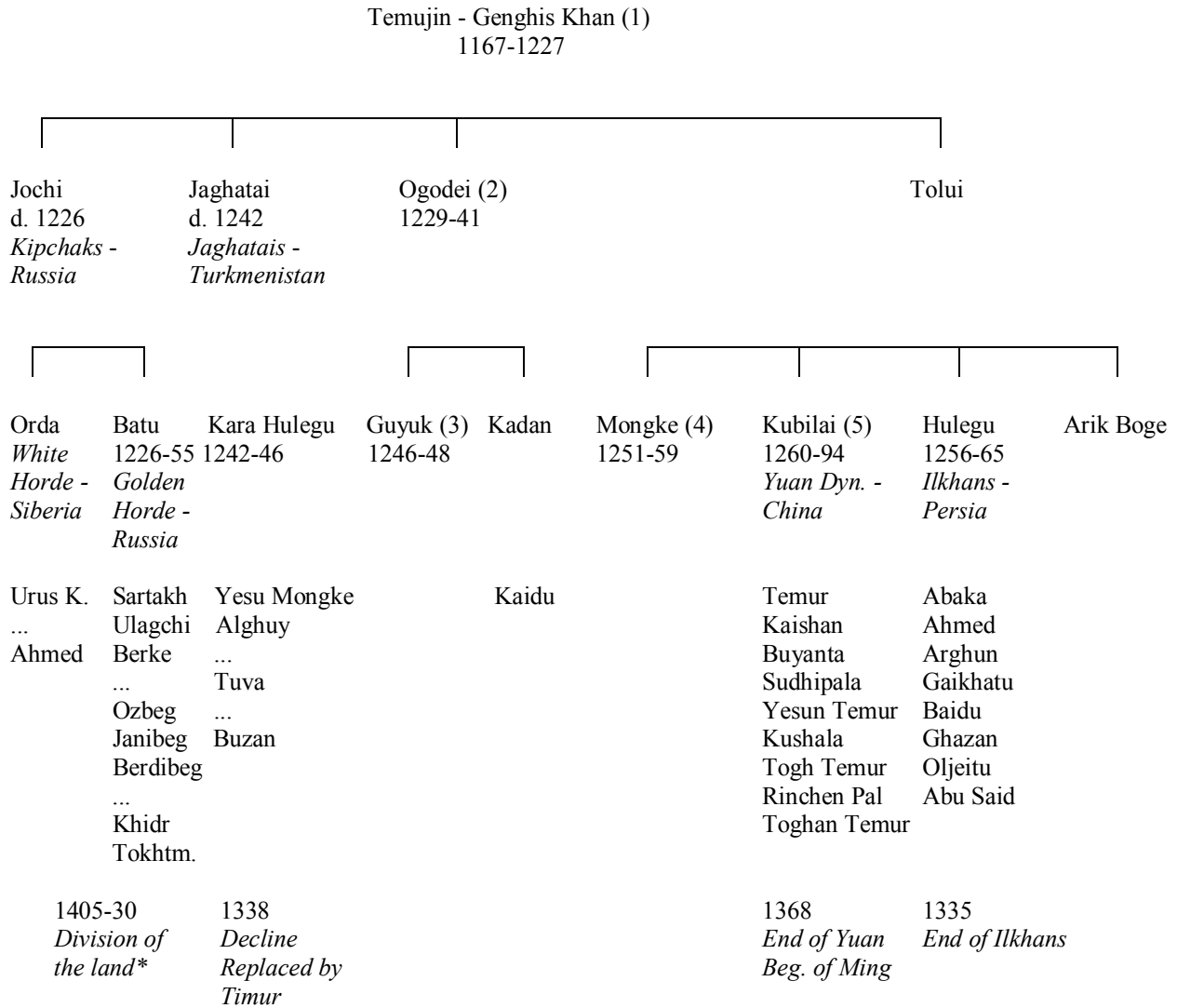
a small Mongol tribe, launched a retaliatory attack against the rival Merkit tribe that had kidnapped his wife Borte, the first step in a quest to unify the various small tribes of the region that had been in constant conflict for centuries. He succeeded in bringing the tribes together under his leadership and was proclaimed Great Khan, or supreme commander, in 1206, after which he turned his attention to neighboring nations. Although the complex unification of the Mongol tribes, from the mid-1190s to 1206, was a necessary path on the Mongol's road to hegemony, the period of interest for the purpose of this study begins in 1206, which marks the evolution of the Mongols' status from an insignificant outsider in the region to a great power contender with real capacity for expansion. The Mongol rise reached its zenith with the reign of Kubilai Khan, Genghis Khan's grandson and the fifth Mongol emperor, during which the boundaries of the empire spread to their furthest extent. The end of Kubilai's reign in 1294 thus constitutes the logical limit of this inquiry (Table 1).

The end of Kubilai's reign is a significant endpoint also because the apogee of the empire's growth symbolizes the beginning of its gradual decline.¹⁰⁹ By the end of the 13th century the empire had spread so far that the Mongols in the various corners of the world eventually lost touch with the central empire, blended into the pre-existing civilizations and, having lost the uniqueness and edge that had allowed their rise, slowly withered away. While the symptoms of decline do not become palpable until the end of Kubilai's reign, its seeds were in reality planted much earlier—and unwittingly—by Genghis Khan himself, who allotted at his death far-apart regions to his sons to avoid bitter conflicts over his succession. This decision might have mitigated internecine divisions, but it also damaged the long-term endurance of the empire as those descendents developed their own sub-empires, or khanates/ulus. At first submissive to the

¹⁰⁹ Komroff introduction to Marco Polo, x.

Great Khan, the khanates progressively grew more autonomous and became quasi-empires of their own.¹¹⁰

Figure 3.1: Genealogy of the Mongol Khans



* One house descending from Batu's brother Togha Timur, the Girai Khans, ruled in Crimea until 1787.
(1) Number in parenthesis denotes order of the five Great Khans (Mongol emperors).

By the end of Kubilai's reign, descendents of Genghis Khan's older son Jochi had settled in Russia and Siberia and become the Golden Horde and the White Horde; descendents of his

¹¹⁰ Genghis Khan mistakenly believed that the khanates would always remain submissive to the central empire (see previous note).

second son Jaghatai had moved to Central Asia and Turkmenistan and formed the Jaghatai Khanate; while descendents of his fourth son Tolui had established the Ilkhanate in Persia. The central empire finally followed the same route. The last Great Khan Kubilai, who was another descendent of Tolui, founded the Yuan Dynasty of China, definitively anchoring his branch of the family into the Chinese civilization. Describing China at the end of Kubilai's reign, Marco Polo observes that the khanates were beginning to have little in common with the original Mongol empire and were increasingly embracing their local civilizations. The "original manners of the Tartar chiefs," namely their battle techniques, lack of logistical hurdles, and unique organizational skills that long gave them victory and superiority, "at the present day ... are much degenerated," Polo writes. "Those who dwell in Cathay [China], forsaking their own laws, have adopted the customs of the people who worship idols, and those who inhabit the eastern provinces [Central Asia] have adopted the manners of the Saracens [Muslims]."¹¹¹ Although some scholars argue that the history of these four khanates into the 14th century may shed light on "processes whereby the Mongol Empire was originally created,"¹¹² that period lies beyond the scope of this study because there is no longer one unified Mongol empire, so that in reality the distribution of power has already reverted from hegemony to multipolarity.

Brief Chronology of the Mongol Expansion

After becoming Great Khan in 1206 and gaining control over the tribes of Mongolia, Genghis Khan made his way toward his larger neighbors, to whom the nomadic tribes, including the Mongols, had traditionally been vassals. In the beginning of the 13th century China was

¹¹¹ Marco Polo, *Milione – Le Divisament Dou Monde (The Description of the World, generally titled *The Travels of Marco Polo*)* (ca. 1298), original Old French and Italian versions published by Arnoldo Mondadori Editore, Milan, Italy, 1982. Transl. into English by Manuel Komroff (New York: Random House/The Modern Library, 1926), 93.

¹¹² Turnbull 90.

divided into three kingdoms: the XiXia Dynasty in the Northwest, inhabited by the Tibet-originating Tangut nomads; the Jin kingdom in the Northeast, populated largely by Khitan tribes but ruled by a Jurchen elite from Manchuria; and the Song in the South, who were the more sedentary and agricultural Han Chinese (Map 3.1). Genghis Khan first turned against the weakest of the three, XiXia, which the Mongols took over 1207-1210, in order to get access to the more powerful Jin. Genghis Khan initiated the conquest of the Jin kingdom in 1211, after refusing to pay the vassal's customary tribute. Because the Jin's territory was so vast, though, this campaign became an enduring enterprise that outlasted him. In 1214 the Mongols captured the Jin's central capital of Zhongdu (modern Beijing), but the Jin emperor fled South and continued the fight from his Southern capital of Daliang (modern Kaifeng). In the North of China, the Mongols took over most of Manchuria by 1216, and obtained tribute from Korea in 1218 after a brief invasion in pursuit of some fleeing Khitans. Genghis Khan's son Ogodei, who succeeded him as Great Khan in 1227, resumed the conquest of the Jin in Southern China and finally defeated them in 1234, absorbing the entire Jin territory into the Mongol empire.

With the ongoing Jin campaign delegated to his commander Mukhali, Genghis Khan decided to simultaneously turn westward, toward the kingdom of the Kara Khitai,¹¹³ populated by Uighurs and Khitans who had fled China when the Jurchen formed the Jin Dynasty and took over the former Liao Dynasty. Genghis Khan was particularly interested in the Kara Khitai for two reasons: they shared their Khitan identity with the inhabitants of the Jin kingdom and could thus potentially support the Jin against the Mongols' ongoing campaign; and their government had been taken over by Kuchlug, a renegade leader of the Naiman tribe Genghis Khan had previously conquered, who might stir up tribal upheavals against the Mongols. In 1218 the

¹¹³ The term means Black Khitans or Black Cathayans in Mongolian.

Mongols defeated the Kara Khitai, occupied their capital, Kashgar, and captured and executed Kuchlug.

Map 3.1: The World before the Mongol Empire, ca. 1200



Source: *Wikimedia*, released into the public domain.

The conquest of the Kara Khitai opened the door to the vast Muslim empire of Khwarezm, which by the 1220s extended over most of modern Iran, Turkmenistan, Uzbekistan, and Afghanistan, and large parts of Tajikistan, Kirgizstan and Kazakhstan. In 1219 Genghis Khan launched a sweeping invasion against Khwarezm's Shah Muhammad II. After a

particularly bloody three-year war, the Mongols subjugated Samarqand, Bukhara and most of Khwarezm, and pursued the fleeing Shah until he eventually succumbed to illness on an island of the Caspian Sea. His son Jelal ad-Din succeeded him and continued resisting the Mongols for a few years, but deserted by his allies, he was eventually defeated by the Mongols on the banks of the Indus river in 1221 and fled into exile in India with a few followers.

While Genghis Khan retired to the Mongolian steppes, he sent his generals Jebe and Subudai to pursue the westward offensive, launching the first of two Mongol campaigns against Eurasia. The generals first conquered the lands between the Black and Caspian Seas: Azerbaijan, Armenia, and Georgia, then moved northward into the territory of the Turkic Kipchak tribes and the Slavic Alans and attacked the Russian principalities. They finally rounded the Caspian Sea, and on their return journey took over Bulgaria, all by the end of 1223. In 1226, shortly before his death, Genghis Khan sent his troops for the second time against his then-vassals XiXia. The Tangut king of XiXia, whom Genghis Khan had left in power after the first invasion, was rebelling and refusing to furnish the soldiers he owed to the Mongols as part of his tributary payment. The conquest was quick and this time the Mongols occupied the land.

Ogodei Khan picked up the Mongols' expansionist agenda just where his father had left it. In addition to continuing and concluding the Jin campaign in Central China, he also returned to Khwarezm, where Jelal ad-Din, back from exile, was attempting to win back the territory conquered by Genghis Khan around Afghanistan. The Mongol army finally defeated Jelal ad-Din in 1231; he was killed shortly thereafter. The same year, Ogodei sent his general Sartaq into Korea, which despite being a vassal was refusing to supply troops for the Jin campaign. When Sartaq was killed in battle there a year later, the Mongols temporarily left the Korean peninsula,

only to return in 1236 as the Koryo king was still refusing to comply with Mongol demands. A final invasion was necessary in 1254 to fully compel Korea to its status as a Mongol tributary.

At the same time, Ogodei made the significant decision to launch the Mongols' second, and bolder, Eurasian campaign, which he placed under the direction of his general Subudai, Jochi's son Batu, and Ogodei's son Kadan. In 1236-37 the Mongols attacked the Bulgars, Kipchaks, and other peoples on the outskirts of Russia. The following winter they took several major Russian cities including Moscow and Vladimir, and turned South just short of Novgorod – the spring mud stopped them from continuing all the way to the Baltic Sea. In 1240 they conquered Kiev, then invaded Poland up to Lithuania, and Eastern Prussia, Bohemia, and Silesia, where Kadan's forces defeated a European coalition army in April 9, 1241, and finally marched into the kingdom of Hungary, where Batu and Subudai's armies crushed King Bela IV and his forces at the Battle of Mohi on April 11, 1241. While resting in Hungary, the Mongols raided Croatia, Albania and the Italian outposts on the Adriatic, their farthest incursion reaching the outskirts of Vienna in July 1241. Ogodei's death in late 1241 brought the European campaign to an abrupt stop as the three campaign leaders and most of the Mongol forces made their way back to Mongolia to attend to succession issues.¹¹⁴

After a five-year interregnum by Ogodei's widow Toregene, their son Guyuk became Great Khan in 1246 for a short, two-year reign. He was followed on the throne by his cousin, Tolui's son Mongke, who ruled from 1251 to 1259 after a three-year interregnum by Guyuk's wife Oghul Qaimish. Kubilai became the last Great Khan in 1260. Despite these rapid successions in leadership, the Mongols pursued Genghis Khan's quest abroad. In 1242-3, Mongol commander Baiju defeated the Seljuk Sultan of Rum, expanding the Mongol grasp to the

¹¹⁴ Batu would later return and settle permanently at Saray on the Volga.

Cilician coast of the Mediterranean. The Mongols also tackled their mightiest opponent yet, the Song Dynasty of Southern China. Mongke initiated the offensive against the Song in 1251, with the goal of uniting all of China under the Mongol empire. After conquering the Yangtze basin and the small kingdom of Dali, the Mongols controlled all of Southwestern China by 1258, confining the Song to the southeast. After first settling for tribute instead of full annexation, Kubilai renewed the efforts against the Song following Mongke's death, but the Song proved rebellious. Finally the Mongols were able to subdue the Song in 1276, proclaiming the advent of the new Yuan Dynasty under Kubilai's leadership.

Mongke also reasserted the Mongols' control of Western Asia in 1251, by sending his brother Hulegu, who would later found the Ilkhanate on these lands, to subjugate the Muslim princes in Persia and Afghanistan, and enforce the payment of the tribute required by their status as vassals. Most of the princes submitted, remembering the Khwarezmian campaigns, but the Mongols faced resistance from the Ismaili sect of the Assassins, who lived in mountain castles in Northern Persia, and from the Abbasid Caliph of Baghdad, Mostassem. After three years of siege, Hulegu's warriors took the Ismaili fortress of Alamut and defeated the Assassins in 1256, then moved on to rout the Caliph and his army and occupy Baghdad in 1258. The Mongols then subdued Syria, conquering Aleppo and Damascus in 1259-1260, and raided parts of Palestine.¹¹⁵ But as Hulegu was preparing an offensive against the Egyptian Mamluk Sultan,¹¹⁶ he received news of Mongke's death. Like Batu twenty years earlier, he interrupted the campaign and returned to Mongolia, leaving only a small force behind that would be defeated by the Mamluks at the Battle of Ain Jalut (Palestine) on September 3, 1260.

¹¹⁵ Samaria, Nablus, Hebron, and Gaza; possibly Jerusalem.

¹¹⁶ The successor of the Ayyubid Sultans.

Kubilai, leaving further Western advances to the khanates of Batu, Jaghatai, and Hulegu, focused his campaigns mostly on East Asia, sending new expeditions to broaden the empire beyond China (Map 3.2). His first target was Japan, but despite several attempts he was never able to conquer the islands. Starting in 1267 Kubilai sent several emissaries to demand the submission of Japan, and all were sent back without even being allowed to talk to Japanese officials. Exasperated, Kubilai ordered two successive naval expeditions against the islands in 1274 and 1277, but both failed without causing much damage to Japan, victims of storms. A third expedition was planned in 1286 but never took place. Kubilai was only slightly more successful with his other target, the Southeast Asian kingdoms, partly because of another weather-related factor, namely the jungles and humid tropical climate that proved a major impediment for his Northern steppe warriors. Starting from the Southwestern Song territory in 1253, the Mongols reached the frontier of Dai Viet (also called Annam) and invaded the country as far as Hanoi in 1287 but at such costs that they had to retreat. Another force that attacked Burma in 1275 failed to conquer the country but managed to obtain tribute from its king. At the same time another Mongol expedition was sent by sea to Champa in 1281, which submitted at first but then rebelled two years later, also forcing the Mongols to turn back. One last mission sailed to Java in 1292, and was eventually expelled by Javanese warriors. Only Siam and a few smaller Southeast Asian states in the Malaysian archipelago and Southern India peacefully accepted to become tributary to the Mongols.¹¹⁷

¹¹⁷ Phillips, E.D., *The Mongols* (London: Thames and Hudson, 1969); Turnbull; Saunders; Weatherford.

Map 3.2: The World after the Mongol Empire, ca. 1300



Source: Ian Mladjov, University of Michigan, reprinted with the permission of Ian Mladjov.

This chapter, the first of two aiming to explain the unlikely ascension of the Mongols to hegemony, focuses on the first set of actors playing a role in the Mongols' successful hegemonic bid—the potential balancers (IVs 1 through 4). It argues that the inability of other powers to effectively counteract the Mongol rise directly enabled the Mongols to achieve hegemony. The next chapter will emphasize the role of the second main actor, the Mongols themselves, in producing that outcome (IVs 5 and 6). It will show that the Mongols' unparalleled skills, both in the military and non-military realms, drove them to surpass their neighbors and competitors in power and appeal and propelled them to hegemonic status.

The first players in the equation, the potential balancers, could have put a halt to the rise of the Mongols and preserved the balance of power, but they were unable to act, whether individually or collectively. Two factors were key in triggering their failure to balance: they faced tremendous communication hurdles, their balancing efforts being particularly disrupted by the Mongols' extensive use of deception and propaganda, and trust-related collective action problems prevented them from collaborating as the Mongol threat grew. Other factors, such as misperception, bandwagoning, or buckpassing, played a minor role, if any at all.

1. Communication Problems (IV1)

Dysfunctional communication contributed significantly to the Mongols' easy progress. The Mongols' targets repeatedly lacked the necessary information to correctly assess the threat they faced and the urgency of balancing. As a result, many failed to act and chose not to balance, allowing the Mongols to advance without major obstacles. Surprisingly, however, imperfect information was not so much due to physical communication barriers, which one would expect to be widespread in the 13th century, or to misperception, but was mostly a product of an ingenious campaign of deceptive, psychological warfare orchestrated by the Mongols.

a. Physical Difficulty (IV1.1)

The absence of modern communication technology did not prevent the Mongols' neighbors and targets, even in distant lands, from learning about the Mongol expansion and the danger facing them. News spread through the slower and less reliable channels of refugees, merchants, spies and ambassadors traveling by foot and horse, or more rarely by boat, but nevertheless reached distant destinations. Most Mongol targets were well aware of the existence

of the Mongols, their army, their conquests and quest for further growth, and the great likelihood of an attack against them, yet still did not balance. Because of the slow pace of communications, one could think that the information reached the Mongols' targets too late for them to organize a riposte—balancing efforts, particularly internal balancing efforts, require time—but in most cases, the targets learned of the danger years before the Mongols attacked them, leaving ample time for preparation. Clearly, something else prevented them from balancing.

Information was easily accessible because the Mongols did not seek to limit it; in fact, they even encouraged it and purposefully allowed refugees to flee after sieges and battles and merchants and ambassadors to observe and travel, since they thought knowledge of their conquests would prompt targets to cooperate and capitulate rather than fight.¹¹⁸ The Mongols preached openness and welcomed foreign missions and exchanges. Even ambassadors and delegates from countries antagonistic to the Mongols were always guaranteed free and safe passage. Whenever the Mongols held a kuriltai, an electoral meeting open to all Mongols that attracted huge crowds, usually after a Great Khan's death or when other crucial military or political decision had to be made, they invited delegations from various countries. At the spring 1246 kuriltai ratifying the choice of Guyuk as the next Great Khan, for example, an impressive list of high-ranking emissaries from every region of the world attended the ceremonies and festivities: the Russian principalities sent Grand Duke Yaroslav; the Seljuks Sultan of Rum sent his son and heir Kilij-Arslan IV; Armenia sent Constable Sempad, the supreme commander of the army; the Ayyubid Sultan of Egypt sent his brother; the Caliph of Baghdad was represented; Korea sent a prince from the royal family; King David VI of Georgia sent his two sons; and Friar

¹¹⁸ See IV1.3, below.

John de Plano Carpini attended as the Vatican's envoy.¹¹⁹ The Mongols took great care on such occasions to display their might, which the emissaries reported on their way home.

Not only did the Mongols encourage reports about their capabilities, they also made their intentions very clear. Carpini took back a letter from Guyuk to the Pope where the Great Khan explains the Mongols' divine mission to take over the world and subject every country to their rule in unambiguous terms. Guyuk writes, "From the rising of the sun to its setting, all the lands have been made subject to me." He then orders the Pope to "submit and serve" him, warning that "if you ignore my command, I shall know you as my enemy. Likewise I shall make you understand."¹²⁰

In reality there is ample evidence that virtually all of the Mongols' targets were aware of their rise and growing might. Immediate neighbors undoubtedly kept an eye on the Mongols already well before their capabilities became threatening. The Chinese kingdoms of the Jin and XiXia that were adjacent to the Mongol steppes knew the Mongol tribes for centuries before Genghis Khan came to power and were monitoring their evolution closely since they were repeatedly interacting in border regions and experiencing recurrent skirmishes and raids. The Jin were especially attentive to the Mongol situation since Kabul Khan, a Mongol clan leader with a particularly strong following, agreed to become a vassal of the Jin emperor around 1135/1140, before rescinding and unsuccessfully attacking the Jin.¹²¹

Even more distant neighbors knew in the first decade of the Mongol empire how threatening the Mongols were. Muhammad, the Shah of Khwarezm, was reportedly very suspicious when Genghis Khan proposed a peace treaty in 1218 to promote peaceful trade

¹¹⁹ Saunders 94-95.

¹²⁰ "Guyuk Khan's Letter to Pope Innocent IV" (1246), printed and edited in *The Mongol Mission* by Christopher Dawson (New York: Sheed and Ward, 1955), 85.

¹²¹ Phillips 24-25.

between Khwarezm and Mongolia. Although he agreed to the treaty, the Mongols' belligerent reputation had obviously not escaped him and prompted him to question the apparently friendly intent of the treaty.¹²² Arabic scholar Ibn al-Athir, a contemporary of Genghis Khan and Muhammad, testifies that all inhabitants of Khwarezm, Transoxania, and the Caucasus region had heard of the Mongols before they attacked: "everybody living knew of this disaster [i.e., the rise of the Mongols], both the learned and the ignorant, all equal in their understanding of it because of its notoriety."¹²³ Yet they failed to stop the invader.

Still further, the Eastern Europeans were particularly aware of the approach of the Mongol warriors because of the mass of refugees that constantly poured into their countries, disseminating tales of horror. During the second Eurasian campaign under the command of Batu, tens of thousands of Cuman (Western Kipchak) refugees sought safety in Hungary with their king Kotyan in 1238, a full three years before the Mongols arrived in Hungary. Nevertheless, King Bela and the Hungarians were not ready and did not stop the invasion. As King Kotyan sent emissaries to his father-in-law Mtislav Mtislavich in Novgorod to warn him, the news of the Mongol advance sped through Northern Europe and reached the Baltic, triggering great alarm but few active preparations. The best example proving that archaic communication technology was not to blame for the lack of balancing is that the Russian provinces and Eastern Europe had already been invaded by the Mongols a first time and still did not react when Batu's army arrived again in 1236. In 1223, in pursuit of the fleeing Shah of Khwarezm, Genghis Khan's army led by Jebe and Subudai pushed into the Caucasus and around the Caspian Sea into Russian, Kipchak, and Bulgar territory, causing much destruction and leaving defeated Russian armies in disarray.

¹²² Saunders 55.

¹²³ Ali 'Izz al-Din Ibn al-Athir al-Jazari, *Al-Kāmil fī'l-Ta'rikh (The Complete History)* (ca. 1231), partly transl. into English (*The Chronicle of Ibn al-Athir for the Crusading Period, Parts 1, 2, and 3*) by D.S. Richards, Part 3: "The Years 589-629/1193-1231 – The Ayyubids after Saladin and the Mongol Menace" (Oxford: Ashgate, 2008), 215.

Thus, these armies and their leaders had first-hand experience of the Mongols' weapons, battle techniques, and attack strategies, and were well aware of the Mongols' power and danger, but they were no more ready thirteen years later when the Mongols returned. In other words, the Russian and Eastern Europeans possessed the correct information and still failed to balance.¹²⁴

Tales of the progress of the Mongol invaders even reached Western Europe, similarly via merchants, refugees, and ambassadors. Even so, the Western European powers would not have been ready or able to stop a Mongol invasion of Europe. In fact, most scholars of the Mongols concur that Batu's army would most likely have taken over Western Europe had Ogodei Khan not died in December 1241. News of the death reached Batu a few months later, and abruptly brought the second Western campaign to an end as Batu left for Mongolia to settle the succession of Ogodei. Thus, "Europe was saved, not indeed by her own exertions but by an event in remote Mongolia."¹²⁵ Lack of information was not the problem. Western Europe learned as early as 1219-1220 of the existence of the Mongol conquerors and their ambitious plans. Caliph Mostassem of Baghdad handed over a Western crusader army he had captured to Genghis Khan, to assist the Mongols in their war against of the Shah of Khwarezm, Mostassem's regional enemy. Having no use for infantry, Genghis Khan released the crusaders, and some made it back to Western Europe with first-hand accounts of the new Mongol enemy.¹²⁶

The next wave of information came to Western Europe through refugees and travelers who had survived the Mongols' first campaign in Eastern Europe in 1223. The Western Europeans were also fully aware of the Mongols' second Western campaign led by Batu. The news even traveled fast and far. British chronicler Matthew Paris reports that in 1238, less than

¹²⁴ *The Chronicle of Novgorod* (ca. 1016-1471, by successive writers), transl. into English by Robert Mitchell and Nevill Forbes (London: Camden Society, 1914; reprint by Academic Intl., Hattiesburg, MS, 1970), 64, 81; Phillips 74; Turnbull 74; Weatherford 142.

¹²⁵ Saunders 88-89.

¹²⁶ Weatherford 110.

two years after Batu's armies entered the Kipchak and Russian territories, fishermen from Gotland and Friesland (northern Germany), fearing a Mongol attack, declined to attend the annual herring fisheries in Yarmouth, England. Their absence resulted in huge overstocks and a considerable drop in the price of herring on the British market but also served to warn the British and other Western Europeans attending the fisheries about the Mongol progression.¹²⁷

In addition, substantial amounts of information flowed through Western Europe's diplomatic channels. The European monarchs were directly warned by King Bela of Hungary and Holy Roman Emperor Frederick II, who sent letters in 1241 describing the Mongol advance and suggesting a common response.¹²⁸ After Batu's armies withdrew the following year and abandoned the Mongols' Western front, European kings and popes sent envoys to the Mongol capital of Karakorum to find out more about their enemies and their intentions toward Europe. Two of the envoys in particular, John Plano de Carpini, who made the long journey to Mongolia between 1245 and 1247 on behalf of Pope Innocent IV, and William of Rubruck, who visited Karakorum from 1253 to 1255 on orders of French king Louis IX (St. Louis), wrote extensive accounts of their encounters with the Mongols and served as messengers between Guyuk and Mongke, respectively, and the European courts. Both men describe the military accomplishments and capabilities of the Mongols and warn the European courts of the Mongols' intention to continue their expansion. In the opening page of his report, Carpini writes to the Pope:

...with your welfare in mind, we shall write to you to put you on your guard... We did not spare ourselves in order to ... be of some service to Christians, that, at all events, having learned the truth about the desire and intention of the Tartars [i.e., the Mongols],

¹²⁷ Matthaei Parisiensis (a.k.a. Matthew Paris), *Chronica Majora* (ca. 1240-1253), edited by Henry Richard Luard, Rolls Series, Vol. 3 (London: Longman & Co., 1876), 488.

¹²⁸ King Bela of Hungary, *Brief König Belas IV an den deutschen König Konrad IV (Letter from King Bela IV to German king Konrad IV)* (1241), transl. into German (*Der Mongolensturm – Berichte von Augenzeugen und Zeitgenossen 1235-1250*) by Hansgerd Göckenjan and James R. Sweeney (Graz, Austria: Verlag Styria, 1985), 285-287.

we could make this known to the Christians; then if ... they [the Mongols] made a sudden attack they would not find the Christian people unprepared ... and inflict a great defeat upon them.

Writing just as Hulegu and the Mongol armies were taking Aleppo, Carpini further cautions that "it is their intention [the Mongols] to attack other countries beyond this" and that Europe was at risk. Carpini even claims hearing from Guyuk that Batu's army did not mean to stop at the outskirts of Vienna but really planned to invade Western Europe, starting with Livonia and Prussia, if Ogodei had not abruptly died.¹²⁹ During his stay at Karakorum, Rubruck witnessed Mongke's preparations for the next phase of conquests, against the Song in the East, and the Ismailis and the Caliphate in the West, and seeing the danger, he advised the French king to take up arms and stop the Mongol advance. "...Indeed," Rubruck writes, "if I were given leave, I would preach war against them [the Mongols] throughout the whole world with all my strength."¹³⁰ Clearly, physical barriers to communication did not prevent the Mongols' targets, even those countries furthest away from the Mongol steppes, from learning about the Mongols' power and expansionist goals on time to plan a counteraction. So what else could have precluded them from taking the necessary preparations?

¹²⁹ Giovanni da Pian del Carpine (a.k.a. John de Plano Carpini), *Historia Mongalorum quos nos Tartaros Appellamus (History of the Mongols which we Call Tartars)* (ca. 1250), transl. into English by a nun of Stanbrook Abbey, UK and edited (*The Mongol Mission*) by Christopher Dawson (New York: Sheed and Ward, 1955), 3, 32, 45.

¹³⁰ Willielmi de Rubruquis (a.k.a. William of Rubruck), *Itinerarium Fratris Willielmi de Rubruquis de Ordine Fratrum Minorum, Galli, Anno gratia 1253 ad partes Orientales (Journal of Friar William of Rubruck of the Order of the Minorite Friars into the East, 1253)* (ca. 1256), transl. into English by a nun of Stanbrook Abbey, UK and edited (*The Mongol Mission*) by Christopher Dawson (New York: Sheed and Ward, 1955), 150.

b. Misperception (IV1.2)

Perhaps the Mongols' targets, despite having the correct information, failed to interpret it correctly. There are in fact a few examples of misperception, but they remain localized and without major consequence. Those that did misperceive information pertaining to the Mongols either overrated their own capabilities, underestimated the threat posed by the Mongols after dismissing the source of the information (i.e., 'shoot the messenger'-type of misperception), or misjudged the Mongols so much that they were paralyzed by fear.

Some leaders, who were fully aware of the Mongol threat, failed to view the danger seriously because they overestimated their own strengths. Caliph Mostassem of Baghdad, for example, reportedly suffered from an inflated sense of self-esteem which triggered a skewed appraisal of the loyalty of other Muslim nations towards him and of his forces' ability to combat the Mongols. When Hulegu approached Baghdad and sent envoys with a request for submission, Mostassem treated them with disdain and expelled them. As Jack Weatherford explains, "the Caliph seemed as incapable of understanding the danger of the Mongols as the Imam [of the Assassins] had been" a few years before, taunting Hulegu from behind the walls of his fortress at Alamut which he believed invulnerable. Mostassem's analysis of the facts was distorted; Hulegu rapidly defeated his army and the center of Islam fell while other Muslim nations passively watched.¹³¹

The Hungarians also misinterpreted the warnings they received about Batu's arrival, not because they were overly confident in their abilities but rather because they dismissed the source of the information, according to the account of Roger of Torre Maggiore, an Italian monk who witnessed the conquest of Hungary and was taken prisoner by the Mongols. When the Cumans

¹³¹ Weatherford 180-181.

(Western-Kipchaks) sought refuge in neighboring Hungary in 1238 while fleeing Batu's armies, they left little doubt that Hungary would be the next target. Yet, because the Hungarians were at odds with the refugees, who caused damage to the Hungarian countryside by trampling farmland with their herds and plundering villages in search for food, the Hungarian psyche fabricated the story of a secret association between the Cumans and the Mongols. Convinced themselves of the duplicity of the Cumans, the Hungarians played down the threat that the refugees had described. They did not even take it seriously when in December 1240, the Hungarian border outposts reported that the Mongol armies were roaming around destroying the adjacent Russian territory. As Torre Maggiore explains, "even as the news spread through all of Hungary, the Hungarians did not believe it." Rather, they focused on their ongoing religious celebrations. When news came in March 1241 of the first border clashes with the Mongols, they "could ... still not believe it" and it took King Bela several days to get his act together and call for the mobilization of his army—much too late, of course.¹³²

Finally, there is evidence suggesting that some countries, though they knew of the imminent threat of the Mongols, exaggerated the threat to such an extent that they were too paralyzed to act. This was the case particularly in Russia and the Caucasus, where much of the information about the Mongols was obtained through the tales of refugees who often gave inflated descriptions of their devastating experiences to justify their defeat. Such accounts encouraged people to misjudge the Mongols as invulnerable, supernatural warriors and forgo rational appraisals of their situation. Believing they had no chance of winning against such unearthly creatures, many simply panicked and concluded that fighting was hopeless. For

¹³² Rogerius de Apulia de Torre Maggiore (a.k.a. Roger of Torre Maggiore), *Carmen Miserabile super Destructione Regni Hungariae per Tartaros* (*Sad Song about the Destruction of the Kingdom of Hungary by the Tartars*) (ca. 1243), transl. into German (*Der Mongolensturm – Berichte von Augenzeugen und Zeitgenossen 1235-1250*) by Hansgerd Göckenjan and James R. Sweeney (Graz, Austria: Verlag Styria, 1985), 149-151, 162, own translation.

example, Armenian chronicler Grigor of Akanc' clearly saw the Mongols as a non-human species:

... the first [Mongols] who came to the Upper Country [Northeastern Armenia] were not like men. They were terrible to look at and indescribable, with large heads like a buffalo's, narrow eyes like a fledgling's, a snub nose like a cat's, projecting snouts like a dog's, narrow loins like an ant's, short legs like a hog's, and ... with a lion's strength... Death does not appear among them, for they survive for three hundred years. They do not eat bread at all.¹³³

Likewise, the chroniclers of Novgorod, in Northern Russia, believed on the eve of the Mongols' first attack in 1224 that they were facing an invincible, mysterious enemy: "at the end of time, those are to appear whom Gideon scattered, and they shall subdue the whole land from the East to the Efrat [Euphrates] and from the Tigris to the Pontus Sea."¹³⁴ This sense of inalterable destiny and fatality was reinforced by those lands' Christian beliefs. Many failed to try to stop the Mongols because they interpreted the Mongol cataclysm as a divine sign and a deserved punishment. As Akanc' exclaims: "Wherefore the Lord roused them in his anger as a lesson to us, because we had not kept his commandments."¹³⁵ All in all, however, such examples of misperception are not widespread—the few mentioned above seem to remain the only ones—and thus, we can conclude that misperception played a minor role at most in the failure of balance of power. What else could be to blame, then?

¹³³ Grigor of Akanc' (a.k.a. Maghak'ia the Monk), *History of the Nation of the Archers – the Mongols* (ca. 1271), transl. into English by Robert P. Blake and Richard N. Frye (Cambridge, MA: Harvard University Press, 1954), 27-29.

¹³⁴ *Novgorod* 64.

¹³⁵ Akanc' 23.

c. Deliberate Deception (IV1.3)

While physical communication barriers and misperception were not a major cause of failed balancing, deceptive tactics, mostly from the part of the Mongols, played a significant role. The Mongols conducted brilliant psychological campaigns in parallel to their military campaigns, strategically manipulating information to their benefit and tricking their opponents into changing their behavior and renouncing to balance. The Mongols' psychological warfare relied on three pillars: the use of spies to learn and exploit the weaknesses of the enemy; a divide-and-conquer strategy of disseminating targeted information to create rifts and disagreements among the enemy; and terror and intimidation tactics exploiting the target's fears and aiming at its surrender without resistance. As Urgunge Onon, Mongolian translator of the *Secret History of the Mongols*, the most important surviving Mongol manuscript to date, points out, "they were masters of the art of deception, espionage, and psychological warfare."¹³⁶ Persian chronicler Minhaj al-Siraj Juzjani writes in 1259 that Genghis Khan "was an adept in magic and deception, and some of the devils were his friends."¹³⁷

1. Spies

From very early on Genghis Khan used spies to outwit his unsuspecting opponents and extract crucial military and other information that would give the Mongols an edge. He was already employing spies shortly after he became Great Khan, while still fighting to unify the various tribes of Mongolia. Searching for a plan to defeat the Ong Khan Toghrul, the powerful leader of the Kerait tribe who was once his ally, Genghis Khan sent two servants of his brother

¹³⁶ Anonymous, *Yüan Ch'ao pi-shih* or *Tobchi'an* (ca. 1228), transl. into English (*The Secret History of the Mongols*) by Urgunge Onon (Richmond, UK: Curzon Press, 2001), Introduction, 12.

¹³⁷ Minhaj al-Siraj Juzjani, *Tabaqāt-i-Nāsirī* (ca. 1259), transl. into English (*A General History of the Muhammadan Dynasties of Asia, 2 Vol.*) by Major H.G. Raverty (New Dehli: Asiatic Society of Bengal, 1881; reprint by Oriental Books Reprint Corporation, New Dehli, 1970), Vol. 2, 1077.

Khasar, who had only recently defected from Toghrul to Genghis Khan, to ask the Ong Khan to take him back and forgive his defection. Phillips argues, however, that the purpose of the servants' mission was "really to discover whether he [the Ong Khan] was ready for battle." Toghrul did not see the trap, and Genghis Khan, "guided by the returned spies," won the crucial 1206 battle of Jejēr-ündür that destroyed the Kerait forces, who were expecting a reconciliation with Genghis Khan and were not the least ready for battle. A few years later, before beginning his campaign against the Jin, Genghis Khan obtained information about the capabilities of his target through a network of non-affiliated merchants and Chinese officials alienated with their governments. He also sent an emissary to the Jin—purposefully choosing a Khitan, not a Mongol, to attract less suspicion—under cover of an official diplomatic exchange, which enabled him to put Jin officials in confidence and learn about their various plans before he attacked them.¹³⁸

The Mongols not only sent their own spies to gather crucial information, but they also frequently hired men working for their enemy as double agents. The Chinese Annals, for example, report that as a result, Genghis Khan had ears everywhere during the 1211 campaign against the Jin, and often learned ahead of time of the movement, position, and intentions of the Jin army, giving the Mongols a clear advantage in battle. The annalist mentions one case in particular. "[Jin] General Wanyen Kyukin detached Ming Gan [one of his officers] to scout out his [Genghis Khan's] position, but this officer entered in the service of Genghis Khan and informed him of everything that took place in the Jin camp." The annalist claims that Genghis

¹³⁸ Phillips 38, 54-55.

Khan's subsequent attack and defeat of the strong Jin general was a direct consequence of the double agent's intervention.¹³⁹

As the Mongols advanced, their use of spies became more widely known and their enemies grew more suspicious of possible spies. To evade the growing mistrust, Genghis Khan would often disguise spies as traveling merchants. In addition to appearing harmless, merchants were favorably positioned to both catch and spread the latest rumors, because markets and caravans were always a prime center of information exchange. Eventually, many nations became apprehensive of peaceful Mongol trading missions. Such concerns may well have played a key role in launching the Mongol campaign against Khwarezm, which started in 1219 after a provincial governor of the kingdom killed Mongol merchants, claiming the trading mission was full of spies. But in response to the growing awareness about spies, the Mongols just used increasing bluff and trickery to place their spies in crucial locations. One particularly striking example occurred further into the Jin campaign, during the Mongol siege of Daqing in Manchuria. Unable to subdue the city, the Mongols thought of sending a spy into the city to coax the inhabitants into opening the gates. They secretly overtook a Jin convoy with a high-level official sent to take charge of Daqing and disguised one of their own who spoke Chinese with the official's clothes. The Mongol took the place of the envoy, took his papers and credentials, and went on to the city. The Mongols arranged to lift their siege just as the fake envoy reached the city, so that the city people would think the convoy defeated the Mongols and welcomed the envoy as a hero. Once in town, the fake envoy ordered the disarmament of the city's defenses and

¹³⁹ Ssu-ma Kuang et. al., *Tong Kien Kang Mou (Annals of the Empire)*, Vol. 9 (11th - 13th century), transl. into French (*Histoire Générale de la Chine ou Annales de Cet Empire*, Vol. 9) by Père Joseph-Anne-Marie de Moyriac de Mailla (Paris: Abbé Grosier, 1779; reprint by Ch'eng-wen Publishing Company, Taipei, 1969), 47, own translation.

the withdrawal of the army, after which he signaled the Mongols, who came out of their hiding places in the surrounding countryside and easily conquered Daqing.¹⁴⁰

2. Divide-and-Conquer

In addition to spies, a second instrument of deceit frequently used by the Mongols was the spread of strategic information meant to create chaos and dissent among the enemy. The Mongols tried to stir up trouble at the international level to prevent alliances, and at the domestic level both among officials to prevent efficient counters to their attacks and between officials and the population to create mistrust in their enemies' leadership.

First at the international level, the Mongols took good care to subtly emphasize and exploit any disagreement and differences to ensure that their enemies would distrust each other rather than ally against the Mongols. Stressing religious rivalries was a particularly successful divide-and-conquer strategy for the Mongols, who alternatively played on the quarrels between the Papacy and the Holy Roman Empire, Christians and Muslims, and Sunnis and Shiites. For example, they convinced both Sunnis and Shiites that they were on their side by destroying their respective enemy, in other words each other, and both sides recurrently fell for the ruse. In the 1250s, Hulegu got the Sunnis on his side by annihilating the dreaded Shiite sect of the Assassins. A few years later, however, he gained the support of various Shiite sects for destroying the Abbasid Caliphate of Baghdad, the core of Sunni Islam, by encouraging the Shiites to believe they would benefit from the end of the Caliphate and presenting the Mongols as saviors against Sunni oppression. To reinforce their claim, the Mongols provided the Shiites with guards for their holy places and employed Shiite notables, tricking them into believing that they were sympathetic to their religion. They similarly played Christians and Muslims off each other.

¹⁴⁰ Saunders 65; Weatherford 93-94.

Hulegu coaxed the Christian kingdom of Georgia into supplying troops for the Mongol invasion of Baghdad by promising to liberate and spare the Christians minority living in Baghdad.¹⁴¹ Ibn al-Athir also describes how Batu cunningly broke up the alliance between the Alans and the Kipchaks, persuading the Kipchaks to support the Mongols against the Alans instead by stressing the two nations' fundamental ethnic and religious differences. As Ibn al-Athir reports, "the Tartars [i.e., the Mongols] sent to the Qipjāks to say, 'We and you are of one race [i.e., nomads]. These Alān are not the same as you that you should aid them, nor is their religion the same as yours!'" As a result the Kipchaks abandoned their ally, which the Mongols promptly defeated.¹⁴²

The Mongols purposefully fueled divisions not only between international actors, but also between officials of a given target country, brewing dissent at the domestic level. During the first campaign against Khwarezm (1219-onwards), Genghis Khan learned about a deep preexisting antagonism between Shah Muhammad's officials and tried to take advantage of it so the Khwarezmians would not present a united front. Thus, when approaching Otrar, the first city to be attacked, he sent propaganda statements emphasizing the divisions between various leaders. But Genghis Khan went further. Upon learning from a Khwarezmian double agent of the sour relations between the Shah and his powerful mother, who effectively ruled parts of the country and had the allegiance of a large portion of the population, he came up with a daring plan. Muhammad Al-Nasawi, secretary of the Shah's son, writes that Genghis Khan had the double agent write a letter on his behalf to all the emirs (princes) of Khwarezm. In the letter Genghis Khan claimed he had originally come to Khwarezm to put himself at the service of the Shah and his mother, but "because today the Shah's sentiments toward his mother have changed, and

¹⁴¹ Saunders 111. The Mongols held their promise. The Christians of Baghdad were forewarned and assembled in a church that the Mongols left untouched when taking the city and slaughtering the inhabitants.

¹⁴² Ibn al-Athir 222.

because he appears rebellious and ungrateful toward her, she asks you to abandon her son. Consequently we await your arrival to place yourselves under our orders and follow our instructions." Of course that was a lie but it served to spur further internal division between the Shah and his mothers' supporters. Genghis Khan followed up by sending an envoy to the Shah's mother offering her an alliance to dethrone her son and put her in his place—another blatant false promise. Although she refused to commit such an obvious act of treason, she fled the country with many of her followers, effectively leaving a major part of the territory undefended for the Mongols to invade—the divide-and-conquer strategy bore its fruit.¹⁴³

Finally, the Mongols disseminated information with the purpose of rousing populations against their leaders and stirring up rebellions in their target states. This was particularly easy when there were preexisting social disturbances or ethnic frictions. During the campaign against the Kara Khitai (1211-1218), Genghis Khan exploited the population's dislike of their leader Kuchlug, a Naiman refugee who had hijacked the throne of the Kara Khitai by marrying old ruler Gur Khan's daughter and overthrowing his father-in-law. Knowing that Kuchlug had alienated the Kara Khitai population by usurping power and persecuting religious minorities and political opponents, Genghis Khan presented himself as the population's savior. Not only did he meet virtually no resistance in conquering the Kara Khitai, but his invasion was even acclaimed for liberating the county from Kuchlug. Genghis Khan took advantage of another ethnic division during the campaign against the Jin. The Jin kingdom was populated by a large majority of Khitans, an Eastern Mongolian tribe that had settled there centuries ago, but it was ruled by the Jin, a small minority tribe that had invaded the area in 1114 and subdued the Khitans. As a result,

¹⁴³ Shihāb al-Dīn Muhammad Al-Nasawi, *Sīrat al-sultān Jalāl al-Dīn Minkbrati* (ca. 1241), transl. into French (*Histoire du Sultan Djelal ed-Din Mankobirti, Prince du Kharezm*) by O. Houdas (Paris: Ernest Leroux, 1895), 64-65, own translation. It goes without saying that the Queen mother was defeated and taken prisoner after the Mongols had conquered Khwarezm.

the Jin ruled over a large population of Khitans, which happened to be very close ethnically to the Mongols, coming from the same general area and sharing the same nomadic roots. Genghis Khan immediately thought about exploiting that distinction, and when the Mongols entered the Jin lands, they presented themselves as the liberators of the Khitans and announced their intention to restore the old Khitan royal family that had been overtaken by the Jin a century before. As a result many Khitans switched their allegiance and joined the Mongols even before the hostilities with the Jin had really begun. Genghis Khan held his promise and restored the Khitan monarchy in 1212, which voluntarily became a vassal of the Mongols.¹⁴⁴

But even without preexisting divisions, the Mongols found ways to stir up civic disobedience among their targets' populations. During the 1258 siege of Baghdad, for example, they sent arrows over the city walls with letters attached to them, delivering the message that anyone renouncing combat and surrendering would be spared, encouraging desertion. In fact, deceptive propaganda and the control and manipulation of public opinion were innate for Genghis Khan from very early on, and he already used them during the unification of the Mongol tribes. After defeating the armies of his Kerait enemies Ong Khan and Jamuka a first time in 1203 but failing to catch either leader, Genghis Khan (then still Temujin) spread rumors that Ong Khan had been killed by guards from the rival Naiman tribe, another tribe not yet subdued by the Mongols. The stories circulated were very elaborate and quite offensive to Ong Khan, and the fake details convinced many that Ong Khan was truly dead and that it was time to switch allegiance to the Mongols. Similarly, to undermine the leadership of the Naiman leader Tayang Khan, who was quite old, the Mongols circulated stories portraying him as senile and

¹⁴⁴ Arthur Waley, introduction to Chih-Ch'ang Li, *Ch'ang-Ch'un Chen-jen Hsi-yu-chi (The Journey of the Adept Ch'ang Ch'un to the West)* (ca. 1228), transl. into English (*The Travels of an Alchemist – The Journey of the Taoist Ch'ang Ch'un from China to the Hindukush at the Summons of Chingiz Khan*) by Arthur Waley (London: George Routledge & Sons, 1931; reprint by Greenwood Press, Westport, CT, 1976), 1-3; Weatherford 89-90; Saunders 55.

effeminate, "using gossip as a way to build confidence in their own men and to weaken the enemy's resolve."¹⁴⁵

3. Terror and Intimidation

In addition to distributing information strategically to divide their adversaries, a third tool of psychological warfare favored by the Mongols was the use of terror and intimidation to exploit the fear of populations and thwart their balancing efforts. The Mongols used violence strategically to deter balancing, by deliberately spreading information that reinforced their image as ruthless, bloodthirsty warriors, and by consistently conveying the message that any rebellion would be met with merciless brutality. Terrorized populations and leaders were left with a choice: try to stop the Mongols and undergo the horrendous end they had heard about, or surrender without a fight and, with some luck, be spared the atrocities. Thus, the goal of the Mongols' terror warfare was to scare their targets so much that they would submit without fighting and remain submissive out of fear of bloody reprisal. The Mongol generals were particularly attentive to cost-efficiency and avoided battles and blood-shedding whenever possible for the simple reason that the Mongol warriors were not numerous compared to their enemies—in fact, they were often at a great disadvantage in raw numbers, hence their lives were precious. In light of their limited human capital, terror warfare was very attractive because it enabled them to conquer whole areas without combat.

To put this into practice, the Mongols usually presented their enemy with the choice of surrendering or being annihilated, in no unclear terms. Persian chronicler Ala-ad-Din Juvaini, an Ilkhan official under Hulegu, reports that the Mongols regularly sent envoys into towns before attacking to lay out clearly the options for the population. For example, Juvaini writes, Genghis

¹⁴⁵ Turnbull 58; Weatherford 60.

Khan sent a Muslim envoy, Danishmand Hajib—to win over the trust of the population—to the towns of Khwarezm before attacking them. To the residents of Zuraq, near Bukhara, Hajib reportedly said: "If you are incited to resist in any way, in an hour's time your citadel will be level ground and the plain a sea of blood. But if you ... become submissive and obedient to his [Genghis Khan's] command, your lives and property will remain in the stronghold of security." The Mongols generally kept their promise to towns that submitted, at least to a certain extent—they would order the population out, execute the soldiers to prevent any danger of future rebellion, plunder the town, take a few prisoners to serve in their army and other posts, but most frequently left the population alive.¹⁴⁶ The same warnings were reiterated a few years later when the Khwarezmians were spurred to revolt by the deceased Shah's son, Jelal ad-Din. According to Al-Nasawi, Genghis Khan's son Jochi, then in charge of the campaign, "sent a message to the Khwarezmians, threatening to become angry and warning them of the consequences of resistance, while on the other hand promising them survival if they surrendered without combat." Nevertheless, Nasawi says, "the opinion of the mad ones prevailed over the wise" and many towns chose to fight, ending up razed and their population slain.¹⁴⁷

Similar intimidation was used to deter most of the Mongols' targets. Akanc' credits Genghis Khan with saying to his delegate in Armenia: "It is the will of God that we take the earth and maintain order, and impose the [tribute] ... Those, however, who do not submit to our command and give us tribute, slay them and destroy their place so that the others who hear and see should fear and not act thus." Akanc' then describes the towns that obeyed and were spared,

¹⁴⁶ 'Ala-ad-Din 'Ata-Malik Juvaini, *Ta'rikh-i Jahān Gushā (The History of the World Conqueror)* (ca. 1252-3), transl. into English by John Andrew Boyle, 2 volumes (Manchester, UK: Manchester University Press, 1958), 99-101.

¹⁴⁷ Al-Nasawi 154-155, own translation.

and those that resisted and were destroyed without mercy.¹⁴⁸ Ilkhan historian Rashid al-Din recounts that Hulegu used a comparable plan to obtain the submission of the Assassins, offering their leader Khurshah amnesty if he capitulated and emphasizing to him that "such a step would save many lives." Khurshah dragged his feet but eventually conceded, and Hulegu kept his word. Rashid al-Din also recalls that upon marching onto Baghdad, Hulegu sent Caliph Mostassem an envoy with a letter carrying the same deterrent undertones and offering no violence and destruction if the Caliph surrendered peacefully and became a Mongol vassal. The letter concludes, "naturally if you obey our orders, it would be disgraceful of us to show hate, and you will remain in possession of your territory, your troops, and your subjects." Mostassem haughtily rejected the offer, and Baghdad was destroyed and set on fire, and all its inhabitants slaughtered, as well as the Caliph and his entire family. Hulegu reportedly concluded with philosophy: "Any man who submits to us with sincerity can be assured to preserve his goods, his wife, his children, and his life. But a man who acts with hostility will not reap any benefits from it."¹⁴⁹

To reinforce the credibility of their message, the Mongols did not hesitate to build up their ruthless reputation. As Turnbull underlines it, "it is quite clear that the terror inspired by the Mongol name was deliberately used as a psychological weapon." The Mongols sought to develop this brutal reputation and at times deliberately engaged in massacres to reinforce it. Terror was always used as a political and military strategy, however, for the ultimate purpose of enticing future targets to surrender. There is no evidence of indiscriminate killings just for the sake of killing. While some historians and the popular psyche tend to describe Mongol leaders enjoying random mass murders and sitting on top chariots filled by piles of human skulls, their

¹⁴⁸ Akanc' 33, 44-45.

¹⁴⁹ Rachid al-Dīn Tabīb, *Jami 'al-Tawarikh (Compendium of Chronicles)* (ca. 1307-1316), transl. into French (*Histoire des Mongols de la Perse*) by Etienne Marc Quatremère (Paris, 1836; reprint by Oriental Press, Amsterdam, 1968), 195-197, 211, 213-214, 217, 231-239, 305, own translation.

descriptions in reality relate to the Huns and Tamerlane, who were adept at perverse torture and butchery. Both the Huns, who preceded the Mongols by some eight hundred years, and Tamerlane, founder of the 14th century Timurid Empire, are often confused with the Mongols because they were also nomadic invaders and shared the same geographic and ethnic origin.¹⁵⁰

The Mongols used terror as a systematized strategy. They always allowed some refugees to flee to the next town to spread the news, frighten residents, and destroy the enemy's morale before the attack even started, knowing that refugees often exaggerated apocalyptic tales. Some stories are quite inflated, actually describing the Mongols as devils from hell and adding gory details.¹⁵¹ Some areas really suffered awful destruction as punishment for rebelling against the Mongols, especially in XiXia and in Khwarezm by the Oxus River, and tales from there were particularly gruesome. During a battle in the town of Nishapur in Persia, because Genghis Khan's son-in-law Toghachar was killed by the enemy, the entire town was slaughtered, including the cats and dogs. Rashid al-Din writes that all the inhabitants of Aleppo, Syria, had their throats cut because "fully confident in the strength of their citadel, they [had] refused to surrender."¹⁵² Akanc' recalls the story of a Georgian prince who had drunk too much at a banquet and boasted he would take up arms and defeat the Mongols. Word got out to the Mongols and they annihilated him before the king was able to clear the misunderstanding, which deterred the "terrified and distressed Christians" from even considering rebellion.¹⁵³

The Mongols used many tricks to spread terror. One of their most common stratagems was to use local prisoners as human shields—a stratagem described by witnesses of the Jin, Khwarezm, and Eastern European invasions. The Mongols would collect prisoners, from one

¹⁵⁰ Turnbull 48; Weatherford 252-253.

¹⁵¹ Matthew Paris, Vol. 1, 314.

¹⁵² Rashid al-Din, 333-337, own translation; Turnbull 24-26.

¹⁵³ Akanc' 53-55

town or from the surrounding countryside and send them ahead of their warriors in the front line of assault against the next town. Anyone retreating or refusing to fight would be executed. People inside the ramparts would often recognize family in the assailant's army and refuse to fight. Turnbull argues that this trick made many Jin cities fall on the Mongol's way to Zhongdu (Beijing) in 1213-1214. Similarly, the *Voskresensk Chronicle* reports that upon attacking the Russian city of Vladimir, the Mongols displayed their prisoner Prince Vladimir in front of the gates. He was the son of Grand Duke Yuri II, whose army was in charge of Russia's defense. Seeing this, people inside the city walls deduced that Yuri's army must have been defeated and began losing their morale—although Yuri had in reality not yet been defeated. To put further psychological pressure on the city, the Mongols sent several hundred Russian prisoners in plain sight to build a protective palisade around the city to shield the Mongol attack. Once the military attack finally begun, Vladimir did not last very long.¹⁵⁴

The *Secret History* mentions yet another trick. After becoming Great Khan, Ogodei resumed the campaign against the Jin in 1231, and he "put Jebe in the vanguard," according to the ancient text. But the *Secret History's* translator, Onon, correctly points out that this is factually impossible since Jebe, one of the great Mongol generals, died in 1225 on the way back from the Mongols' first European raids around the Caspian Sea. Onon interprets this as yet another deceptive tactic of the Mongols. Because Jebe was highly feared, Ogodei purposefully did not publicize his death. Instead, "to scare the enemy, Ogodei was still using Jebe's name in 1231," Onon writes.¹⁵⁵ The same way, the Mongols would circulate fake stories of their exploits to increase the fear of their targets. Against the Jin, the Mongols circulated a story that they had promised a city to lift their siege if the city people gave them their cats and birds, and that when

¹⁵⁴ Turnbull 31, 48.

¹⁵⁵ *Secret History* 264.

the city people had done so, the Mongols had attached torches to the animals and lit them on fire, and the panicked animals had fled right back into the city, setting it ablaze. Though unrealistic, this kind of propaganda, coupled with word-of-mouth from witnesses, recurrently intimidated entire populations into surrender.¹⁵⁶

The Mongols' many deceptive tricks and stubborn psychological pressure worked. There are numerous examples of cities that surrendered out of fear, deterred by the Mongols' terror warfare. Terror tactics worked better the further the Mongols went. It was particularly effective against the people of Persia, Central Asia and Eastern Europe, to whom the Mongols appeared as "strange, alien savages," allowing Genghis Khan and other Mongol leaders to "exploit... the 'barbarian factor' so well," Turnbull explains. Terror warfare was comparatively less effective against China, which was used to attacks by nomadic horsemen from having had them as neighbors for centuries. Ibn al-Athir testifies to the success of the Mongol terror tactics in Persia. "I have been told stories about them [the Mongols], that the hearer can scarcely credit, about the fear of them that God Almighty cast into people's hearts," he writes. He goes on to describe incidents where the population would voluntarily surrender even when the Mongols arrived to their town in very small numbers—in one episode where a lone Mongol warrior entered a town, "nobody dared to raise his hand on that horseman."¹⁵⁷ As Genghis Khan approached Bukhara in 1219, the second large city on the Khwarezm campaign, some twenty thousand soldiers left their posts in the city and fled in a panic, leaving only about five thousand troops behind to barricade themselves in the city's citadel. They were easily defeated when the civilian population spontaneously surrendered and opened the city's gates to the Mongols.¹⁵⁸

¹⁵⁶ Weatherford 93-94.

¹⁵⁷ Turnbull 77-79; Ibn al-Athir 307.

¹⁵⁸ Juvaini 99-100.

Similar accounts are reported from the Caucasus. Carpini describes that after the city of Barchin in Crimea refused to submit, Batu subdued it with great destruction, and "the inhabitants of another city called Sakint, hearing of this, came out to meet the Tartars [i.e., the Mongols] and of their own accord surrendered to them." The same happened in Eastern Europe, where word of the devastation of major cities like Kiev spread like wildfire. Batu and the Mongol armies were able to take Krakow with virtually no opposition, for example. Boleslaw the Chaste, Prince of Krakow, fled before their arrival, along with many officials and soldiers, so that no siege was required. No siege was necessary either at their next target, the Silesian capital of Breslau, where the inhabitants burnt the city themselves and abandoned it to plunder before the Mongols arrived. During the later campaign against Persia and the Middle East, Hulegu had a similar experience. When the inhabitants of Damascus, Syria, learned of the slaughter at Aleppo, "they themselves, of their own will, gave over the city and the key of the city into the hands of Hulawu Tan [Hulegu Khan]," Akanc' explains. Rashid al-Din even writes that "some of the officials and high-ranking persons from the town went to meet the conqueror, bringing with them presents, valuable objects, and the key to their houses."¹⁵⁹

Thus, there is ample evidence of dysfunctional information during the Mongol conquests that directly hampered or prevented balancing efforts, but it was primarily a result of the Mongols' own deceptive tactics rather than a consequence of archaic means of communication or misperception. By waging a skillful psychological campaign, which included the use of spies, divide-and-conquer strategies, and widespread terror and intimidation, the Mongols were able to stall their opponents' balancing efforts and facilitate their rise to hegemony. But their balancing-free rise was also greatly helped by their opponents' inability to organize a collective response, leaving the Mongols essentially free to conquer with impunity. The Mongols' targets were

¹⁵⁹ Carpini 29; Turnbull 51; Akanc' 81; Rashid al-Din 339-341.

impaired by collective action failure and were mostly unable to cooperate at all to stop the Mongols. A handful of instances of laggard balancing occurred, but most were too late or too weak to bear any fruit.

2. Collective Inaction (IV2)

Collective inaction—the absence of any cooperative balancing effort—was rampant among the Mongols' targets and clearly contributed to their seemingly effortless rise. It was mostly due to a lack of trust, following the classic collective action scenario, because the Mongols' enemies continually focused on local conflict and short term gain at the expense of the larger threat of the Mongols. There was also some evidence of communication problems that prevented the formation of balancing alliances, but there was no evidence of interest-, or buckpassing-related collective inaction. The absence of countervailing coalitions bears special salience in the case of the Mongols, because they rapidly became a very strong aspiring hegemon. The more powerful a rising hegemon becomes, the more vital it is for balancers to concentrate their force if they want to be successful. In other words, when facing an overwhelming and impenetrable rising hegemon like the Mongols, external balancing constitutes a much more realistic option than internal, individual balancing. In addition, given the power distribution, which consisted mainly of medium- and small-size states and lacked significant great powers, internal balancing efforts against the Mongols were virtually futile. Because external balancing was thus necessary to counter the Mongol advance, its absence—collective inaction—was a pivotal cause of balance of power failure.

a. Communication Problems (IV2.1)

Attempts at deliberate deception by the Mongols not only prevented internal balancing by individual states, as mentioned in the previous section, but also indirectly affected external balancing efforts, encouraging collective inaction. If the Mongols' terror campaigns worked well in deterring individual states and towns' balancing efforts and prompting them to surrender, the same campaigns were by extension also effective in preventing common balancing efforts.

And just in case the psychological campaigns were insufficient for that, the Mongols made sure there would also be physical barriers in place to bar leaders from various regions and kingdoms from communicating and exchanging information, thus rendering most attempts to ally materially impossible once the conquests had started. The Mongols used targeted physical communication barriers aimed specifically at preventing official interaction and planning between potential balancers, while continuing to encourage the spread of information about their rise and terror campaign so as not to undermine their psychological effect. When planning an invasion, Genghis Khan and other Mongol leaders would first and as soon as possible send scouting troops into the target area, which would not only map out the best routes for troops to follow, but also immediately take control of critical roads and interior lines of communications, and seize any official correspondence. This isolated their targets and, with no communication or negotiations between their enemies' camps possible, forestalled any military cooperation against them. Once the Mongols had invested a particular region, it became virtually impossible for anyone to travel without being intercepted by their outposts, as Carpini and other envoys testify. All travelers not only needed to carry letters of safe-conduct approved by Mongol officials, but their passage also had to be allowed by Mongol stations everywhere along the way, which at each stop required bribing one's way through with an array of gifts and providing lengthy

explanations for one's whereabouts— a situation hardly amenable to secret military coordination between countries.¹⁶⁰

b. Lack of Trust (IV2.2)

But much more strikingly than communication, it was the traditional collective action hurdles that prevented the Mongols' targets from cooperating and balancing together. As game theory models suggest, trust issues routinely lead to collective action failure among actors. Actors fear that they might be cheated or double-crossed if they agree to cooperate and share the costs of a collective action. Trust issues are particularly difficult to mitigate when the actors are pre-modern states, since those states lacked elaborate agents like today's international organizations or press corps to monitor compliance and discourage defection. States in the Mongol times operated in a rudimentary international system with no institutional or normative mechanism to guarantee, or at least promote, compliance between allies. The risks of being cheated for relative gains were thus especially high.

In addition, trust issues are exacerbated when the actors have a history of conflict, because the potential partner has a reputation for hostility and thus the risk of being cheated increases dramatically. It just so happens that all of the Mongols' targets were at some point or another engaged in local conflicts around the time of the Mongol conquests. Trust issues becoming particularly salient, local quarrels almost always took precedence and hampered alliance efforts, even though it was in the actors' interest to collaborate to face the larger threat of the Mongols. These local conflicts occurred simultaneously between neighboring kingdoms, but also within specific kingdoms. In the feudal societies of the time, local power was much more

¹⁶⁰ Saunders 66.

prominent than in the modern nation-states. Each king had to deal with more or less autonomous regions led by local princes and warlords, with their own populations and armies, who would occasionally rebel or turn against the king's authority and added to the difficulty of creating a united front against the invader. Not only were kings mostly unable to cooperate with neighboring kings, but they were often also incapable of placing all their local feudal lords under one banner. This political and military fragmentation and the presence of such strong decentralized centers of power just made international alliances all the more difficult to arrange because it multiplied the number of allegiances and thus the sources and risks of defection. The constant belligerence and lack of trust not only between neighbors, but within kingdoms, basically ruined any possibility for an alliance.

Not one region was spared of those inter- and intra-state conflicts that naturally played to the Mongols' advantage and greatly facilitated their rise, since as a result the Mongols faced an array of weak, easily beatable adversaries rather than one powerful balancer.

1. Mongolia

A Yuan-dated history based on a now-lost Mongol royal manuscript reports that Genghis Khan already benefited from local quarrels that hampered collaboration between his enemies even before becoming Great Khan. The Tartars, one of the most powerful tribes of the Mongolian steppes whose name is often—wrongly—equated with the Mongols', broke off their alliance with the neighboring Jin just as the Mongols, their traditional enemy, were gaining power. Given the strong capabilities of the Tartars and the Jin, fighting both at the same time would have proved dangerous for the new Mongol army. Genghis Khan might not have been able to defeat the Tartars if their petty disagreements with the Jin had not led them to renounce

their alliance, or he might at least have had to postpone an attack against them until he had strengthened his own forces. Their fallout was just the lucky break he was waiting for. The Yuan document says that "when the Emperor [Genghis Khan] learned this [the breakup of the alliance], he immediately gathered his troops that he kept nearby and ... went to meet [the Tartars] to attack them." The Mongols went on to defeat the Tartars, who did not receive any assistance from the Jin.¹⁶¹

2. China

The Chinese theatre was particularly prone to inner tensions that impeded the ability of the various actors to trust each other and work together to stop the Mongols. The Chinese annals mention that the XiXia and the Jin had enjoyed relatively peaceful relations for about eighty years prior to the Mongol conquests, but because the Tanguts of XiXia were the Jin's vassals, the hierarchical relations necessarily fostered resentment from the part of the Tanguts and indifference at best from the Jin, for whom the Tanguts were not an integral part of the kingdom but just another vassal. Thus, when the Mongols headed for XiXia, the Tanguts appealed to the Jin for help, thinking that their vassal status implied protection, but the Jin refused to get involved and as an excuse, pointed out that the Tanguts had not always displayed loyalty toward them in the past. The Tanguts, too weak to resist the Mongols on their own, were forced to settle and submit to the Mongols, but they were "irritated by [the Jin's] refusal." In retaliation, they started attacking the Jin in 1210, just as the Jin were under attack from the Mongols, and kept on encroaching upon the Jin territory for years, in effect subjecting the Jin to a multiple front war.¹⁶²

¹⁶¹ Wang Kuo-wei, *Cheng-wou Ts'in-tcheng Lou (Tale of the Campaigns of Genghis Khan)* (ca. 1285), transl. into French (*Histoire des Campagnes de Gengis Khan*) by Paul Pelliot and Louis Hambis (Leiden, Netherlands: E.J. Brill, 1951), 191-192, own translation.

¹⁶² Ssu-ma Kuang et. al., 57, own translation.

In 1213, an additional front opened for the Jin, as they were also faced with domestic insurgency from the Khitan people, a tribe which constituted the majority of the population of the kingdom and whom they ruled, and a Khitan leader, Zhi Zhong, assassinated the Jin emperor Wei Shao Wang in the capital of Zhongdu.¹⁶³

Local conflicts did not end there for the Jin. During the entire period of their ordeal at the hands of the Mongols, from roughly 1211 to 1234, their neighbors to the South, the Song, refused to help in any way and were instead satisfied by the Jin's struggle against the Mongols, since they considered the Jin as their traditional enemy. The Jin and Song shared a long history of hostility. The two kingdoms fought a lengthy war over the control of China in the 12th century, in which the Jin took Kaifeng from the Song and considerably reduced the Song's lands, confining them to the South of China. But the Song periodically sent retaliatory raids against the Jin, doubling their efforts after the Mongols started attacking the Jin. The Jin were thus facing multiple attacks from the North and South. The Chinese annals report that as a result, the Jin's ability to counter the Mongols was severely weakened. "Upset that he had to divide up his forces to resist both to the Mongols and to the Chinese [Song]," the Jin emperor offered the Song a truce in 1218, hoping that they would then be able to cooperate to stop the Mongol incursions into China. But the Song emperor "was well aware of the [weakening] position of the Jin and of the rapid progression of the Mongols" and refused. Angered, the Jin emperor then re-started the war against the Song, giving himself a full-blown two-front war engaging both extremities of this vast territory because he let a local dispute distract him from the greater threat of the Mongols. In 1219 the Song inflicted a severe defeat on the Jin army and forced it to retreat, and thereafter the Song continued to regularly raid the Jin from the South. At the end of 1219, the annals conclude,

¹⁶³ Phillips 55.

because of his foolish inability to rise beyond local rivalries to cooperate with his neighbors, "the king of the Jin [was] beaten from all sides."¹⁶⁴

Even if the Jin had sought to coordinate a response to the Mongols with their neighbors, it would have proved impossible since they were in conflict with virtually every one of their neighbors and even with their own population. The focus on local rivalries and short term gain did not serve their neighbors well either. Almost half a century later when the Mongols had achieved their conquest of the Jin, they simply moved on to subdue the Song. Again, the local rivalry played to Mongke and Kubilai's advantage since the Song could expect no sympathy and hence no aid from the rest of China that was already under Mongol rule. Though the former Jin could have rebelled and helped the Song retain the last morsel of Chinese land against the foreign invader, the local rivalry was too deep to transcend. Just as it eased the Mongols' conquest of the Jin and the Song, the lack of trust between neighbors also greatly facilitated the Mongols' invasion of Korea. In 1216, the Mongols pursued the fleeing Khitan inhabitants of the Jin kingdom all the way to the Yalu River. The Khitans asked the neighboring Koryo kingdom for help but the Koreans refused to even send them basic supplies, let alone troops. Partly out of resentment and partly because they desperately needed supplies, the Khitans crossed the Yalu River and helped themselves by pillaging Korean villages. The Mongols, undeterred by the border, crossed the Yalu River after the Khitans in the winter of 1216 and invaded Korea.¹⁶⁵

3. Middle East

The lack of trust, exacerbated by local conflicts, was even more rampant in Persia and Central Asia, and prevented any type of collaboration to balance the Mongols. Sunni and Shiite

¹⁶⁴ Ssu-ma Kuang et. al., 83-89, own translation.

¹⁶⁵ Turnbull 35; Saunders 122.

populations were unable to ally despite the threat of the Mongols and instead focused on their own quarrels. Eventually both were destroyed by Hulegu's armies. Arabic scholar Ibn al-Athir describes how the Muslim world—Persia, Iraq, Syria, Turkestan etc.—never attempted to combine their forces. He clearly identifies the root of the problem as the constant discord within the Islamic world: "[The] misfortune was that those who escaped these two hordes [the Mongols and the Franks, who attacked the Muslim world from the East and West, respectively] were at daggers drawn between themselves and dissention was raging... The cause of these Tatars [i.e., Mongols] only prospered because of the lack of a strong defender." Ibn al-Athir blames Khwarezmian Shah Muhammad in particular for being a warmonger in the fifteen years leading the Mongols' arrival and aggressively seeking expansion at the expense of other Muslim states and princes, aggravating many and stirring up local conflicts.¹⁶⁶ But to Muhammad's credit, he was also unable to find allies in the Muslim world because of his different ethnic roots. He was of Turkic origin while most other actors around Khwarezm were Arabic or Persian and viewed the Turkic leader as a barbarian.¹⁶⁷

As Phillips remarks, "this enmity was a source of weakness for the Khwarezmian Empire and for the whole of Islam in Asia." Muhammad started a war with the Kara Khitai, his neighbors to the East, instead of seeking their support against the Mongols, because their leader Kuchlug was persecuting his Muslim Uighur subjects. The Uighurs, however, seeing the danger posed by Muhammad's belligerence, appealed instead to the Mongols for protection. Jebe took immediately advantage of these divisions to attack Kuchlug, oust him, and restore the old Kara Khitai monarchy as a vassal, before turning against Khwarezm.¹⁶⁸ Muhammad then conquered various Transoxanian cities (in modern Uzbekistan) in an effort to expand his kingdom to rule

¹⁶⁶ Ibn al-Athir 204.

¹⁶⁷ Weatherford 110.

¹⁶⁸ Phillips 58-59.

from the Indus River to the Caspian Sea, at the same time collecting more enemies instead of gathering allies. Muhammad was also at odds with the Caliph of Baghdad, his neighbor to the Southwest. Ibn al-Athir adds, however, that Caliph al-Nasir, who preceded Mostassem until 1225, also contributed to stirring up the rivalry, citing a secret message that the Caliph sent to the Mongols 'inviting' them to attack Khwarezm in retaliation against Muhammad.¹⁶⁹ In the end the Muslim world was so fragmented that the Mongol invasions were effortless compared to other theatres like China, where subduing the enemy took significantly longer.

In addition to provoking his neighbors, Muhammad also constantly picked fights with his own regional princes. As a result, by the time the Mongols attacked, his power was so diluted that any resistance against the invader was hopeless. There was so much antagonism between Khwarezmian princes that Muhammad refused to put his army in the hands of one general, for fear that he might turn against him. The result was a highly undisciplined army, with little allegiance to Muhammad.¹⁷⁰ The domestic antagonisms were reinforced by the ethnic diversity of the country. While Muhammad and his army were mostly of Turkic origin, Khwarezm also held large pockets of Persian and Tajik minorities that felt no allegiance to either the Shah or the army despite their common belief in Islam and sought to take advantage of them rather than facilitate their task. In return the Shah's soldiers cared only little about protecting Persian and Tajik lands since they had little in common with their population.¹⁷¹

As briefly mentioned above, even the members of the Shah's own family were in a feud, raging on the question of how to run the country, between the Shah and his mother, a Kipchak from the North of Khwarezm who had the support of the Persian minorities and controlled most

¹⁶⁹ Ibn al-Athir 132, 205, 261. There is most likely some truth to that story because it is seconded by other, less Persian-friendly, accounts.

¹⁷⁰ Turnbull 20.

¹⁷¹ Weatherford 110.

of the Northern regions herself. The imminence of the Mongol invasion seemed to heighten their quarrel rather than soothe it as one would expect in the face of external danger. Al-Nasawi, the secretary of Muhammad's son Jelal ad-Din, credits the family dispute for ultimately leading to the downfall of Khwarezm. The governor of the Eastern city of Otrar, who in 1219 seized the Mongol merchants and their cargo in violation of the peace treaty with the Mongols, triggering of the Mongol invasion, was none other than Inal Khan, son of Muhammad's maternal uncle. Al-Nasawi implies that Inal Khan purposefully provoked the Mongols into attacking to bring further trouble to Muhammad on behalf of his mother. The domestic sabotage continued after Genghis Khan took over Otrar when another official of the city, Bedr ed-Din el-Amid, who hated Muhammad virulently for ordering the killing of some of his family members, disclosed the rivalry to Genghis Khan. "Bedr ed-Din informed ... Genghis Khan of the hostility and revulsion that existed between the Sultan [Muhammad] and his mother," which Genghis Khan did not fail to exploit.¹⁷²

Just like his father, Muhammad's son Jelal ad-Din, who took over after his father's death in 1220, stirred up conflict and distrust both in and out of Khwarezm instead of seeking a united front and allies to tackle the Mongols. After returning from a three-year exile in India and in the midst of his fight against the Mongols, he turned his army against the Caliph of Baghdad, because of whom the Sultan of Dehli had refused to welcome Jelal ad-Din during his exile. The Caliph fended him off, so Jelal ad-Din moved into the Caucasus and invaded Azerbaijan in 1225 and Georgia in 1226, then fought the Seljuks of Rum, attacked the Ayyubids, and harassed the Ismailis before eventually being defeated by the Mongols just as he was planning to capture Armenia and re-attack the Caliphate. By his erratic behavior, Jelal ad-Din turned every neighbor into an enemy, leaving the door wide open for the Mongols.

¹⁷² Al-Nasawi 57-62, 64, own translation.

He also alienated the local Khwarezmian princes against him. They were aware that his aggressiveness was ruining any chances of survival they may have had. As Ibn al-Athir underlines it, "Jelal ad-Din was a bad ruler who administered his realm abominably. Among the princes ... he did not leave one without showing hostility to him and challenging him for his [land] ... Every prince abandoned him and would not take his hand." Ibn al-Athir concludes that Jelal ad-Din "revealed an unparalleled lack of good sense" verging on incompetence. He acted with such hostility that the Ismailis wrote letters to the Mongols asking them to destroy him, and that eventually "his troops had turned against him and his vizier, along with a large section of the army, had cast off their loyalty," Ibn al-Athir notes incredulously. In 1227-1228, Jelal ad-Din's own half-brother, Ghiyath al-Din, "became fearful of his brother." Ibn al-Athir continues. "Several emirs were also afraid of him." These princes left Khwarezm and took refuge with the Caliph, while Ghiyath sought the protection of the Ismailis, which almost triggered a retaliatory war from Jelal ad-Din.¹⁷³ Of course Al-Nasawi, Jelal ad-Din's personal secretary, presents a slightly different account of these events, biased by his position, in which a heroic and adventurous Jelal ad-Din is resisting against challenges and attacks from his neighbors and is forced to re-conquer renegade provinces of Khwarezm and to purge his own ranks because of internal rebellion.¹⁷⁴ Yet regardless of which side one takes, these events just provide repeated evidence of the predominance of local discord and immediate gains in relative power and the consequent failure of collective action in the face of a massive threat. "The Muslims were broken because of this," Ibn al-Athir concludes, and stood no chance in their division of withstanding the Mongol conquests.¹⁷⁵

¹⁷³ Ibn al-Athir 288-289, 303-304.

¹⁷⁴ Al-Nasawi 185-223.

¹⁷⁵ Ibn al-Athir 229.

Domestic frictions damaged the Muslim world's chances beyond just Khwarezm, though. The Ismailis, Rashid al-Din argues, would have had a hard time finding allies against the Mongols even if they had tried because their sect was considered heretic by the rest of Islam and they spread terror throughout the Muslim world. In fact, their neighbors had been meeting and trying to devise a plan to destroy the Ismailis when the Mongols invaded them. But the Assassins were not even able to seek allies, because they were too consumed by their own domestic struggles for power. In addition to isolating the Assassins even further, this in-fighting really helped the Mongols because the Assassins killed their own leader, Ala-Eldin Mohammed in 1255, in the midst of their struggle with the Mongols. Rashid al-Din writes that Mohammed's own son Khurshah commissioned his assistant to kill his father—which exemplifies the absence of loyalty and the high risk of defection allying with the Assassins would have meant—then executed the assistant and his whole family to erase any trace of his own culpability. In the Caliphate also, the reign of Mostassem was contested by various domestic forces at the time of the Mongols' arrival, which similarly denotes his lack of legitimacy and of any capacity for collective action. The Djemris, a minority that had acquired a significant military force and that was led by two high-ranking officials serving the Caliph, were plotting to overthrow the Caliph, whose authority was weak.¹⁷⁶ Juzjani adds that "enmity had arisen" between Mostassem's eldest son and his Shiite vizier, Ahmad, over a massacre in a Shiite village.¹⁷⁷ Moreover, as Rashid al-Din explains, "the inhabitants were tired and disgusted of the government of the Abbasids." The situation was particularly chaotic, with the dynasty on the verge of collapse and a quasi-civil war erupting when the Mongols arrived. Instead of pulling together and presenting a united front as the Mongols were about to attack, the dissidents took advantage of the situation to further

¹⁷⁶ Rashid al-Din 157, 185, 189-191, 227-229, own translation.

¹⁷⁷ Juzjani 1229-1234.

undermine the Caliph's rule.¹⁷⁸ The Muslim world was thus affected by such dire political divisions, both domestically and between states, that their collective inaction in the face of the Mongol threat, which greatly facilitated the Mongol invasions, is hardly surprising.

4. Russia and Europe

In Russia, the Caucasus, and beyond, collective inaction was similarly spurred by local conflict and the inability to build relations of trust with neighbors. Akanc' explains that the Europeans and Muslims in the Caucasus were more interested in their own conflict than in working together to fend off the Mongols. When the Mongols arrived in the Caucasus, the king of Georgia had just died and the Seljuk Sultan of Rum captured Dawit, his only son and heir, from a Seljuk stronghold in Georgia, thus virtually handing their neighbor over to the Mongols since Georgia was deprived of a leader. After invading the kingdom, the Mongols cleverly sent a Mongol detachment to liberate Dawit and restored him on the throne, effectively buying off the loyalty of its new vassal and exacerbating the Seljuks' disdain for the Georgians. Thus, they ensured that these two neighbors would never combine forces against them.¹⁷⁹

The *Chronicle of Novgorod* also documents with countless examples that at the eve of the first Mongol invasion of Eastern Europe, in 1223, cooperation was unthinkable among the Russian principalities and between the Russians and their neighbors. "The Lithuanians made war round Toropets, and Yaroslav with the men of Novgorod went in pursuit of them," the chronicler reports. He continues: "the same year ... Khyaz [Prince] Yaroslav ... went with all the province to Kolyvan, and conquered the whole Chud land, and brought back countless plunder." Even the year of the Mongol attack, 1224, divisions were running deep: "the Nemtsy [i.e., the Germans—

¹⁷⁸ Rashid al-Din 245-247, own translation.

¹⁷⁹ Akanc' 25; 27; 49; 51.

likely Teutonic knights from Livonia on the Baltic coast] killed Knyaz Vyachko in Gyurgev and took the town." Or again: "The same year ... Posadnik [Governor] Fedor rode out with the men of Russa and fought the Lithuanians." The Russian princes also fueled the rivalry with the neighboring Cumans (Western Kipchaks) by blaming them for the Mongol attacks of 1224 because the Mongols followed the fleeing Cumans into Russian territory.¹⁸⁰

The Russian princes were still fighting amongst themselves in 1237 when Batu led the second campaign against Eurasia, even though they had severely suffered from the first Mongol incursions. As Weatherford eloquently puts it, "the Europeans quickly forgot about the Mongol victories and returned to their own squabbles." And of course, Saunders adds, "the Mongols found aid in the ... quarrels of their opponents." Even after hearing that the Mongols were approaching for the second time, the Russians did not coordinate their military strategy and forces. As a result the attack in the late 1230s was devastating. The Golden Horde ruled over Russia for almost two centuries.¹⁸¹ The *Chronicle of Novgorod* describes how interestingly, even while a vassal of the Golden Horde in the late 13th and 14th century, the Russian principalities failed to act commonly to oust their invader. Instead, they repeatedly requested the Mongols' intervention to settle their quarrels, again showing that short term, limited gains in power remained their priority.¹⁸²

Further West, much of the same evidence is repeated. In Eastern Europe internal dissent and internecine conflicts virtually thwarted any possibility of collective action. In Hungary in 1241, right before the Mongols' second invasion, King Bela faced such tremendous dissent from all parts that Torre Maggiore characterizes it as "hatred between King Bela and the Hungarians," which he believes was "the reason for the destruction of Hungary [by the Mongols]." In other

¹⁸⁰ *Novgorod* 63-64.

¹⁸¹ Weatherford 142; Saunders 82; Turnbull 44.

¹⁸² *Novgorod* 104-105, 108, 110-111.

words, Hungary was so caught up with its own conflicts that it was incapable of preparing for the arrival of the Mongols and effectively seeking help outside. The population was in a quasi-insurrection since 1239 because of the damage from the plundering by the violent Cuman refugees, which the Hungarians blamed on King Bela since he allowed the Cumans onto Hungarian land without consulting the population. Numerous fights erupted between Hungarians and Cuman refugees, generating widespread chaos. An angry mob assassinated the Cuman king Kotyan after suspecting him of a crime he was later exonerated of. By the time the Mongols arrived, King Bela had lost most of his authority. The rich and aristocrats were angry at him too because he jailed many of his political opponents, almost exclusively members of the aristocracy, and confiscated the territory and belongings of countless others. Thus, Torre Maggiore writes, by the end of 1240, roughly a year before the Mongol attack, the people of Hungary were contemplating violent action against their own king.¹⁸³ The political fragmentation and general mistrust both in Hungary and between Hungary and its Cuman neighbor was such that collective action could not even be envisaged.

Western Europe was not spared by internecine divisions either, which precluded any cooperative action against a possible Mongol invasion. Most scholars of the Mongols recognize that the fragmentation of Western Europe was so severe that if Batu had not left the European theatre in 1242 and returned to Mongolia to attend the kuriltai following the death of Ogodei, Western Europe would have most likely been overtaken by the Mongols because it was simply unable to put up a collective defense. The main conflict lay between the two highest authorities of the continent, the Pope and the Holy Roman Emperor, who each called for crusades to be sent against the Mongols but at the same time kept accusing each other of supporting the Mongols. They could not agree to send help to King Bela when the Mongols attacked Hungary in early

¹⁸³ Torre Maggiore 141-151, 157-159.

1242. Similarly, while the Mongols were attacking the Russian principalities the previous year, Novgorod was simultaneously assailed by an army of Teutonic knights from Livonia. Though the knights were eventually defeated, they forced Novgorod to significantly reduce the troops assigned to stopping the Mongols.¹⁸⁴

The powerful Holy Roman Emperor Frederick II, probably the last figure to be able to shield Western Europe from the Mongols, fiercely clung to his animosity with the Pope and the clergy and out of fear of being double-crossed, did not make any overtures toward Rome, the other strong power in the region. He nevertheless sent a letter to the kings of Europe in July 1241, describing the fall of Kiev and Hungary and the danger it presented for Germany and the rest of Western Europe. In the letter he requested from each king a number of arms and troops to organize a common defense. But the European kings, just as suspicious as Frederick II, ignored the plea. King Bela simultaneously sent an appeal to Pope Gregory IX from exile, but the Pope died a month later and the cardinals in Rome refused to take any steps in the absence of a new pope. Even after Batu had returned to Mongolia, trust issues continued to plague the Europeans, who were unable to organize themselves in the event that the Mongols came back to conquer their lands. The new pope, Innocent IV, tried harder than his predecessor to transcend regional disputes, by calling together all the European kings and the Holy Roman Emperor at the Council of Lyon in June 1245 to discuss a Christian alliance against the Mongols. But they were unable to reach an agreement and the Council was adjourned without any plans for an alliance.

In his report to the Pope, Carpini clearly warned the Western Europeans that if they did not ally, they stood no chance of staving off the Mongols. It is impossible for one isolated king or prince to balance an enemy that powerful, he writes:

¹⁸⁴ Turnbull 55.

If one province is not prepared to help another, then the country the Tartars [i.e., the Mongols] are attacking will be vanquished ... Therefore if Christians wish to save themselves, their country and Christendom, then ought kings, princes, barons and rulers of countries to assemble together and by common consent send men to fight against the Tartars [i.e., Mongols] before they begin to spread over the land.¹⁸⁵

Carpini clearly makes an urgent call for collective balancing. Nevertheless, the Europeans, just like other potential balancers of the Mongols, sabotaged their chances of overcoming the rising hegemon by focusing on their numerous internal conflicts and pursuing relations of mistrust that impeded all collaboration.

c. Lack of Sufficient Interest (IV2.3) and Buckpassing (IV2.4)

While the lack of trust and constant bickering among the Mongols' opponents clearly played a significant role in preventing balancing and facilitating the Mongols' rise, the absence of balancing interest and buckpassing appear much less relevant.

In general some states may decline to organize or participate in a balancing coalition because they feel less imminently threatened by the rising hegemon, but this was evidently not the case with the Mongols' enemies. Both weak and strong countries, and even those countries furthest away from Mongolia, notably in Western Europe and Southeast Asia, showed concern and in many cases a will to take action to stop the Mongols, but were unable to cooperate with their neighbors because of the aforementioned collective action hurdles. Very few failed to understand the emergency of the situation and the necessity to balance. One example of insufficient interest may have been the Song. Though forewarned by a Jin ambassador in the early 1230s that they would be attacked by the Mongols, the Song did not understand the

¹⁸⁵ Carpini 45-46.

closeness of the threat until decades later when attacked by Mongke's armies. Their distance and power gave them a false sense of security, and thus they did not anticipate the Mongol advance and made no overtures toward the Jin.¹⁸⁶

There is even less evidence of buckpassing behavior. Given the sweeping power of the Mongols and the gap in power between them and the next most powerful kingdoms, passing the buck was hardly an option to ensure that the Mongols be stopped. As the strength of the Mongols increased, the participation of all became increasingly important, if not vital, for successfully fending off the invader. In addition, the lack of trust between the Mongols' enemies guaranteed that no one would sit back and rely on their neighbor to take on the task of blocking the Mongol advance.

Thus, collective inaction played a major role in enabling the Mongols' rise to hegemony. In particular, the considerable difficulties of cooperation—the lack of trust between neighbors, spurred by a generalized focus on local disputes and short term gains—prevented any collaborative effort that would have been necessary to make balancing successful.

3. Laggard Balancing (IV3)

During the Mongols' rise collective action mostly failed altogether (IV2). There is surprisingly little evidence of cooperative balancing efforts to stop the Mongols. In the few instances where the Mongols' targets attempted to form balancing coalitions, their efforts constituted clear examples of laggard balancing—balancing undertaken too late, or displaying insufficient skills and capabilities, or simply too weak, to stop the Mongols. In most cases the evidence does not permit to know for sure the cause of the laggard balancing, but it is easy to

¹⁸⁶ Turnbull 35.

extrapolate that just like with collective inaction, traditional cooperation hurdles—lack of trust and focus on immediate gain—were the main factors that dampened the few balancing coalitions that were put together. For the reasons outlined above, again, nothing points toward lack of sufficient interest or buckpassing.

a. Communication Problems (IV3.1)

The Mongols did not hesitate to use deceptive psychological tactics to try to breakup already formed alliances, as the example of the Russian-Cuman alliance shows. Upon having heard that after the 1223 invasion of the Kipchak territory, a remnant of the Cumans (Western Kipchaks) had fled to Russia and allied with the Russian princes, the Mongols tried to lure the Russian princes into breaking up the alliance by reminding them of their previous fallouts with their Cuman neighbor—a classic example of the divide-and-conquer tactic. According to the Chronicle of Novgorod, the Mongols sent the following message to the Russian princes:

Behold, we hear that you are coming against us, having listened to the Polovets [i.e., Cuman] men; but we have not occupied your land, nor your towns, nor your villages, nor is it against you we have come. But we have come sent ... against ... the Polovets men ... If they escape to you, drive them off thence, and take to yourselves their goods. For we have heard that to you also they have done much harm; and it is for that reason also we are fighting them.¹⁸⁷

The Mongols were obviously attempting to stir up the old animosity between the Russians and the Cumans, which the allies had managed to overcome in the face of the imminent, larger threat posed by Jebe and Subudai's approaching armies. In this case, however, the Mongols' psychological tricks did not bear fruit. The Russians paid no attention to the message, killed the

¹⁸⁷ *Novgorod* 65.

Mongol envoys that had delivered it, and went on to fight the Mongols alongside the Cumans. The allies, however, did not succeed in stopping the Mongols. They did not pose a real problem for Jebe and Subudai, who flatly defeated them at the Battle of Kalka in Crimea on May 31, 1223. Since the Russians did not fall into the trap set by the Mongols, this particular alliance obviously did not fail because of the Mongols' deceptive tactics but because of some other factor. Nevertheless, the Mongols likely used similar tactics to disband other alliances. Other leaders may have hesitated after receiving a comparable message from the Mongols and delayed coordinating plans with their allies long enough to jeopardize their chances of success.

b. Lack of Trust (IV3.2)

While some alliances may have been formed too late and may have subsequently failed because of the Mongols' deceptive tactics and other communication-related issues, most collaborative efforts against the Mongols were victims of trust issues and disunity, which prevented allies from efficiently coordinating their efforts. Traditional cooperation problems bear an enormous potential for undermining military effectiveness, especially when facing a unitary, rigidly and hierarchically organized actor like a Mongol army. Allies need to have a unified command that synchronizes maneuvers and strategies and assigns tasks. Various battalions and individuals need to support each other and overcome the lack of a strong, common allegiance. Trust is a vital requirement for a coalition army. Expectations of defection, double-crossing, and gains at the expense of one's ally will immediately undermine the cooperative enterprise. The few among the Mongols' targets that considered or even managed to put together an alliance suffered principally from this defect. Unable to build a relationship of trust, their efforts often

stalled early and they did not to build a working coalition until it was too late, or their coalition, once built, was too weak and impractical to slow down, let alone stop, the Mongol advance.

1. Mongolia

Such trust issues already hampered coalition building against the Mongols when Genghis Khan was still Temujin, attempting to unify the Mongol tribes. Several tribes tried to cooperate to enhance their chances against the Mongols. But they were mostly former enemies, and by the time they were ready to commit to an alliance, they were usually right about to be attacked by Temujin, and the alliance at that point was too late to be effective. For example, the Yuan Dynasty history of Genghis Khan relates that "the Khatagin, Salji'ut, Dorben [three clans related to Temujin's], Tatars and Onggirat [Temujin's wife Borte's clan] met at the Alui source to seal an alliance; they sacrificed a white horse and swore to attack our [the Mongol] army." Temujin preempted the attack and took the offensive shortly after the meeting, inflicting a severe defeat upon the alliance and integrating most of its members into his growing empire.¹⁸⁸

2. China

The same trust issues marred alliance formation and effectiveness later on in the Chinese theatre. When Genghis Khan targeted the XiXia kingdom in 1207, the Tangut king appealed to the Jin to jointly take on the Mongols but they refused. The irritated Tangut attacked the Jin in retaliation instead of insisting upon an alliance, ruining his chances of obtaining assistance and greatly facilitating the Mongol attack. When the Mongols invaded XiXia for the second time in 1226, the Tangut king had—this time successfully—assembled an alliance, which included forces from the former Jin and from Tibet. He therefore thought he was ready to fend the

¹⁸⁸ Wang Kuo-wei, 393, own translation.

Mongols off, but the Mongols easily defeated the alliance. There is no clear record of the causes of the alliance's failure, but one can only speculate that a hastily organized coalition of neighbors having just fought a war could not possibly provide a robust line of defense against the invader.¹⁸⁹

3. Middle East

Further West in Khwarezm, the same difficulties of cooperation also impaired alliance efforts against the Mongols. Jelal ad-Din managed to enlist various surrounding tribes to fight the Mongols, and even inflicted a temporary defeat upon the Mongols at the Battle of Parwan. His victory was short-lived, however, because of his inability to keep his patchwork of forces united. Juzjani explains that the different tribes making up the coalition eventually started fighting with each other. "In the army of Sultān Jalāl ud-Dīn were a great number of the Ighrāk [Afghan and Ghur] ... men, and between that body of the Ighrāk and the 'Ajāmis [Turks] and Khwārazmis, a quarrel arose ... and hostility ensued; and that body of Ighrāk troops separated from Sultān Jalāl ud-Dīn," Juzjani writes. As a result of the alliance breaking down, Jelal ad-Din was forced to retreat, and his much smaller army was defeated by the Mongols at the Battle of the Indus in the spring of 1221 that forced him into exile.¹⁹⁰

Jelal ad-Din failed to draw the right conclusion from this episode, however, and upon his return from exile erratically attacked his neighbors instead of seeking their collaboration. When the Mongols came back to stop his rampage in 1230, Jelal ad-Din understood, but too late, the need for allies. As Ibn al-Athir explains, "he saw ... his own feeble and weak state ... [and] his plan was to ... visit the Caliph's court to ask for aid from him and all the princes against the

¹⁸⁹ Phillips 63-65; Turnbull 29.

¹⁹⁰ Juzjani 1021.

Tatars [i.e., Mongols] and warn them of the result of their failure to act." He also reached out to the Seljuk Sultan of Rum. But because of his focus on local wars and beating his neighbors, this attempt to find allies came much too late. Jelal ad-Din and the Seljuk Sultan were unable to agree on an alliance, and the Mongols attacked Khwarezm before Jelal ad-Din's envoys even reached Baghdad. Moreover, since he had attacked just about every one of his neighbors just months before, including the Caliph and the Seljuks, it is unlikely that the Caliph or any prince of the region would have consented to an alliance. Other alliance efforts in the region failed for similar reasons—local conflicts remained too widespread to trust anyone with military cooperation. In the 1240s, the Caliph of Baghdad managed to enlist the Ayyubid Sultan of Egypt into a defensive alliance against the Mongols, but it was an alliance in name only. When Hulegu's armies marched onto Baghdad in 1258, no Egyptian army came to the rescue of the Caliph.¹⁹¹

4. Russia and Europe

The same story was repeated in the Caucasus and Russia—trust issues hampered and delayed all collaboration efforts against the Mongols until it was too late. After their first invasion of Khwarezm, the Mongols moved toward Georgia. The Georgians, after a brief encounter with the Mongols where their army was severely defeated, reached out to their neighbors, but these neighbors were suspicious because they had not been in friendly terms. Ibn al-Athir describes the effort: "The Georgians sent to Uzbek [i.e., Uzbek Ibn Pahlawan, lord of Tabriz, Azerbaijan], asking him to make peace and form an alliance to resist the Tatars [i.e., the Mongols] ... The Georgians also sent to Al-Ashraf Ibn Al-'Adil, lord of Khilāt and the Jazīra, asking him to reach an agreement with them." Both leaders consented but showed their

¹⁹¹ Ibn al-Athir 305; Phillips 71-72, 84.

reluctance by holding off their troop commitment until after the winter had passed. Of course, that was too late; Jebe and Subudai invaded Georgia in January 1221.¹⁹²

As mentioned above, when the Mongols moved onto Kipchak and Russian territory, some Russian princes and Cumans formed an alliance to safeguard the Russian plains against the Mongol advance. Prince Mtislav Mtislavich of Galich, upon hearing from his fleeing Cuman father-in-law King Kotyan that the Mongols were arriving, allied with Princes Mtislav Romanovich of Kiev, Mtislav Svyatoslavich of Chernigov, and with the princes of Smolensk, Volhynia, Kursk, and Suzdal. The Chronicle of Novgorod reports that Mtislavich stressed the importance of an alliance to the other Russian princes. "If we ... do not help these [Cumans], then they will certainly surrender to them [the Mongols], then the strength of those will be greater," Mtislavich reportedly said, displaying a clear understanding of balance of power mechanisms. Together they gathered some 82,000 troops to face the Mongols. The Cumans also contributed a cavalry numbering several tens of thousands.

But when they met the armies of Jebe and Subudai at Kalka in Crimea in 1223, lack of trust and quest for personal gain at the expense of the others was apparent and led to their downfall. First, the alliance suffered from a lack of clear leadership. Likely unable to reach common ground on who was to lead the coalition army, the members of the alliance put Mtislavich, Romanovich, and Svyatoslavich jointly in command of the troops, resulting in major coordination problems and confusion of movement as such a joint command structure was incapable of responding promptly to the Mongols' intricate battle tricks. On the Mongols' side, in comparison, Genghis Khan had clearly put Jebe in charge, with Subudai as his second-in-command. Then, the armies of the various princes tried to outsmart each other to steal the honors of victory, instead of acting in concert, and effectively ruined the common effort. The armies of

¹⁹² Ibn al-Athir 214-215.

Volhynia and Kursk launched the attack without waiting for the other Russian armies and the Cuman reinforcements. The next army, from Chernigov, arrived at the scene just when the Mongols counterattacked, and, unaware that the battle had started, ran into their retreating comrades and blocked their escape route. The chaos that resulted greatly benefited the Mongols, who easily ambushed the remaining Russian forces upon their arrival.¹⁹³

The same problems of cooperation—mistrust and obsession with short term gain—continued to plague the Russians and prevent successful external balancing against the Mongols when the Mongol armies returned under Batu in 1238 for their second campaign in the region. The Chronicle of Novgorod clearly underscores these problems: "... the Knyazes [i.e., princes] of Ryazan sent to Yuri of Volodimir [i.e., Vladimir] asking for help ... but Yuri neither went himself nor listened to the request of the Knyazes of Ryazan, but he himself wished to make war separately." Intent on taking all the credit for beating the Mongols, the Russian princes decided to fight the Mongols individually, with no chance of success. The chronicler goes on to list the principalities, which fell one after the other, including the largest ones. After the invasion had started, with the Southern region of Ryazan, some princes reacted and sought to cooperate. But of course, their efforts were too late and too isolated to be effective. Duke Yuri II of Vladimir eventually sent a relief force, but it arrived after Ryazan had been destroyed. Vladimir was soon invaded too and Yuri killed. Another force was sent from Suzdal, but was similarly defeated by the Mongols.¹⁹⁴

Eastern Europe faced the same cooperation hurdles that doomed their alliance efforts. After invading Russia, Batu's armies continued toward Eastern Europe in 1240, where one of the most significant collective attempts to stop them occurred. Henry the Pious, Duke of Silesia,

¹⁹³ *Novgorod* 65; Weatherford 139-142; Turnbull 74; Phillips 63-65.

¹⁹⁴ *Novgorod* 82; Turnbull 47.

succeeded where so many others had failed and managed to assemble a coalition army of about 30,000 men under his sole command. The army was comprised of Henry's own Polish and Moravian troops, which were mostly conscripts, as well as Czech, French Templar, and Teutonic knights sent by the Pope, and even a battalion of Bavarian gold miners enlisted in haste as Henry was trying to gather all the manpower he could. He also made plans to join forces with the army of King Wenceslas of Bohemia. Yet the coalition was defeated at the Battle of Liegnitz in Silesia on April 9, 1241, not even by the whole Mongol army but by a side force of no more than 20,000 men from Subudai's contingent. Despite the more unified leadership, the total lack of communication between the heteroclitic allied battalions led to such disorganization that the various battalions undermined each other. The Mongols, who to the contrary benefited from a brilliant communication system based on colored flags, took advantage of the allies' absence of coordination and executed their usual battle tricks, such as fake withdrawals and diversionary movements, creating massive confusion among the allies. Without joint preparation and collective responses, the allies did not stand a chance.¹⁹⁵

After King Bela's Hungarian army was defeated only two days after Liegnitz in the Battle of Mohi by another side force of the Mongols, Bela thought that the Europeans would finally overcome their divergences and cooperate to confront the Mongols. Nevertheless, suspicions still triumphed. Bela immediately wrote to Holy Roman Emperor Frederick II, the most powerful ruler in the region and thus the most likely to have the authority to put together a coalition army against the Mongols. In his letter Bela described the horrors inflicted by the Mongols to neighboring regions and claimed to know "from a secure source" that the Mongols' plan after invading Hungary was to take over German lands. Therefore, Bela wrote, "we plead ... that you ... come to our rescue without delay." The letter is not dated but from the context one can

¹⁹⁵ Weatherford 152; Saunders 85; Phillips 75; Turnbull 51.

assume that it was sent in the spring of 1241, shortly after the Mongol attack, thus much too late to put together an alliance that could save Hungary. Moreover, far from being eager to help, a mistrustful Frederick II replied that he would cooperate, but that in the meantime Bela should ask his son Konrad IV of Bavaria. By the time the request reached Konrad, the Mongols had invaded all of Hungary. Another incident highlights the deep mistrust between Europeans. As the Mongols swept through Hungary, Bela fled with the remainder of his army and reached Pressburg (modern Bratislava), seeking to take refuge with Duke Frederick of Austria. But Frederick, skeptical of Bela's intentions because of prior disputes, charged him a large fee and refused him asylum, sending him forward to Croatia instead of helping him form a common line of defense.¹⁹⁶

As Saunders rightly concludes, the Europeans' lack of trust and their focus on immediate gains undermined all the cooperative efforts that were undertaken: "the reaction of Europe to the Mongol invasion proves that the most alarming danger is incapable of compelling bitter rivals to sink their differences and unite if each party is persuaded that it can extract an advantage from the discomfiture of others."¹⁹⁷ Thus, just as trust issues and traditional collective action problems prevented most collective balancing efforts from occurring in the first place, they also ruined the few positive attempts to put together an alliance against the Mongols. Given the Mongol enemies' manifest disregard for the approaching Mongol danger and their innate inability to trust their neighbors and ally amongst themselves, one must wonder whether they may not also have been more inclined to flock to the enemy despite the danger.

¹⁹⁶ King Bela of Hungary, 285-287; Turnbull 52-53.

¹⁹⁷ Phillips 80; Saunders 86, 88-89, 91-95.

4. Bandwagoning (IV4)

Besides the inability to coordinate balancing efforts, one further reason why balancing may fail is that potential balancers, instead of balancing, engage in exactly the opposite behavior and side with the rising hegemony. There is relatively little evidence of bandwagoning among the Mongols' targets. Few countries willingly joined the Mongol ranks, except in some isolated instances. While there were many examples of Mongol enemies cooperating with the Mongols once they were defeated and occupied and forced to become vassals, this does not constitute bandwagoning. Bandwagoning is restricted to voluntary collaboration with the Mongols and does not include coerced collaboration after conquest. Those that voluntarily cooperated with the Mongols were mostly not whole countries but rather individuals and sub-national groups lured by the Mongols' psychological tricks and deceitful reward promises. Most bandwagoners joined the Mongols for profit, and very few out of fear.

a. Fear (IV4.1)

Examples of bandwagoning out of fear are surprisingly rare. One would expect, given the extensive Mongol terror campaigns, that many would be scared into siding with the Mongols as a survival strategy, hoping the Mongols would spare them. But perhaps the Mongols' use of intimidation was so pervasive and shocking to their targets that they entertained no hope for survival even by bandwagoning. Indeed, if they took the descriptions of the Mongols that reached them seriously—depicting the Mongols as inhumane, devilish creatures—few would have been inclined to side with them.

Juzjani mentions the isolated case of one emir who likely bandwagoned with the Mongols out of fear during their conquest of Eastern Persia. Hoping to be spared, Sar-i-Zarrad of

Ghur "entered into an accommodation with the Mughals [i.e., Mongols]" just as they were about to attack him. Although his territory did not constitute a significant addition for the Mongols, they managed through flattery to turn the collaboration in their interest in another way. "Chingis Khān ... bestowed upon him [Sar-i-Zarrad] the title of Khusran [prince] of Ghūr, showed him great honor, and sent him back ... in order that he might, by means of accommodation, cause the other strongholds to be given up," Juzjani explains. Thus, Sar-i-Zarrad was instructed to undermine the other Khwarezmians who attempted to balance against the Mongols. This example also underscores the danger and futility of bandwagoning. After Jelal ad-Din's defeat, Sar-i-Zarrad realized that his alliance with the Mongols might have proved costly to the Khwarezmian by depriving them of additional power. But once he started balking in his support of the Mongols, Genghis Khan showed no mercy for his ally and had him promptly executed.¹⁹⁸

Some Armenian and Georgian princes also sided with the Mongols, resignedly convinced that the Mongols would win and hoping that cooperation would spare them, according to Akanc'. Their cooperation went even further, as they agreed to subject troops to the Mongols, which were likely used against their resisting compatriots.

When the wise princes ... realized that God was giving power and victory to them [the Mongols], to take our countries, then they became obedient to the Tat'ars [i.e., the Mongols], and agreed to give the tribute ... and to come out to them with their cavalry wherever they [the Mongols] led them.¹⁹⁹

Except for these few examples, though, the majority of cases of bandwagoning with the Mongols stemmed not from fear but from a calculated aspiration for gain.

¹⁹⁸ Juzjani 1006-1007.

¹⁹⁹ Akanc' 23, 29.

b. Profit (IV4.2)

The prospect of gain played a major role in motivating some local and national leaders to join forces with the Mongols. In general the expected profit was accrued power and/or territory at the expense of neighbors, another sign that local disputes were so prevalent that they justified any strategy, even virtually suicidal bandwagoning with the Mongols. To the bandwagoners' credit, though, the decision to side with the Mongols often resulted from the Mongols' own clever psychological pressure. In some cases, even personal defections to the Mongols, a less spectacular form of bandwagoning, greatly helped the Mongols' progress by undermining their targets' morale and balancing abilities. Quite a few military leaders switched sides and embraced the Mongols, anticipating their success and expecting rewards for their support. Interestingly, the East and Southeast Asian theatre accumulated most evidence of bandwagoning behavior.

1. East Asia

During the Kara Khitai campaign, two major groups in the kingdom bandwagoned with the Mongols. Under such circumstances the Kara Khitai resistance against the Mongol invasion, led by Naiman leader Kuchlug and his Naiman troops, was destined to fail. First, in 1209, a year before Kuchlug even took power in the Kara Khitai kingdom, the large Uighur Turk minority seceded from the kingdom because the majority Khitai, a declining force, were taking advantage of them and imposing particularly high taxes on them. The Uighurs decided to serve the Mongols and voluntarily became their vassals. In other words, the Uighurs bandwagoned out of profit because the Mongols promised a better deal. Then, a little less than a decade later, with the Mongol invasion of the kingdom underway, the Khitai population began undermining Kuchlug and sabotaging his efforts to repel the Mongols, because he had hijacked the leadership of the

kingdom and ruled ruthlessly. The Khitai, sensitive to the Mongols' appeals to resistance against Kuchlug, supported the Mongols to free-ride on their destruction of Kuchlug and his army. After the Mongols had caught and executed Kuchlug, the Kara Khitai became their vassals.²⁰⁰

Bandwagoning for profit was also rampant in the Chinese kingdoms. As a mirror image of the Kara Khitai, the Khitan, a vast ethnic majority in Northern China that was ruled by the Jin, succumbed to the Mongols' attractive promises of restoring their land taken away by the Jin. In 1213 the Khitan leader voluntarily submitted to the Mongols, along with about 100,000 troops, and combated alongside Genghis Khan against the Jin. The official Chinese records even mention that the Khitan gave the Mongols all the bounty they confiscated from the Jin, keeping none for themselves. In 1214, the Jin emperor, whose position was weakening, sent an envoy to the Khitan to request help. The annals report that not only the Khitan refused, but the Jin envoy also defected and joined the Mongol side. He was not the only Jin to switch sides. According to the annals, "the year 1215 proved even more grievous for the Kin [i.e., the Jin] because of the defection of Poussa Tsitsin, one of their best generals, who entered the service of the Mongols along with all the troops he commanded. Tchinkis-han [i.e., Genghis Khan], to encourage others to follow this example, made Poussa Tsitsin a Mongol general and rewarded all those who had followed him." The general in question passed major Jin military secrets onto the Mongols. The Khitan force and Jin deserters played a major role in the Mongol capture of the Jin capital of Zhongdu that same year. Eventually the Khitan obtained their reward; the Mongols returned their lands and reinstated their royal family on the throne, though as vassals. But since all of China was later incorporated into the vast Mongol state, the Khitan still did not obtain the sovereignty they were hoping for.²⁰¹

²⁰⁰ Phillips 52; Turnbull 16.

²⁰¹ Ssu-ma Kuang et. al., 50-51, 59-60, 68, own translation.

Once they had abandoned Northern China and retreated South to Kaifeng, the Jin were further undermined by their Southern neighbor's bandwagoning with their enemy. During Mongol general Mukhali's campaign against the Jin from 1221 to 1228, the Song colluded with the Mongols to defeat the Jin. They agreed to attack the Jin from the South while the Mongols attacked from the North, to tear the Jin apart in a two-front war. The Song were clearly bandwagoning for profit –they were engaged in a long-standing conflict with the Jin and saw their enemy's difficulty with the Mongols as an opportunity for their own territorial expansion at the expense of the Jin. The Song were also instrumental a few years later in helping Subudai ultimately defeat the Jin by winning the 1233 siege of Kaifeng. But bandwagoning generally comes at a price; because it reinforces the power of the rising hegemon, it may turn very dangerous, as the Song soon found out. When the Song emperor claimed former Jin land as a prize for having helped the Mongols, Ogodei and his successors conquered his kingdom and destroyed him.²⁰²

Korea also bandwagoned with the Mongols, at least at first, to extract local advantages, and once more, the Mongols were able exploit an ongoing local conflict, between the Koreans and their Khitan neighbors, to their advantage. In 1216, the Khitans invaded parts of Korea's Northern territory while fleeing from the Mongols. The Mongols thereafter presented themselves as Korea's defender and cleverly asked for the Koryo government's assistance in expulsing the Khitans out of their own land, in effect 'inviting' themselves onto Korean soil. The Koreans, who were prudent and understood the slyness of the Mongols, first limited themselves to providing the Mongols with supplies and watching them closely. But slowly their participation escalated and soon Korean archers joined in the fight to dislodge the Khitans from the peninsula. Gradually the Koreans became more involved and in 1218, a sizeable Korean army teamed up

²⁰² Phillips 67, 70-71; Weatherford 136.

with the Mongol army and expelled the Khitans. The collaboration proved costly, here too. It did not seem to be a bad deal at first for the Koreans, because the Mongols simply asked for a large tribute payment for 'saving' Korea and then left. In reality, however, the Mongols left parts of their army stationed at the Korean border to collect more tribute, while planning to come back and invade Korea entirely, which they did in 1231. Though the official Korean government switched their stance from bandwagoning to balancing, the damage had been done. In 1236, the Mongol advance was greatly helped by a high-ranking Koryo official who bandwagoned on his own and switched to the Mongols, handing Pyongyang over to the Mongols.²⁰³

2. Southeast Asia

The Southeast Asian theatre also experienced an unusually high number of bandwagoners compared to the rest of the Mongols' targets. When Kubilai sent expeditions down to Vietnam and beyond, several countries voluntarily submitted and sent tribute to the Mongols. Although fear might have played a role in their decision since most Southeast Asian countries were small and relatively weak, profit is their most likely motivation for bandwagoning. Most of these countries were interested in trade partnerships with the Mongols, and the exchange of tribute and gifts was a disguised form of commercial venture. Champa, the modern Vietnam, and the island of Java, joined the Mongol side at first before rebelling and expelling the Mongols. Java's Prince Vijaya bandwagoned to solve his domestic problems. Faced with an insurgency just as the Mongol expedition arrived in 1292, Vijaya sought to collaborate with the Mongols to obtain their assistance to defeat the rebels, and thus granted them tribute. The Mongols, fighting alongside the Javanese troops, managed to destroy the rebels and their fortified base. Realizing he did not need the Mongols anymore, Vijaya then ambushed them. Though not defeated, the Mongols left

²⁰³ Turnbull 35-36, 40.

the island to avoid further losses, in one of the rare instances where bandwagoning remained overall positive for the bandwagoner.

Other Southeast Asian countries and principalities sided with the Mongols without rebelling later. Malabar, on the Indian coast, is one example, and so is Ceylon, the modern Sri Lanka. Ceylon even willingly pushed the relationship with the Mongols one step further than most Southeast Asian states, who generally just agreed to be tributary to the Mongols. In addition to paying tribute, Ceylon formally accepted Kubilai as its suzerain, "no doubt with some hope of advantage" in mind, Phillips adds. Siam, the present-day Thailand, also went a step further for strategic reasons. King Ramkhamhaeng signed a treaty of friendship with the Mongols in 1282 and even personally visited Kubilai in China a decade later. Ramkhamhaeng was hoping that the presence of the Mongols in the region would keep his hostile neighbors Burma, Lanna (a Northern Thai kingdom) and Champa occupied, since they were at odds with the Mongols at the time. This would allow him to invade parts of the weak Khmer kingdom while his other neighbors had their backs turned. Bandwagoning with the Mongols enabled him to expand unnoticed for a while.²⁰⁴ In Southeast Asia, bandwagoning had less dire consequences than in the rest of the Mongols' target areas because of its difficult accessibility and uncomfortable, humid climate that kept the Mongols at bay.

3. Middle East

In Central Asia and Persia, bandwagoning was almost inexistent, and the few recorded bandwagoners were local, or smaller, groups. After Genghis Khan captured the main cities of Khwarezm and Shah Muhammad fled, several emirs sympathetic to Muhammad's Khitai mother joined the Mongols in order to accelerate Muhammad's defeat, al-Nasawi writes:

²⁰⁴ Weatherford 212-213; Turnbull 85, 87-89; Phillips 109.

Seven thousand Khitai belonging to the tribes of his [Muhammad's] cousins on his mother's side joined the Tatars [i.e., Mongols]. 'Ala ed-Dîn, prince of Qondouz, also went and accepted the authority of Djenghiz-khân and united with him to fight the Sultan [i.e., Muhammad]. Thus did also Emir Djâh Reri, one of the lords of Balkh. From all sides defection and treason multiplied.

Genghis Khan's new allies helped him pursue Muhammad to the island on the Caspian Sea where he eventually sought refuge. Similarly, decades later during the attack on Baghdad in 1258, the Christians in the city turned against the Muslims and joined the Mongols in looting and destroying the city, after the Mongols had promised to free them from Muslim dominion.²⁰⁵

4. Russia and Europe

Further West, Mongol targets were more tempted by bandwagoning. As Ibn al-Athir mentions, not only did the Mongols under Batu manage to breakup the alliance between the Kipchaks and Russian Alans in 1236-1237 by emphasizing to the Kipchaks how different they were from their ally, but they also tricked the Kipchaks into bandwagoning with them by stressing their common nomadic origins and offering them rewards. The Mongols reportedly told the Kipchaks: "We will promise you that we will not trouble you and we will bring you whatever money and clothing you want if you leave us to deal with them [the Alans]." Thanks to their deceptive tactics, the Mongols killed two birds with one stone—ridding themselves of a balancing coalition and transforming it into a partnership through the lure of profit. Yet, the fate of the Kipchaks once again demonstrated the futility of bandwagoning. The Mongols did not keep their promise, and in addition, "they slaughtered many of them [Alans], plundered and

²⁰⁵ Al-Nasawi 75, own translation; Weatherford 183-184.

made captives and then moved against the Qipjaq, who were feeling secure and had disbanded because of the peace that had been agreed," Ibn Al Athir concludes.²⁰⁶

The Christians in the West were periodically interested in bandwagoning with the Mongols for the purpose of defeating the Muslim world, whom they considered—wrongly, according to balance of power theory—to be their most threatening adversary. Some Christian princes already sought a common alliance with the Mongols against Islam under Ogodei's rule, having observed how the Mongols successfully defeated Khwarezm and Seljuk strongholds and thinking that the Mongols shared their animosity against Islam. They also hoped to open up Central Asia to Western influence and convert its population to Christianity. But the Mongols, not interested in religious rivalries or the promotion of Christianity but contemplating much larger, territorial goals, remained rather cold when an alliance promised little gain to them. For example, when in 1254 King Hayton of Armenia visited Mongke's court, requesting a Christian alliance with the Mongols against Islam, Mongke simply gave him a "diploma of safe-conduct and a letter of enfranchisement for the Christian churches" in his region, in order to pacify him. After the fall of the Caliphate in 1258, King Hayton renewed his efforts for an anti-Muslim alliance and volunteered 16,000 Armenian troops to Hulegu's army, which contributed to the campaign against Syria in 1259-60. King Hayton's interested generosity, again, did not pay off. When Hulegu left the Middle East in 1260 after the death of Mongke, the Muslims destroyed Armenia and re-captured most Christian outposts in the region, forcing King Hayton to abdicate in 1269. The Mongols did not do anything to help him.²⁰⁷

Western Christians failed to understand that the Mongols' conquest of Muslim lands was simply part of a larger expansionist scheme, and had nothing to do with their own religious

²⁰⁶ Ibn al-Athir 222.

²⁰⁷ Saunders 79-80, 105, 112-113, 117.

quest. As a result, they did not realize the looming threat posed by the Mongols, and time and again exultantly sought to collaborate with them. Several groups of Western crusaders had managed to retain outposts in small cities and castles on the coast of Palestine, Lebanon, and Syria. Like King Hayton, when the Mongols took the Arab heartland and the Caliphate in just a few years, they failed to see the danger the Mongols may pose to them and instead "saw an opportunity for themselves to ally with the Mongols and share in their victories," Weatherford argues. While the Mongols marched onto Syria from the East, the crusaders attacked Syria from the West to help the Mongols, and even took food and other supplies along with them that they pass on to the Mongols. But just like Armenia, the Mongols abandoned them to their fate and destruction by the Muslim armies shortly after.²⁰⁸

But the Christian attempts at bandwagoning with the Mongols went far beyond their outposts in the Middle East. Most of the diplomatic exchanges between various Western kings and Popes and the Mongol Khans starting in the 1240s aimed at a large, sweeping Christian-Mongol alliance. At first the Mongols were open to an alliance with the Christians, because it would create a welcome diversion in the Middle East that could facilitate their own conquests. For example, Mongke reasoned that if the forces of French king Louis IX invaded Egypt, it would occupy the Mamluks there and keep them from coming to the assistance of the Caliph of Baghdad, whom Mongke was planning to attack. However, the first exchanges—between Pope Innocent IV and Guyuk in 1245-47 via Carpini, between Innocent IV and Baiju in 1247-49 via the Dominican Ascelin of Lombardia, and between Saint Louis (Louis IX) of France and Mongke via Rubruck 1253-55—utterly failed because the Western message insisted upon the conversion of the Mongols to Christianity, while the Mongols demanded immediate submission. Even after the Mongols made clear their intentions to subject the entire world to their rule, the

²⁰⁸ Weatherford 184; Turnbull 58.

Europeans still persisted in forming an alliance with the goal of taking Jerusalem, Syria, and Egypt from the Mamluks.²⁰⁹

The next initiative came from the Ilkhan rulers, who were closer to the West than Mongke's successor Kubilai and also more open to cooperating with the Christians against the Muslims as they had started to slowly drift away from the central Mongol Empire and were thus more in need of reinforcements than Kubilai. Ilkhan ruler Arghun sent Mongol Nestorian monk Rabban Bar Sauma as his diplomatic representative on a tour of European capitals in 1287-88 to evaluate the possibilities of an alliance, hoping that the Europeans would still want to profit from Mongol support in tackling the Muslim world. Sauma, in his travel memoirs, writes that he obtained many positive reactions and support for an alliance, but that in the end, the alliance project did not come to fruition. After visiting King Andronicus II of Constantinople, Sauma went to France and met with King Philippe IV "Le Bel," who pledged his support. "If it be indeed that the Mongols, though they are not Christians, are going to fight against the Arabs for the capture of Jerusalem, it is meet especially for us that we should fight [with them]," King Philippe assured Sauma. King Edward I of England also promised his allegiance, though in a more muted tone: "My mind is relieved on the subject about which I have been thinking [Jerusalem], when I hear that King Arghôn thinkest as I think." Pope Nicholas IV, whom Sauma met on Easter 1288, showered Sauma with gifts but seemed surprisingly less interested in a common military mission. After the mitigated results of these meetings, both Arghun and the Europeans abandoned the idea of an alliance, though exchanges continued.²¹⁰

²⁰⁹ Phillips 80-81.

²¹⁰ Rabban Bar Sauma, *History of Rabban Sâwmâ and Yahb-Allâhâ III* (ca. 1290-1294), transl. into English (*The Monks of Kûblâi Khân, Emperor of China or the History of the Life and Travels of Rabban Sâwmâ*) by E.A. Wallis Budge (London: The Religious Tract Society, 1928; reprint by AMS Press, New York, 1973), 165-168, 182-183, 186, 190-191.

Overall, bandwagoning left a mixed impact on the Mongols' hegemonic quest. While some bandwagoning clearly occurred and may have occasionally contributed to the Mongols' easy rise, bandwagoning cannot be considered a major factor in the Mongols' achievement of hegemony for two reasons. First, bandwagoning remained sparse, and was most often undertaken by non-critical actors, such as weak, remote, or local groups, and not by the actors that could have made a major difference against the Mongols. Second, bandwagoning covered the map irregularly, occurring frequently in some regions and not at all in others. The regions devoid of bandwagoners did not fare any better than the regions where bandwagoning was common, which leads us to conclude that bandwagoning was not a crucial causal factor explaining balance of power failure.

In fact, the resisters' actions—or lack thereof—only constitute a partial cause of balance of power failure. Collective inaction, deception, laggard balancing, and bandwagoning account for the absence of real resistance against the Mongols' rise, but alone they cannot explain the unprecedented scope of the Mongol conquests. As the following chapter demonstrates, the exceptional skills of the rising hegemon significantly contributed to its success in transcending the balance of power. The Mongols' ability to match and surpass both the military and non-military skills of their targets was without doubt a vital cause of their hegemonic success. The tremendous balancing hurdles faced by their targets were necessary factors to ease their rise, but without special talent of their own, the Mongols would not have been able to create such an unprecedented empire.

The Mongol Conquests II The Rising Hegemon

The Mongols' unique talent laid partly in their aptitude to innovate in a variety of fields, which gave them an unconditional advantage over their opponents, but also in their unparalleled propensity to learn from other states and civilizations, embracing and replicating at once what they deemed useful and improving upon other aspects with their own unique twist. After all, the Mongols started their rise from scratch. As a collection of conflict-prone, family-based clans of herders and hunters, they started their ascent devoid of almost any basic technology and large-scale military, political and social organization, and managed to conquer far superior, long-established civilizations within a few decades. Such a momentous evolution requires above-average innovative skills and adaptability, both in the military arena and with regards to non-military aspects such as socio-economic and administrative structures. In fact, the Mongols' distinctive nomadic background gave them the flexibility to design their own unique military techniques and unprecedented state-building efforts that gave them a decisive edge over their competitors.

1. Military Innovation (IV5)

On the military side, the Mongols gained the advantage through an amalgam of clever innovations—in logistics and battle techniques, based on their nomadic traditions, and in army

organization and communication—and useful adaptations and improvements—mostly in weaponry, thanks to the purposeful incorporation of foreign expert advisers into the army leadership and corps of engineers. Thus combining their innovative skills and their versatility, the Mongols created an unmatched, state-of-the-art military structure that left them always one step ahead of their targets, who were utterly unable to catch up or cope with the Mongols' constant progress.

a. Advantage of the Steppes: Logistical Innovations and Adaptability to Foreign Theatres

The Mongols' distinctive background as nomadic hunters and herders enabled them to create an army of horsemen devoid of the traditional armies' usual logistical hassle and whose swiftness and effectiveness was unmatched. The first innovation stemming directly from the Mongols' nomadic lifestyle was the unique composition of the Mongol army—a cavalry made up of dexterous, self-sufficient warriors. Unlike every army they encountered, the Mongols had no infantry. Instead, every soldier rode, and the army was thus able to cover large distances in record time. Every Mongol, girl and boy, grew up as a warrior and was taught to ride a horse and handle a bow and arrow from an early age, since hunting was a necessity to procure food and survive on the steppes. Since the Mongol society did not engage in any division of labor, every Mongol man was a herder, a hunter, and a warrior, resulting in a highly skilled, professional army that needed very little to operate. Even when women and children occasionally followed the army into a campaign, they never represented a hassle because they could defend themselves.

On the warpath every warrior carried everything he needed to survive on his own, and nothing superfluous: clothes designed to help him withstand any weather included a wool robe and pants, a fur hat, and resistant leather and fur boots; other necessities besides weapons were

flints for fire, a leather water container, sharpening files for arrowheads, a rope, a sewing kit to mend clothes, a knife and a hatchet, all carried in a waterproof skin bag. Besides these individual items, all the Mongols needed on their campaigns was a supply of extra horses to relieve their tired mounts, usually about a dozen per warrior. The mares would also provide milk, which the Mongols fermented or dried and thus never lacked. Besides dairy products the Mongols would eat hunted animals or animals from their herds that followed them on campaigns—fresh or dried—and thus never lacked food. As Carpini describes it, "they subsist entirely upon flesh and milk."²¹¹

Because of these arrangements, they never required a supply train, which was traditionally long and cumbersome and constituted a logistical nightmare for any army at the time. Soldiers from settled, agricultural societies needed to be followed by a whole convoy of food supplies and were heavily dependent upon it, which put them at a disadvantage if they were temporarily or permanently separated from the main army, and also considerably slowed down their progress. On the contrary, Ibn al-Athir writes, "the Tatars [i.e., the Mongols] do not need a supply of provisions and foodstuffs, for their ... horses, and other pack animals accompany them and they consume their flesh and nothing else." The absence of supply train also made the traveling armies look much more inconspicuous, as they were often able to pass as herders or hunters and thus attack in surprise. They did not even need to carry fodder for their horses and other animals, since their short, sturdy steppe horses fed themselves by grazing and were used to provide for themselves even in the winter by digging up their own grass under the snow and ice. As Marco Polo writes, "their horses are fed upon grass alone, and do not require barley or other grain" like horses from other cavalries. Ibn al-Athir confirms that "the animals they ride dig the earth with their hooves and eat the roots of plants, knowing nothing of barley." The Arabic

²¹¹ Marco Polo 89; Weatherford 86-87; Phillips 32.

scholar concludes: "thus, when they [i.e., the Mongols] make a camp, they require nothing from without."²¹²

A diet of dairy and meat carried many more advantages besides outmoding the supply train. It was also easier to transpose to other parts of the world than a diet of grain, which may not always be available elsewhere and would then require time and effort to grow. Moreover, the Mongols' nomadic lifestyle made them independent of crops and harvests, while their opponents were often seriously hurt by their reliance on agriculture. The Chinese annals mention, for example, that the Jin society and conscript army was composed of hordes of peasants who ate mostly rice. When in 1212, in the midst of the Mongol invasion, a severe famine hit several Chinese provinces, the price of rice soared, the annals explain, and "the countryside was littered with the corpses of the countless underprivileged who died of starvation." Food shortages constituted a serious disadvantage for agricultural societies that the Mongols never experienced, and it proved particularly dire in war-time when the enemy did not face the same problem.²¹³ Furthermore, even as some witnesses report with disgust that the Mongols preferred rodents to bread or rice, the dairy and meat diet reinforced the Mongols' physical resistance and increased their strength. Because of the high protein-content of their food intakes, the Mongol warriors were able to survive on relatively little amounts of food for longer periods of time and were considerably leaner than the cereal-fed men they fighting against. As Marco Polo points out:

They [i.e., the Mongols] are capable of supporting every kind of privation and ... can live for a month on the milk of their mares, and upon such wild animals as they may chance to catch ... No people upon earth can surpass them in fortitude under difficulties, nor show greater patience under wants of any kind. They are ... maintained at small expense.

²¹² Ibn al-Athir 203; Marco Polo 89-90.

²¹³ Ssu-ma Kuang et. al. 50, own translation.

From these qualities so essential to the formation of soldiers, it is, that they are fitted to subdue the world.²¹⁴

Just as the Mongols' nomadic diet constituted a major improvement upon the traditional complications of army logistics, the portability of their dwellings also made the Mongol warriors more mobile. All Mongols lived in transportable felt tents. Generally, the tents were organized by family, with one tent for each mother since the Mongols were polygamous. Carpini describes the tents' mobility: "Some can be speedily taken down and put up again and carried on baggage animals; others cannot be taken down but are moved on carts." When on campaign, however, the Mongol warriors shared one tent per ten warriors and used the most portable and easily packable tents, which allowed for even quicker movement. But when necessary, they also knew to comfortably rest without a tent. Marco Polo mentions that "the men are trained to remain on horseback during two days and two nights, without dismounting; sleeping in that situation while their horses graze." When a covert advance was required in the middle of an invasion, they often simply spread out in the countryside at night, eating dried meat and dried mare's milk that did not even require lighting a fire. The army was able to blend into its surroundings and move with unusual ease.²¹⁵

Their nomadic roots and exclusive reliance on a cavalry thus gave the Mongols unique mobility, reactivity, and efficiency that other armies did not have and made them particularly adapted for long-distance travel and battle in foreign lands. Even though the Mongols almost always faced much larger armies than their own, their mobility and efficiency gave them a decisive advantage. Thanks to the self-sufficiency of the nomadic lifestyle, the Mongol warriors were adaptable to any environment and able to fight across huge distances, in drastically

²¹⁴ Marco Polo 91-92; Weatherford 88.

²¹⁵ Carpini 8; Weatherford 87.

different cultures, and over very long periods of time—sometimes decades-long campaigns. Not only were the Mongols not handicapped by an infantry, but they also did not allow their pace to be slackened by heavy artillery. Even though they recurrently needed siege engines to subject cities, they never carried them along on a campaign. Instead, the army always traveled with a corps of engineers, who would custom-build artillery pieces on site shortly before reaching the destination, after evaluating the need and using logs cut down by the warriors.

This innovative, nomadic army arrangement was particularly successful because it made other armies obsolete. While the Mongols were lifelong warriors, archers, and riders, they faced anachronistically ill-equipped enemies—mostly slow, undisciplined assortments of conscripts and mercenaries who combined a lack of experience, motivation, and nourishment. Peasant conscripts were often not even given appropriate weapons, and attempted to fend off the adroit Mongol archers with whatever farm utensils they owned, be it a simple club or a pitchfork. Those who did have a proper weapon were not used to handle it and were thus just as inefficient as their counterparts without weapons. Since the Mongols' opponents relied mainly on infantries, they constituted easy targets for the Mongol archers on horseback who towered over them. They could not defend themselves against the Mongols, who simply remained at a distance and sent "hales of arrows" on them, as the Chinese annals describe it. The annals underline that the Mongols always attacked collectively in whole regiments and never engaged in individual combat; they did not need to since they were archers. In fact, the Mongol warriors never moved close enough to their enemies to be in reach for one-on-one combat. Many armies, especially mercenary armies, were expecting individual combat, however, and the result was catastrophic for them. The European armies, for example, mostly stood paralyzed not knowing what to do and waited to be showered by arrows. The Japanese samurais, with their long tradition of carefully

regulated personal sword combats, were similarly confused by the storming Mongol horsemen. Although the samurais engaged in a few heroic attacks on the Mongols, their army would have been useless if the Mongols had undertaken a large scale invasion of the Japanese islands.²¹⁶

The rare archers that the Mongols' enemies had among their troops generally used shorter-range, much less powerful bows than the Mongols and were therefore easy to beat. The few armies that used horses were likewise helpless against the Mongol warriors. They were for the most part small cavalries—often just the high officers were on horseback—riding large, heavy, slow, and awkward horses meant to display force and carry heavily armored and decorated riders. They could not contend with the smaller, faster, and better-trained Mongol steppe horses. Thus, the main reason the Mongols were able to rise with such rapidity, medieval historian J.J. Saunders argues, is because "an army of light cavalry, of trained professional mounted archers, could sweep from the field the clumsy levies of ill-equipped ... peasants, mere temporary soldiers bereft from their farms, who for long constituted the defense forces of civilized states." As Ibn al-Athir concludes, "one of the local inhabitants said to me [during the campaign in Khwarezm against Jelal ad-Din]: 'If we had had five hundred cavalry, not one of the Tatars [i.e., Mongols] would have survived'." It is clear that the logistical and military superiority of the Mongols' innovative cavalry-based army, enhanced by their mobile nomadic lifestyle, left their opponents no chance.²¹⁷

The mobility and resistance their army, achieved thanks to their nomadic roots, also allowed the Mongols to tackle challenges no one else could surmount. For instance, it made winter an asset rather than an obstacle for them. Where other armies were decimated by ice, snow and cold, the Mongols thrived, as they were used to surviving outside on the frigid steppes.

²¹⁶ Ssu-ma Kuang et. al. 50, own translation; Turnbull 65-66.; Weatherford 140-141.

²¹⁷ Ibn Al Athir 307; Weatherford 140-141; Saunders 28.

In fact, the Mongol army often waited for winter to attack tough enemies or cross difficult areas such as deserts or mountains because in the winter they needed to carry less as men and horses used less water, allowing for even lighter and faster travel. The winter also drew game further out where it was easier to catch. For example, during the Hungarian campaign, Subudai chose to attack in the middle of January when the opponent was least expecting it. The Mongols were also the only army in history to successfully invade the Russian heartland, where other major would-be hegemony like Hitler and Napoleon failed because of the winter weather. Batu and Subudai again purposefully chose to initiate the invasion of Russia in the winter and to start with a Northern route—preferring to travel over frozen steppes rather than during the soggy and flooded spring and summer months. The Mongol army easily adapted to any winter situation. During the campaign against XiXia in September 1226, for instance, Genghis Khan's army met the Tangut army by the flooded banks of the Huang Ho (Yellow) River, which had entirely iced over and become an ice plain. The Mongols simply attached cloth slippers to their horses' hoofs to prevent them from slipping, and easily crushed their dismayed opponent.²¹⁸

Growing up as nomads on the steppes instilled the Mongols with an exceptional capacity to survive and adapt to very different, changing, and challenging circumstances and military situations. The hardest test for their adaptability skills was without doubt Kubilai's Southeast Asian campaigns, which constituted the southernmost expansion of the Mongols. In Southeast Asia they were confronted with a hot and humid climate and heavy rains, thick jungles, exotic animals and insects, and unfamiliar diseases. Yet, even though Southeast Asia was entirely different from anything they had ever encountered, it did not stop them from attempting the conquest, though their success was mixed in the end. Problems with adaptation to hot and humid climate started already during Mongke's campaign against the Song some time earlier, when the

²¹⁸ Phillips 65-66; Saunders 84; Turnbull 45, 49; Weatherford 4.

Mongol army was reportedly hit by epidemics of diarrhea, very likely due to dysentery. Mongke himself died very suddenly during that campaign in August 1259, and Persian accounts speculate he may have succumbed to the disease. The Mongols eventually surmounted their unfamiliar surroundings, and although it may have taken more time and efforts than in other campaigns, they successfully defeated the Song in 1276.²¹⁹

Unlike other armies of the time, the Mongols' logistics were not set in stone and they integrated changes that could improve their efficiency without the usual inertia of other armies. They were even open to fundamental changes. A perfect example is the introduction of elephants in combat, a major change for an army riding steppe horses. Juvaini provides the first mention of elephants in a Mongol campaign when he describes the siege of Samarkand, where he says Khwarezmian soldiers used a few elephants. Genghis Khan, having never heard of elephants before, took the animals prisoners then released them to graze with his horses, but with no special fodder or trained keeper to care for them, they starved and died.²²⁰ The next encounter with elephants occurred much later, in the mid-1270s, against the King of Burma. Marco Polo gives an enlightening account of how Kubilai's warriors, despite having never faced such monsters before, acted with considerable calm and ingenuity instead of panicking, even though the Burmese were also much more numerous than them—roughly 60,000 to 12,000:

The Tatar [i.e., Mongol] horses, unused to the sight of such huge animals, with their castles [baskets carrying soldiers atop the elephants], were terrified, and wheeling about endeavored to fly [i.e., flee]; nor could their riders by any exertions restrain them ... As soon as the commander perceived this unexpected disorder, ... [he] order[ed] his men to dismount and their horses to be taken into the [nearby] wood, where they were fastened to the trees.

²¹⁹ Turnbull 82; Weatherford 187-188.

²²⁰ Juvaini 119-120.

Continuing on foot, the Mongols bombarded the elephants with arrows:

It soon became impossible for their drivers to manage them. Smarting under the pain of their wounds and terrified by the shouting of the assailants, they were no longer governable, but without guidance or control ran about in all directions, until ... they rushed into ... the wood ... The consequence of this was, that the branches of large trees wrecked the ... castles that were upon their backs, and destroyed those who sat upon them.

The confusion gave the Mongols new strength, and despite the bad odds in numbers, they were able to win the battle and obtain tribute from Burma. Instead of executing the dangerous animals, Kubilai's men captured over two hundred of them, learned from the Burmese how to care for them, and brought them back to Kubilai. "After this battle the Great Khan has always employed elephants in his armies," Marco Polo concludes. His account is corroborated by other stories, which show that the Mongols not only had no problem adapting to the use of elephants, but even employed them in innovative ways. Under Kubilai, the elephants carried only one driver and one warrior, who would operate a bow or crossbow from the castle. The animals were not simply used as vehicles, but also taught to engage in fighting themselves. The Mongols trained them to trample enemy soldiers and sweep others off the ground and throw them. After the Burmese battle, the Mongols widely used elephants for combat in the rest of Southeast Asia, and also on the Chinese mainland. One account describes Kubilai setting off to war against a Mongol renegade, Nayan, in 1287 on a palanquin set on the back of an elephant.²²¹

Another example of a substantial logistical adaptation is the Mongols' embrace of naval warfare. Coming from an area far remote from the ocean, where small, easily fordable rivers constituted the only water obstacles, the Mongols faced at first tremendous difficulties in front of

²²¹ Marco Polo 199-203; Turnbull 84, 87.

water, which some historians have even termed an innate "fear" of water. But here again, the Mongols eventually showed the ease at learning and adapting to novelty that they inherited from their self-sufficient, itinerant culture. At first, water constituted a real weakness for the Mongol army. When Khwarezm Shah fled onto an island on the Caspian Sea with a few followers in 1220, the Mongols were unable to follow. Ibn al-Athir describes their vulnerability to water: "When he [Khwarezm Shah] and his men embarked on the boats, the Tatars [i.e., Mongols] arrived and, seeing that Khwārazm Shāh had put to sea, they halted at water's edge. Despairing of catching up with him, they withdrew." Nasawi adds that they helplessly tried to shoot arrows at the fleeing boat, albeit unsuccessfully. Similarly, when Jelal ad-Din crossed the Indus River after having been defeated by the Mongols the following year, Genghis Khan had to abandon the chase and was reportedly very impressed by Jelal ad-Din's prowess in the water. During the invasion of the Jin, the Mongols were similarly forced to let their enemies escape after they boarded boats on the Yellow River.²²²

The Mongols were likewise helpless during the invasion of Korea in 1232, when the Koryo government escaped from the capital of Kaesong and sought refuge on the fortified island of Kangwha. Although the water between the mainland and the island only covered a half a mile's distance, the Mongols were immobilized by the obstacle and, unable to dislodge the Koryo government, had to withdraw. In Europe and Russia they encountered fewer issues with water because of the winter timing of the invasions, which enabled them to cross most rivers when they were frozen. But the problem resurfaced during the pursuit of King Bela, who got on a ship upon reaching the Adriatic, sailed along the coast to Trau, in Croatia, and simply anchored offshore when he realized that the Mongols were blocked by water. Batu's army was unable to dislodge him, and had to abandon the pursuit. Finally, when a contingent of Kipchaks found shelter on an

²²² Ibn al-Athir 211; Al-Nasawi 107-108.

island on the flooded Volga, Batu's army was forced to wait for the water to subside before crossing the river, although the Kipchaks had had no problem crossing earlier.²²³

Undaunted by these setbacks, the Mongols sought a way to overcome their water problem. They understood the need to acquire a navy and learn about the sea and naval warfare and were conscious that this constituted an area of weakness. Thus, they bounced back by taking over their vassals' and defeated enemies' fleets—Jin, Song, and Koryo—and by improving upon them to continue expanding once they had exhausted virtually every land venture. Upon reaching the Jin's Pacific shore they immediately confiscated all vessels they found. Similarly, when they won one of the first battles against the Song at Diaoyu (Szechuan province) in early 1265, they captured 146 Song ships and rapidly started to copy them and build more. The Song navy was soon transformed from a simple coastal defense navy into a high seas commercial and military fleet. The Mongols also forced the Korean king to surrender his naval force, including vessels and sailors, in 1274, and Korea was thereafter used for shipbuilding and as another naval base of operations from which to conquer Japan and the islands of Southeast Asia. Kubilai commissioned nine hundred warships from Korea for the expedition to Japan alone, which were operated in majority by Korean sailors. For the second expedition to Japan, Kubilai also ordered six hundred extra ships from the Song. Although part of the Mongols' lack of success in Japan and Southeast Asia may be due to their inexperience with naval operations, the Mongols certainly learned from their early mistakes and adapted to naval warfare quickly for horsemen from the steppes with no prior knowledge of the sea.²²⁴

Overall, the Mongols' unique natural skills as nomadic herders and hunters gave them a logistical superiority over their opponents that transcribed in a highly mobile, effective, and

²²³ Saunders 61, 81-88; Turnbull 33, 49, 51.

²²⁴ Weatherford 210; Saunders 122; Turnbull 39, 61, 65-72.

resistant army particularly well-suited for the wide-range of conquests it undertook. Their distinctive background as self-reliant nomads dealing with the harshness of the steppes lifestyle also gave them the versatility and resourcefulness to adapt and improve under unfamiliar circumstances, which underscores their constant willingness to learn and subsequent reliance on foreign advisers.

b. Learning and Foreign Input

The key to the Mongols' unusual adaptability and incessant military progress lay in their interest in learning and borrowing from other cultures and civilizations. Instead of being protective and wary of interactions that would make others aware of their assets, they sought any foreign contact they could benefit from. Because they started with virtually no technological and other material knowledge, the Mongols necessarily needed inspiration from abroad to grow and maximize their military abilities. Thus, they continuously used foreign experts and workers to improve their skills, and also their numbers, always learning the best strategy and weaponry from allies and enemies alike. They soon filled in their gap with other leading powers and became increasingly efficient with new weapons and techniques of all kinds. Yet, those same allies and enemies whom they learned from made no efforts to simultaneously learn from the Mongols, and hence were unable to catch up to the Mongols as the Mongols had caught up to them. The Mongols used three types of foreign advisors: technical experts and engineers to advance their military technology; high-ranking officers and leaders to form the best military and political strategies; and scholars, interpreters, and skilled craftsmen to understand and foresee the enemy better and beat it at every advantage it might have had.

1. Technical Experts

Foreign engineers were particularly prized by the Mongol leaders because they lacked technical knowledge about the latest weapons. Chinese engineers excelled at siege warfare and had invented new artillery to defeat fortified towns: regular catapults throwing various projectiles like stones, liquids and flames; more advanced trebuchets, which were counterweight catapults that could throw projectiles faster and farther; and ballistas, massive crossbow-catapults that threw arrows at great distances with remarkable impact. They also invented gunpowder and from that derived a whole array of new weapons, including the thunder-crash bomb, the precursor of modern bombs, and the fire-lance, a rudimentary bamboo tube filled with gunpowder—which were both highly effective in frightening enemies into surrendering. Genghis Khan and his successors did not simply appropriate these weapons, but they imported the engineers that could create them and thus constantly upgraded the weapons. For example, thanks to the efforts of their Chinese, Persian, and European engineers who used a combination of Chinese gunpowder, Middle-Eastern flame-torch, and European bell-making techniques, the Mongols created an explosive device that was the ancestor of the cannon and would have a vast impact on modern weaponry. This kind of innovation was only possible because of the Mongols' unique reliance on a diverse pool of foreign experts. It could only have been created by their hybrid, transnational corps of engineers combining their respective techniques.²²⁵

The Mongols made a deliberate effort to lure—or force, if necessary—foreign technicians into their army. They provided various rewards for enemy engineers defecting to the Mongols, and incentives for vassal engineers to join them. They also singled out engineers among captives to be spared and brought along with them. Genghis Khan already managed to bring massive numbers of foreign engineers onto his side. Mukhali, for example, who was in charge of the Jin

²²⁵ Weatherford x-xi, 94.

campaign under Genghis Khan's leadership, had an army composed of almost one third of non-Mongols, mostly Khitans and Chinese, and mostly engineers. Such a high proportion of foreign advisors in the Mongol army was not uncommon. Rashid al-Din describes how in preparation for his two simultaneous campaigns against China and the Middle East, Mongke first assembled a solid corps of foreign engineers. Mongke "sent couriers into Kathāi [i.e., Cathay, the former Jin territory] to bring from that country one thousand war-machine [i.e., siege engine], oil projectile and crossbow experts and their families," Rashid al-Din writes.²²⁶ Marco Polo also famously tells us how his father and his uncle, who had some knowledge of Western weapons, personally offered to lend their technical skills to Kubilai to help take over the key city of Sayan Fu (the modern Xiangyang) from the Song around 1270. The approach to the city was difficult because it was surrounded by water on all but one side. Niccolo and Maffeo Polo helped construct a siege engine "such as were made use in the West, capable of throwing stones three hundred pounds in weight," Marco Polo writes. They built a mangonel, a type of catapult with a lower trajectory but more speed than a trebuchet and meant to destroy walls, and the city was defeated.²²⁷

2. High-Ranking Leaders

In addition to technical experts, the Mongols also sought to attract high-ranking, valorous military and political leaders that could provide them with advice on military and political strategy. They frequently hired defeated enemy leaders to help them plan further offensives, obtain inside-information, and learn more about their enemies' practices. Juzjani reports that Genghis Khan already had a Muslim leader, Ja'far, as a military advisor in the late 1100s and early 1200s when he was battling the Altan Khan, leader of the Tatar tribe. Ja'far played a crucial

²²⁶ Rashid al-Din 129, 133, own translation; Phillips 66, 88.

²²⁷ Marco Polo 221-222.

role in defeating the Tatars. After the sack of Kiev in December 1240 during the second Russian campaign, Batu decided to spare the life of Dmitri, the prince of Kiev, hoping to win over a leader who had shown great bravery in defending his city. But when a leader refused to join the Mongols, they did not hesitate to coerce their cooperation. During the campaign against the Jin under Genghis Khan, Mukhali simply captured a Jin general he wanted on his staff. The general later gave Mukhali valuable advice, encouraging him to limit pillage and slaughter to gather more popular support.²²⁸

Some foreign leaders were even at times entrusted with major positions in the Mongol army. Waley mentions in his introduction to Ch'ang Ch'un's travels, for example, that two Jin generals joined the Mongol ranks in 1219, Shih T'ien-hsiang and Li Ch'i-ko. Sometimes foreign leaders were even entrusted with directing combat operations. During the conquest of Khwarezm in the early 1220s, Juzjani writes, "Arsalān Khān of Kaiālik, who was a Musulmān, and had [under him] about six thousand [horse]men, all Musulmāns and 'Ajamis, along with ... a Mughal [i.e., Mongol] force, was sent against the fortress of Walkh of Tukhāristān [i.e., Turkestan]." Even more significantly, in 1277, Kubilai named Nasir Al-Din, son of a Muslim governor already at the service of Kubilai, to command the Mongols' military operations against the capital of Burma, Pagan. But while incorporating foreign leaders liberally in the civil and military administration of the empire, the Mongols also always made sure to retain close, loyal Mongols for top posts, even though foreign advisors were generally loyal to them. A nomination as prominent as Nasir Al-Din's remained a rare occurrence.²²⁹

²²⁸ Juzjani 954; Phillips 67, 74.

²²⁹ Arthur Waley 5, 7; Juzjani 1004; Turnbull 84; Phillips 109.

3. Scholars and Craftsmen

Finally, and perhaps most importantly, the Mongols were able to appropriate new ideas and techniques and at the same time learn about their targets, and about other civilizations in general, by welcoming prominent scholars and craftsmen from abroad into their ranks. Learning from and about the enemy in every possible way was crucial to Genghis Khan, who never launched an attack or a campaign without clear knowledge of not just the military power, but also the human, economic, social, and psychological strength of the target. Along the way, he picked up anything new that could be appropriated by the Mongols to increase their own abilities. In addition to their vast spy system, the Mongols expanded their knowledge of others by inviting foreign scholars to teach them. Genghis Khan initiated this practice by bringing back a number of Khitan scholars at his court after his victory over the Northern Chinese capital of Zhongdu in 1215.²³⁰

Among those scholars, Yelü Chucai played a particularly significant role in the evolution of the empire. Very impressed by his extensive knowledge of astrology, Chinese literary works, various languages, and Chinese culture, law, and tradition in general, Genghis Khan offered him to work for him. As Phillips stresses it, this had "important consequences for China" because the Mongols showed genuine interest in learning and took the advice from foreign scholars very seriously. For example, Chucai convinced Genghis Khan that the Mongols would benefit more from exploiting Northern China's economic potential in iron and salt production and raising taxes on it than from transforming the area into grazing pastures as Genghis Khan had planned. Though the concept of economic development was alien to the Mongols, Genghis Khan adopted Chucai's plan and transformed Northern China into a very lucrative district. He named Chucai the governor of that district and slowly Chucai was able to instill thriving notions of Chinese law

²³⁰ Weatherford 90.

and administrative practices into the Mongol empire. As René Grousset mentions, Chucai argued that though the Mongols' nomadic lifestyle were most efficient for conquest, it was unsuited to administer an empire of that size. After Genghis Khan's death, Ogodei continued to rely on Chucai and introduced more reforms he suggested, mostly to further enhance the civil administration of the empire. Chucai was only one example of a foreign scholar playing a major advising role, though. Ogodei also relied on Uighur scholars like Chinkai and the Khwarezmian scholar Mahmud Yalavach, who later became governor of the Transoxania region, Grousset writes. Kubilai similarly chose to surround himself with Chinese scholars, having himself been taught by Confucian scholar Yao Chi as a child, whom he later consulted before leaving on a campaign.²³¹

Foreign scholars were used not only as political advisors, but also more broadly as teachers and translators. Through this task, Genghis Khan's Uighur scholars came up with a Mongol alphabet based on the Uighur script, used in military communication and archiving and which has been used in Mongolia to this day. On every campaign, the Mongols always brought interpreters conversant in the target's language and customs along with them. Already in 1219 during the Khwarezmian campaign, Phillips writes, "Moslem interpreters familiar with the West were assigned to all divisions and intelligence was diligently collected." Kubilai went one step further and appointed Chinese scholar Chao Pi to teach the Mongol soldiers who were heading for the Song campaign about Confucian doctrine and Chinese traditions. Kubilai even made him learn the Mongolian language to help him teach better. Marco Polo also describes that during his stay at the Mongol court, there were scholars and translators from virtually every corner of the world, with knowledge of every possible language, working for the Mongols.²³²

²³¹ Grousset 250-252.

²³² Phillips 61, 86.

Besides scholars and interpreters, the Mongols also insisted upon bringing foreign craftsmen along with them and learning their practices. Even if craftsmen seem less directly related to improving the Mongols' military skills, they played a key role by enabling the Mongols to learn more about their targets and to develop new techniques, such as metalwork or ceramics, that they then applied to combat. Learning from foreigners on all occasions was clearly a Mongol tradition, Turnbull argues. When taking prisoners after a defeat, the Mongols would carefully set aside any craftsman with unique skills to take along with them. "Taking away the craftsmen ... was customary," Turnbull concludes. Thus, the Mongols 'imported' alternatively some German miners, Chinese doctors, Middle-Eastern architects, a Parisian metalworker commissioned to build a massive silver fountain, an English aristocrat who became the army's translator, etc. This hybrid accumulation of diverse people with special skills was an incredible vector of progress for the Mongol army and builder of civilization for the Mongol empire.²³³ Overall, the Mongols understood that being inclusionary, incorporating enemy practices and welcoming enemy leaders, experts, and scholars not only taught them what they were missing in military technology, strategy, and assets and propelled them beyond their enemies, but it also increased the loyalty of defeated enemies and vassals and diminished the likelihood of insurrection and balance of power behavior. Loyalty constituted the cement at the basis of the Mongol army and empire, and the Mongols' innovative army organization was specifically designed to enhance it.

c. Decimal Army Organization (IV5.3)

One of the Mongols' most dramatic innovation resided in their new army organization, which significantly contributed to the army's effectiveness and its superiority to its targets.

²³³ Turnbull 76; Weatherford x-xi.

Before even becoming Great Khan, Temujin organized his army of professional horsemen along a strictly hierarchic decimal system. The *Secret History* dates the new system back to 1204, and explains that Temujin set it up before taking on his last major challenge on the way to uniting the Mongol tribes, the Naimans. According to the *Secret History*, Temujin "counted [his] soldiers and formed them into thousands. He appointed the leaders of thousands, the leaders of hundreds, the leader of tens and the six cherbis [leaders of ten thousands]." In other words, the Mongol army was organized in units of tens with each one commander; ten units of ten were led by a commander of one hundred; ten units of one hundred by a commander of a thousand; and ten units of a thousand by a commander of ten thousand. Each soldier obeyed his immediate commander of ten, who was supervised by his commander of a hundred, and so on up to Genghis Khan, who had under his direct command the commanders of ten thousand. The clarity and simplicity of the chain of command made for easy communication and order transmission, and in fact, this hierarchical organization was so ground-breaking and efficient, Onon says, that it "had some features in common with the general staff of a modern army."²³⁴ In any case, the Mongols' enemies had nothing of the sort, as Juvaini points out: "with regard to the organization of their [the Mongols'] army, from the time of Adam to the present day ... it can be read in no history and is recorded in no book that any of the kings that were lords of the nations ever attained an army like that of the Tatars [i.e., the Mongols]." The Mongol decimal system "is, indeed, the best way to organize an army," Juvaini concludes.²³⁵

The Mongols' nomadic lifestyle made such a system particularly efficient. Because the Mongols mixed civil and military life, and every man was thus both a soldier and a herder and hunter, the decimal organization extended from the army to the entire society. Households were

²³⁴ *Secret History* 10,168; Carpini 26, 32.

²³⁵ Juvaini 29.

organized along the same decimal system, and camped, lived, and interacted following the same hierarchy. As a result, each unit was its own communal system able to provide for its own food and responsible for its own share, which precluded problems with provisioning and mass distribution that traditional armies and societies faced. It also enabled the army to be on constant stand-by and ready to go and thus considerably shortened the time necessary for mobilizing. As Juvaini explains, this dual societal and military decimal organization was a great advantage over the Mongol opponents' slower, inefficient arrangements. "Whenever these [other] kings prepare to attack an enemy or are themselves attacked, months and years are required to equip an army and it takes a brimful treasury to meet the expense of salaries and allotments of land," Juvaini writes. In addition, "when they draw their pay and allowances the soldiers' numbers increase by hundreds and thousands but on the day of combat their ranks are everywhere vague and uncertain and none presents himself on the battlefield." The Mongol army, on the other hand, was quick and reliable because civil and military life was tightly intertwined and ran along the same decimal system.²³⁶

The organization within each unit of the army was novel and unique too. Genghis Khan insisted on the diversity and heterogeneity of each unit. In order to avoid having tribal groups or large aristocratic families regroup under units of tens or hundreds that might turn antagonistic to Genghis Khan and foment insurrection under cover of the decimal system, Genghis Khan fundamentally reorganized the households and thus composition of the army. He broke up ethnic groups and families and assigned households to various units and set up each unit with very diverse mix of people. This also enabled him to incorporate foreigners and certain vassals into units in the midst of Mongols. The same way, when defeating the tribes of Mongolia, Genghis Khan systematically dispersed those he defeated into the various Mongol decimal units in order

²³⁶Ibid. 32.

to dilute the danger of revolt by former enemies. Genghis Khan introduced one last safeguard to the system by requiring that once a man and his household were assigned to a particular unit, they could not leave it under any circumstance and the unit became their extended family. Thus, the decimal organization was an inherent watchdog that safeguarded the Mongol leaders against risks of domestic uprising and mutiny.²³⁷

The decimal system also safeguarded them against insurrection in another way. One of the great novelties of the system was its reliance on merit for nominations and promotions. Genghis Khan, having as a youth suffered from ill-treatment at the hands of aristocratic tribal families for his relatively modest origins, was highly suspicious of bloodline privileges and inherited positions and the challenge they could pose to his rule. He knew that local aristocrats and warlords were often the ones who stirred up revolts against central power. While all other armies at the time had only princes and aristocrats, capable or not, in charge of the troops, Genghis Khan set up an unprecedented, purely merit-based command system. As commanders of thousands and ten thousands, "he chose those who were skillful, strong, and handsome," the *Secret History* tells us, and he ordered them to "choose heroes" to command the smaller units. Thus, commanders were the most competent leaders and warriors, and could be just anyone, from simple tribesman to tribal leader or even vassal and foreigner, as long as the position was deserved.²³⁸

The best example is perhaps Subudai, the Mongols' genial general and strategist. The son of a simple herder and blacksmith, Subudai joined Temujin with his brother in the 1190s. He rose within the Mongol hierarchy to command a unit of cavalry, then his success in combat earned him higher positions, and he was eventually given the command of a unit of ten thousand

²³⁷ Saunders 53.

²³⁸ *Secret History* 168-169.

against the XiXia, the Jin, and during the Western campaigns. As Juvaini explains, all men in the Mongol army, "great and small, noble and base," were treated alike and could advance in rank if they distinguished themselves. Promotions depended solely on military and moral merit, the *Secret History* says. This equality of opportunity not only encouraged better behavior from all warriors but also enhanced the loyalty and obedience of the troops. And just as there was no privilege of birth, there was also no privilege of seniority. An older man was no superior to a younger man of the same rank but instead was subject to the same rules. Any favor had to be earned on the battlefield. The *Secret History* enumerates at great length the various qualities of the commanders originally nominated by Genghis Khan to lead each unit. After the nominations Genghis Khan reportedly told them: "I gave favor to you ... [since you] merit such favor."²³⁹ In the *Precious Summary*, a chronicle of Genghis Khan's life, Mongolian Prince Sagan Setsen also emphasizes that Genghis Khan distributed ranks and rewards solely to those who deserved it, and not on the base of heredity. "He [Genghis Khan] granted ... all who had shown him their strength and ability sought-after posts and titles, high ranks and large presents, to each according to his merit," Setsen writes.²⁴⁰

Along the same lines, Genghis Khan made sure to instill the decimal system with a broader sense of justice and equality, reasoning that this would keep his warriors grievance-free and further boost both their morale and loyalty towards him. Tasks were distributed with this goal in mind, as Juvaini explains. "There is a true equality to this [system]," Juvaini writes; "each man toils as much as the next, and no difference is made between them, no attention is paid to wealth or power." Things were very different in other armies, Juvaini argues, with rampant

²³⁹ Ibid. 207-211, 214.

²⁴⁰ Sagan Setsen, *Erdeniin Tobchi (Precious Summary)* (ca. 1662), transl. into German (*Geschichte der Ost-Mongolen und Ihres Fürstenhauses*) by Isaac Jacob Schmidt (St. Petersburg: N. Gretsck, 1829; reprint by Europe Printing, The Hague, 1961), 91, own translation.

privileges and bribery. To underscore the equality of treatment in his army, Genghis Khan decided from very early on to involve the sons of Khans, including of the Great Khan, in direct combat at the front, making sure all would fight and risk their lives, not just the regular soldiers. Ogodei, upon becoming Great Khan, even took this rule one step further and systematized it, the *Secret History* reports. At the suggestion of his older brother Jaghatai who reportedly argued that "if the eldest sons [of Mongol leaders] campaign ... morale will be high and the army will be strong," Ogodei required that all oldest sons of Mongol leading families be sent to battle.²⁴¹

As a result of these egalitarian measures and of the hierarchic structure of the decimal system, obedience and loyalty were particularly high in the Mongol army and society in general. Most foreigners who witnessed the Mongols' military and social organization concur. The Mongols "are more obedient to their masters than any other men in the world," Carpini writes. He notes the Mongols' great respect for their leaders and for each other, and stresses that he observed very few in-fights and no thefts. Lost animals recuperated by someone else were always given back voluntarily to their owner. The Mongols rarely complained, and always shared their food with one another even when there was only little. While they did not show such reverence at all to non-Mongols, Carpini continues, their spirit of loyalty for one another and for the Mongol hierarchy clearly boosted their war-making abilities by enhancing the army's cohesion—in combat the Mongols always fully obeyed higher-ranking commanders, helped their wounded, and sacrificed themselves to save other Mongols. "Whatever command he [the leader] gives them, whatever the time, whatever the place, ... to life or to death, they obey without a word of objection," Carpini concludes. Marco Polo also confirms that "they [the Mongols] are most obedient to their chiefs." For most of Genghis Khan's and Ogodei's rule, very little dissent

²⁴¹ Juvaini 30-31; *Secret History* 263.

occurred. At the death of Ogodei, succession issues began arising, but even when their khans quarreled, warriors generally remained loyal to their immediate hierarchic commander.²⁴²

The Mongol leaders also always made sure to grant no exceptions to the decimal system, in order to preserve its egalitarian goal. They were particularly adamant in upholding the meritocratic nature of the system, which was key to the efficiency of the army. If ordinary warriors were to be held to high standards, so would all commanders. Thus, Genghis Khan and his successors removed incompetent commanders and punished severely those who misbehaved, regardless of their rank, unlike most of their enemies who often simply ignored their black sheep and gave high-ranking officials a blank check. The *Secret History* mentions a few examples of Genghis Khan's intransigence in holding all Mongols to the same code of conduct and obedience regardless of rank. In 1219 during the campaign against Khwarezm, his three best commanders—Jebe, Subudai, and Tokhchar—were sent as vanguards with their forces, with instructions to "skirt the Sultan's [positions] and get on the far side of him ... so that we can attack jointly [from both sides]." While Jebe and Subudai scrupulously followed the order, Tokhchar, instead of waiting in position for the signal to attack jointly, initiated combat and raided and plundered border cities, antagonizing local Khwarezmian leader Malik and prompting him turn against the Mongols to join Jelal ad-Din's forces. Regardless of Tokhchar's defiant action, the Mongols were ultimately victorious against Malik and Jelal ad-Din. Genghis Khan praised Jebe and Subudai for the victory, and, though in the end he did not execute Tokhchar as he first contemplated, the *Secret History* reports that he "fiercely reprimanded him, punished him, and demoted him from his command."²⁴³

²⁴² Carpini 14-15, 27, 33; Marco Polo 91-92; Phillips 72.

²⁴³ *Secret History* 249-250.

Even Genghis Khan's own sons were punished for their missteps. For example, further in the Khwarezmian campaign, his three sons Jochi, Jaghatai and Ogodei took the city of Ortinggechi in the Oxus region, and divided up the inhabitants amongst themselves according to the Mongol usage, but they failed to allocate some inhabitants to their father as they should have. "When they returned and dismounted, Chinggis Qahan [i.e., Genghis Khan] reprimanded them and waited three days before according them an audience," a serious punishment at the court, according to the *Secret History*. Ogodei later also punished high-ranking officials for mistakes that had threatened the Mongols' military success, even members of the Great Khan's family. He had put Batu in charge of the second European campaign and during the campaign, Guyuk disobeyed Batu and even showed disrespect toward Batu and screamed at him in front of a banquet-full of people. Even though Guyuk was his own son and Batu only his nephew, "the Qahan [i.e., Ogodei] became very angry and refused Guyuk an audience." He even considered sending him into exile, but eventually "reprimanded him" instead, leaving punishment as "a field matter for Batu [to decide]," the *Secret History* reports.²⁴⁴ Sometimes the punishment for misbehaving could be very harsh. During the campaign against the Caliphate, Hulegu had ordered a methodical plunder of the city of Baghdad, with fair accounting of all the loot so it could be evenly redistributed among the warriors. But Akanc' writes that after discovering the plunder had resulted in chaos and unnecessary destruction, with most commanders simply taking the loot for themselves, he had some commanders executed and some imprisoned.²⁴⁵

On the other hand, most of the Mongols' enemies did nothing to punish or even stop renegade, corrupt, or disobedient commanders, and this had the opposite effect of decreasing loyalty and support for these leaders. The Chinese annals tell the story of a Jin general who took

²⁴⁴ Ibid. 251, 268-270.

²⁴⁵ Akanc' 69-71.

advantage of the disarray surrounding the defeat of the Jin in Northern China in 1211 to plunder the city of Yu-tcheou, helping himself to its treasure, taking silver and rich clothing, as well as horses that belonged to the inhabitants. The Jin emperor first praised him and bribed him to keep the thefts a secret, then arrested him a year later when he kept complaining that the bribes were insufficient. But in 1213, the Emperor released him and reinstated him at the head of a large detachment in spite of his obvious incompetence. Not surprisingly, the annals conclude, the general "only pursued his pleasures and went hunting" thereafter.²⁴⁶ There are numerous examples of similar behavior in other theatres that went unpunished, and the lack of morale and unwillingness to fight among the Mongols' enemies is thus hardly surprising, as is the very high number of spontaneous defections to the Mongols noted by many witnesses.²⁴⁷ In the end, Genghis Khan's merit-based decimal army organization was a critical innovation that greatly enhanced the Mongol army's loyalty and efficiency and thus undoubtedly gave the Mongols an edge over their enemies' traditional armies, especially when coupled with their advances in weaponry and battle techniques.

d. Weaponry (IV5.1)

Because the Mongols were not a technologically advanced society at the beginning of their rise, they did not so much innovate as adapt to already existing technology when it came to weaponry. Their approach was to adopt the most advanced weapons invented by others, mostly by acquiring them from defeated enemies and by importing foreign experts to help them, and then to improve upon these weapons in their own way to turn them into an advantage over their enemies. As Saunders rightly points out, in terms of weaponry, "Chingis's [Genghis Khan's]

²⁴⁶ Ssu-ma Kuang et. al. 48-49, own translation.

²⁴⁷ Ibn al-Athir 219-220; Al-Nasawi 225-226.

ability consisted, not in startling innovation, but in the uncanny power he showed in adapting and improving existing practices." Their weapon of choice was naturally the bow and arrow, which was the weapon they used for hunting. Little improvement was necessary there, since they had already developed fast, high-impact bows and arrows for hunting that were very efficient in combat. As Marco Polo notes, their skill with the bow met no match. "Their arms are bows, iron maces, and in some instances, spears; but the first is the weapon at which they are the most expert, being accustomed, from children, to employ it in their sports," Marco Polo writes. The bows were usually heavy, with a pull of about 160 pounds, and a range of 200 to 300 yards. The arrows were dipped in brine to pierce through the enemy's armor. Carpini adds that on the warpath the Mongols would generally carry "two or three bows, or at least one good one, three large quivers full of arrows, [and] an axe and ropes" to build more if needed. In addition, to avoid one-on-one contact, "some have lances which have a hook in the iron neck, and with this, if they can, they will drag a man from his saddle," according to Carpini. Because they prized speed above all, the Mongols and their horses generally carried a very light, leather-made armor—a major improvement over the competitors who wore heavy, uncomfortable iron plates that hampered movement and reaction and tired the horses. The Mongols' light cavalry was solely protected by sheepskin coats, while the heavy cavalry wore armor plates made of small leather scales held together with leather strips. All wore high leather boots and an iron helmet with a small iron piece protecting the neck.²⁴⁸

Where most of the adaptation and improvement occurred was in siegecraft, where the Mongols had little experience since they did not themselves dwell in towns and cities. Starting with the conquest of the Jin, the Mongols were confronted with heavily fortified towns that arrows alone could not defeat and that required modern artillery, which they did not have at first.

²⁴⁸ Marco Polo 91; Saunders 63-64; Carpini 33-34; Turnbull 17.

But there are reports that by the time of the siege of Ninqiang, in Manchuria, in 1214, they had copied the latest Chinese catapults and were showering the town with stones and burning torches. They constantly improved their artillery skills, with the help of primarily Chinese and Middle-Eastern engineers who were at the forefront of progress in siege machinery. By the time Batu led the Mongols into Europe a few decades later, they were equipped with a state-of-the-art artillery comprised of various forms of catapults and trebuchets that hurled all sorts of projectiles over fortified walls, including explosive devices.²⁴⁹

When Genghis Khan united the Mongols tribes, he was unfamiliar with gunpowder, which was still in the early stages of development in China. Though it remains debatable when exactly gunpowder was invented and turned into explosive devices to be thrown by catapults, one of the first report of its use in combat was in 1221 by the Jin in the siege of the Song city of Qizhou. Soon the Mongols were faced with the Jin's "thunder-cap bombs" (soft shells stuffed with gunpowder), "thunder-crash bombs" or "heaven-shaking thunders" (iron shells stuffed with gunpowder and sometimes metal scraps, that would fragment and disperse with the force of the explosion), fire-lances (bamboo tubes filled with gunpowder that would spread fire when exploding), landmines equipped with time fuses, and the primitive ancestors of the grenade, the mortar and the cannon. The Mongols adopted gunpowder from the Chinese and developed their own, improved bombs, and were instrumental in spreading the invention of gunpowder throughout the globe through their conquests. When Mongke's army attacked the Caliph of Baghdad in 1258, the Arabs had not yet knowledge of the use of gunpowder coupled with oxygen to trigger explosions, for example, which they picked up from the Mongols.²⁵⁰

²⁴⁹ Turnbull 31; Weatherford 147-148.

²⁵⁰ Weatherford 8, 182-183; Turnbull 18-19.

Starting from a total lack of any artillery pieces or knowledge of siegecraft, the Mongols proved once more their extraordinary versatility by rapidly excelling at siege warfare. One reason they surpassed their targets in artillery effectiveness was that they proved very inventive with the weapons and constantly improvised new means of using them. Al-Nasawi, the secretary of Jelal ad-Din, thus describes their resourcefulness during the campaign against Khwarezm. Having arrived at the fortress of Gurganj, the last bastion of Khwarezm ...

... The Tatars [i.e., the Mongols] started ... to prepare for the siege. They constructed all sorts of siege engines: mangonels, turtles, and wheeled towers. Because the land of Khwarezm lack stones for the mangonels, they discovered, in the area, a great quantity of mulberry trees with very hard trunks, which they began to cut in rounded pieces; then they soaked the wood pieces in water until they acquired the weight and hardness of the stone, and they used them as projectiles instead of the stones, to load their mangonels.

The Mongols reportedly used a similar trick in their campaign against one Jin town where the inhabitants had collected all large stones within a five mile radius around the town to render the Mongols' catapults useless. Rashid al-Din also describes how Hulegu's men came up with an alternative to stones when attacking Baghdad: "Since the Baghdad surroundings did not hold any large stones proper for the machines, they gathered some in the mountain in Hamrin and Djeloula [three to four days away]; in addition, they cut down palm trees and used their trunks instead of stones." The Mongols were thus successful not only because they kept up with new weapons and technology of their enemies, but also because they were never deterred by any unusual and foreign circumstances and instead found ingenious solutions to all obstacles they encountered.²⁵¹

²⁵¹ Al-Nasawi 154, own translation; Rashid al-Din 283-285, own translation; Turnbull 31, 58.

The Mongols' ingenuity extended to defensive measures to protect themselves against technology or surroundings they were not yet familiar with. For example, during Hulegu's siege of Kaifeng in 1232, the Mongols had not yet matched the Jin's skills with gunpowder bombs. But, while working on perfecting their own bombs, they did not let themselves be defeated by the Jin's still superior technology. To counteract the effect of the Chinese bombs and cancel out the Jin's advantage, they came up with a better defense and dug trenches around towns and made cowhide shields to cover their advance toward the walls. Despite the Jin's technological superiority, Kaifeng fell because of the Mongols' persistence and ingenuity. The Assassins also posed an unusual challenge to the Mongols because they barricaded themselves in mountain-top castles from which they controlled all access paths and could easily destroy any would-be attacker. Juvaini reports, though, that while this hurdle had previously prevented anyone from defeating the Assassins, the Mongols simply adapted their weapons to the new geographical setup. Hulegu had special crossbows constructed by Khitan engineers and artisans, specifically tailored to the Assassins' castles. The crossbows were designed to shoot regular as well as larger, burning arrows at the fortresses from adjacent mountain-tops along horizontal trajectories as far as 1.4 mile in distance. To get the heavy artillery pieces up the mountains, Hulegu positioned warriors every 300 yards and each warrior would relay the pieces up the slope to the next warrior. The special crossbow was highly effective, and after heavy bombardment the castles surrendered.²⁵²

Very rarely did the Mongols' enemies show any such inventiveness. The Song were probably the only ones that showed ingenuity that could have matched the Mongols' skills, but in the end their efforts were insufficient to stop the Mongols. They managed to hold up to the Mongols during a five-year long battle for their strategic Northern stronghold of Xiangyang

²⁵² Juvaini; Turnbull 33, 35, 56-57.

(modern-day Xiangyang in Hubei province) in an innovation contest, from 1268 to 1273. For the first three years the city was able to withstand the Mongol siege by getting supplies via paddle boats coming on the Han River that ran between Xiangyang and its twin city, Fancheng, exploiting the Mongols' water weakness. When the Mongols finally managed to seize the paddle boats and interrupt the shipments, the Song built a wood bridge linking the two cities despite the siege. The Mongols reacted quickly too and constructed mechanical saws that they mounted on some confiscated paddle boats. They managed to section off portions of the bridge, and when the Song attempted to rebuild it, they simply burnt it. Still stalled and unable to take over the city, the Mongols had counterweight trebuchets and men capable of operating them brought from their Persian dependencies—counterweight trebuchets were by then widely used in Europe and the Middle East where they had replaced the originally Chinese traction trebuchets. But by 1273, while the Mongols were moving closer to the city, the Song within Xiangyang began to copy the new counterweight trebuchets, adding their own improvement. The Mongols commissioned their best engineers in the faraway capital of Karakorum to further improve on the siege engine, and were soon delivered a larger, stronger counterweight trebuchet that could throw a stone ten times bigger than the previous version of the trebuchet. Unable to counter-innovate or adapt this time to the new, massive weapon, the city eventually capitulated, handing Kubilai a victory that was a turning point in the campaign against the Song.²⁵³

Given the long, difficult battle the Mongols experienced to take Xiangyang, we can only speculate that if more Mongol enemies had contended with the Mongols' constant innovative and adaptative skills in weaponry the way the Song did, the Mongol rise would have been much slower and arduous. Overall, though, the Mongols kept an edge over all their enemies because they adopted and improved the latest weapons, while only few attempted to catch up.

²⁵³ Rashid al-Din; Turnbull 62-64.

e. Strategies and Tactics (IV5.2)

Instead of simply copying their enemies' best strategies as they mostly did with weapons, the Mongols invented most of their strategies. As a result, and also because the Mongols derived them from their unique experience as hunters on the steppes, these strategies were confusing to their enemies. What was unique about the Mongols' military methods was, on the one hand, the thoroughness of their strategic planning, from the grand strategic down to the operational level, with a new, highly effective communication system and use of manpower, and, on the other, the novel tactics they put into practice to achieve their goals, both on the open battlefield and in siege situations. Once more, the Mongols' strategic and tactical choices show their unusual ability to adapt to any unfamiliar, foreign terrain, and again, other states failed to imitate or counter the Mongols' innovations.

1. Strategic Planning

Genghis Khan and the Mongol leaders paid great attention to strategic planning and devised every campaign methodically, down to the smallest detailed. The military approach of their enemies, on the contrary, was mostly reactive and thus less calculated and often chaotic, undoubtedly due to the fact that they were balancers favoring the status quo rather than rising hegemon. Genghis Khan reportedly planned every combat situation to maximize the impact of the Mongol attack and minimize his own casualties. His opponents, on the other hand, who used mostly larger, peasant-staffed feudal armies, often relied on their vast supply of soldiers at the expense of combat quality. When preparing for a campaign Genghis Khan generally constructed

a systematic attack plan, after deliberating with family members and commanders, something rarely practiced by the absolute rulers of the time.

Genghis Khan established the strategic plan to take over the three Chinese kingdoms as soon as he was elected Great Khan, for example, even though most of the plan was not put into action during his lifetime. The plan called on overtaking XiXia first otherwise it could open up a second front against the Mongols once they attacked the Jin. Once XiXia fell the next target was the Jin. Since the Song were at odds with the Jin, Genghis Khan reasoned that there was no risk they would come to the Jin's defense, and the Mongols could exploit the Jin-Song conflict to keep the Song temporarily on their side and to weaken both enemies at the same time before finally turning against the Song, the strongest of the three kingdoms. The strategic planning was particularly complex because the Mongols often conducted several major campaigns at once and great distances apart, and thus the allocation of troops was crucial. For example, at the kuriltai following Genghis Khan's death, Ogodei and the other Mongol leaders made the decision to send their best troops and commanders to the European and Korean theatres, rightly judging that these would be the two most difficult campaigns where they would face most resistance. Thus, elite generals Batu, Subudai and Kadan were commissioned to lead the European invasion.²⁵⁴

2. Tactical Innovation and the Influence of Hunting

After making detailed strategic plans, Genghis Khan and his successors also paid scrupulous attention to the means used to put the plans in practice, and some of their tactical choices proved ground-breaking in comparison to their enemies'. One of the Mongols' advantage stemmed from their combination of various types of warfare to increase the efficiency of their attacks. While other civilizations typically had one fighting specialty, the Mongols used

²⁵⁴ Saunders 54, 66; Turnbull 14, 44; Weatherford 140-141.

simultaneously their versatile cavalry, artillery and corps of engineers, and later navy, for full-blown and diversified attacks from the air, ground, and sea. Not only did the Mongols attack from everywhere, but they also attacked in many different, unpredictable ways, not just by lining their armies in columns and swarming forward onto the opponent as most armies of the time did. Their battle techniques were unusual and unsettling to any opponent because they were modeled after traditional hunting techniques from the steppes, which all Mongol warriors naturally mastered. As Juvaini explains, for the Mongols, "war ... is after the same fashion [as the hunt], and indeed analogous in every detail."

The Mongols' hunting skills included a variety of highly organized and worked out techniques to surround and defeat the prey, which were adapted to human combat. Just like in hunting, the Mongols used a succession of different tightly coordinated tactics on the battlefield, which each warrior knew and their leaders could call on anytime, executed in quick succession one after the other thanks to the strong obedience and methodical training of the warriors. Marco Polo emphasizes that the Mongol warriors and their horses were used to quick tactical changes in the midst of a battle, in sharp contrast to the usual inertia of their enemies' infantry columns. "Their horses are so well broken-in to quick changes of movement, that upon the signal given, they instantly turn in any direction; and by these rapid maneuvers many victories have been obtained," Marco Polo writes. Even though Genghis Khan did not invent these age-old hunting techniques, he combined them in unique ways and implemented them in an organized fashion for the first time on the battlefield, effectively creating a new, modern type of warfare.²⁵⁵

²⁵⁵ Juvaini 29; Marco Polo 93; Weatherford 8, 62.

3. Communication System

In order for these complex strategic plans and fast-changing battle tactics to be relayed properly even in faraway campaigns and thus to be effective, one requirement was a good communication system, perhaps one of the most dramatic innovations of the Mongols. The Mongols invented a comprehensive communication system to transmit military information both from Mongolia to every warzone their warriors were located at, but also within each combat unit on the battlefield. To convey orders and reports between Mongolia and the battle zones, and between the simultaneous campaign theatres that were thousands of miles apart, the Mongols set up the yam trading posts, a system of horse relay stations unprecedented in its scope and land coverage. Because the yams were also put at the disposition of travelers and ambassadors on their way to the Mongol court, Marco Polo experienced them first-hand and gives a detailed account of the system. "Upon every great high road, at the distance of twenty-five or thirty miles, accordingly as the towns happen to be situated, there are stations, with houses of accommodation ... These are called *yamb* of post-houses," he explains. Every yam post was always fully stocked and ready with food, shelter, and fresh horses for travelers, but also primarily for Mongol military envoys and messengers—the primary purpose for the establishment of the system. "At each station four hundred good horses are kept in constant readiness, in order that all messengers going and coming upon the business of the Great Khan ... may have relays, and leaving their jaded horses, be supplied with fresh ones," Marco Polo continues. Even on mountain roads where there were no villages, the Mongols kept post-houses with all the necessities. By the time of Marco Polo, the Mongols had set up over ten thousand post-stations with a reserve of hundreds of thousands of horses. As a result, military information traveled very quickly via mounted messengers who rode without respite between Mongolia and the various fronts. In

addition, to expedite matters every further, the Mongols had foot-messengers wearing bells on their waist who carried notices between yam posts when necessary—for local instructions, for example. The bell warned the next messenger to get ready ahead of the arrival of the missive to take over without delay.²⁵⁶

Because of the decision to expand the conquests westward toward Europe at the 1235 kuriltai, the yam system was also enlarged in anticipation of a larger front, enabling fast distribution of orders for missions in one direction, and timely reports to Mongol leaders in the other direction. The yam system was highly effective in shortening communication times. While for the Mongols' enemies military information generally took months to reach its destination and any army operating far away from its home base was literally on its own, the Mongols kept in constant contact with their hierarchy and their peers in other theatres, easing coordination and giving them a valuable advantage over their enemies. The yam system enhanced the Mongols' reactivity in battle and made their frequent changes of strategies and battle techniques possible, tailored to rapidly evolving circumstances. *Secret History* translator Onon reports that the Mongols even had express messengers for emergencies that could travel at a pace of over two hundred miles a day. For example, Batu's forces learned in early February 1242 that Ogodei had died in late December 1241. The news thus reached Hungary from Karakorum—a distance of over 4,000 miles—in about forty days. To guarantee the efficiency and constant readiness of the system, the upkeep and provisioning of the yams were the responsibility of the army. "Each two tumen [unit of ten] ... share[d] one yam," Juvaini explains, "according to the census." In addition, to ensure that the yams were always kept in perfect condition, every year each yam was inspected and anything missing or damaged had to be fixed or replaced by those responsible for the yam, Juvaini adds. Thus, Marco Polo writes, "in the management of all this the Great Khan

²⁵⁶ Marco Polo 160-162; Weatherford 144-145.

exhibits a superiority over every other emperor [and] king." He concludes that "it [i.e., the yam system] is indeed so wonderful a system, and so effective in its operation, as it is scarcely possible to describe."²⁵⁷

The Mongols established an innovative communication system not only for long-distance communication, but also for short-distance communication, mainly on the battlefield, to be able to implement their frequent tactical changes. Rapid communication on the warpath was also crucial for field intelligence. Since the Mongols did not have maps and thus lacked knowledge of the geography prior to attacks, they used scouts to evaluate the terrain, distances, and enemy force location ahead of the main army, and these scouts needed fast, discreet means of transmitting the information they gathered. During battles, because the Mongols used a variety of attack techniques, warriors and commanders needed to be able to communicate orders, coordinate maneuvers, and synchronize multi-front attacks at all times. Thus, the Mongols came up with a complex system of signals for each maneuver that included smoke, colored flags, and lanterns, Onon writes. Each warrior knew how to interpret the various signals, learning and repeating them over and over again in songs and stories.²⁵⁸

4. Quantitative Improvements

In addition to a new, efficient communication system, the Mongols also used creative ways to make up for their primary weakness on the battlefield—numbers. Even though the Mongols' hunting techniques and sharp coordination skills gave them an unmistakable qualitative advantage in combat, the Mongol army remained quantitatively smaller than most of the armies it was facing. Their first trick, as mentioned earlier, was to enroll war prisoners into the army and

²⁵⁷ *Secret History* 6; Juvaini 33; Marco Polo 162-164.

²⁵⁸ *Secret History* 12.

send them, disguised as Mongols, on the front line to lead the attack. The prisoners swelled the ranks of the Mongol army and gave the impression of a much larger army. In fact, Mongol attacking forces were often comprised of more non-Mongols—prisoners and drafted vassals—than Mongols. This trick was used to subject Samarkand, with prisoners from Bukhara, and to submit Urgench, with prisoners from Samarkand.²⁵⁹ Ibn al-Athir describes the Bukharan prisoners marching onto Samarkand, stressing that "with every ten prisoners there was a [Mongol] banner, so the city's inhabitants thought they were [Mongol] fighting troops." The inhabitants of Samarkand, "certain they were doomed," opened the gates and surrendered. In addition to making the Mongol army look larger, the prisoners also enabled the Mongols to spare their own warriors' lives, since the prisoners generally fell first. Ibn al-Athir gives the example of the siege of Marāgha, Azerbaijan, in March 1221:

It was their [the Mongols'] custom, when they attacking a city, to send the Muslim prisoners they held forward in front of them to carry out the [initial] assault ... The Tatars [i.e., the Mongols] themselves fought behind the Muslims, so the losses were among their Muslim prisoners, while they were safe.²⁶⁰

The Chinese annals report the same practice in the war against the Jin. Torre Maggiore, who wound up himself a prisoner of the Mongols, describes identical proceedings in Hungary:

[The Mongols] sent the Hungarian prisoners to combat first. After those fell, the Russian, Ismaelite, and Cuman [prisoners] attacked. The Tartars [i.e., Mongols], though, stood behind the prisoners and watched with interest the fall of the prisoners; they slain those who tried to flee.²⁶¹

²⁵⁹ Juvaini 92; Saunders 57; Turnbull 77.

²⁶⁰ Ibn al-Athir 209-210, 216.

²⁶¹ Ssu-ma Kuang et. al. 59; Torre Maggiore 179, own translation.

In addition to serving as soldiers to increase the size of the Mongol army, prisoners were also used to increase the Mongol workforce for various other military tasks that the Mongols, lacking manpower, were otherwise slower at completing. The local captives were generally dispatched into the various army units following the decimal system—each Mongol warrior was in charge of up to ten locals—and ordered to gather food and water for the warriors and their mounts, carry supplies, cut trees and gather stones and other raw materials for the artillery, build and dig defensive fortifications, assist engineers in building siege engines, and operate those engines close to city walls to preserve Mongol lives. The Mongol warriors could focus on the complex attacks and war planning rather than waste time on material details. To this aim, whenever the Mongols took prisoners after a town surrendered, Juvaini explains, they immediately divided up the new men in groups according to their skills. Soldiers were always slaughtered because of the danger they posed. Skilled workers such as craftsmen, artisans, and engineers, as well as scholars, were treated the best, as mentioned earlier, and often taken along to Mongolia or major bases. "The young men amongst those remaining were pressed into the levy," Juvaini says, and used in combat and other military tasks.²⁶²

Besides enlisting prisoners, the Mongols regularly used other tricks to swell their ranks, or at least give that appearance. Genghis Khan already faced the problem of small numbers early on when he was battling the various steppes tribes, and out of necessity came up with tactics aimed at scaring the enemy into believing his forces were much more numerous than they actually were. The *Secret History* reports that he often ordered his men to attach branches to their horses' tails to multiply imprints on their path. Similarly, the night before attacking the Naimans in 1204, when both sides were camping in position and getting ready to attack in the morning, Genghis Khan had every man light five fires at a distance from each other, "to scare the Naiman

²⁶² Juvaini 91-92; Weatherford 92-93, 147.

people." The Naiman watchmen spotted the fires and reported to their chiefs that "daily they [i.e., Genghis Khan's forces] appear to grow in numbers" and that "there are more fires than stars," causing Tayang Khan, the leader of the Naimans, to lose courage and doubt his ability to defeat the Mongols.²⁶³

Carpini describes that because they had many more horses than riders, before a battle, the Mongol leaders stood on their horses a distance away from the enemy, with at their side women, children, and sometimes even dummies, set on horses. "They do this to give the impression that a great crowd of fighting men is assembled there," Carpini notes. The opponents "think [they] are combatants; and alarmed by this they are thrown into disorder." Torre Maggiore describes how the Hungarian army was tricked by such dummies by the city of Erlau (present-day Eger). The Mongols set puppets on their extra horses and left them behind a hill with only a few men to direct them. The regular army engaged in the battle with the Hungarians and pushed them toward the hill. When they were close enough, the Mongol warriors directed the horses to move from behind the hill into plain sight, in orderly fashion. The Hungarians, thinking from a distance that this was a reinforcement army and now stuck between the puppets and the real Mongol army, panicked, turned back and fell right into the waiting Mongol army who slaughtered them except for a few.²⁶⁴ Thus, the Mongols were able to overcome their weakness in numbers on the battlefield and considerably enhance the quality of their military operations through creative strategic and tactical planning that gave them an edge over all their opponents.

²⁶³ *Secret History* 170-171.

²⁶⁴ Carpini 36; Torre Maggiore 160-161.

5. Key Tactical Innovations

The Mongols' tactical prowess on the battlefield, derived from their hunting techniques, relied on a few key innovations present in some form or another in all of their successful invasions: speed and surprise, multiple-front attack, and fake withdrawal. In an age of defense dominance, heavily focused on passive fortifications, the Mongols introduced offensive speed and surprise to the battlefield. The typical attack began with the light cavalry units that, at a signal, swarmed onto their opponent out of nowhere and from every possible direction, attacking in a heartbeat and retreating almost immediately, leaving the enemy so confused and disorganized by the suddenness of the attack that it was easy for the heavier cavalry to then finish it off. The Mongols almost always relied on this type of lightning attack, which some scholars argue is reminiscent of the German Blitzkrieg. It is mentioned repeatedly in the *Secret History* and in Wang Kuo-wei's presumed transcription of the lost *Golden Book* of the Mongol royal family, for example in the description of Temujin's attack on the Tayichiud tribe's camp years before he became Great Khan. The swiftness paralyzed most enemies, particularly the widely used heavy infantries, with their tremendous inertia, who were frequently defeated before anyone could realize what was happening.²⁶⁵ Contemporary Mongolian scholar Dalintai reconstituted a list of sixteen military tactics recurrently used by Genghis Khan and argues that by analyzing these tactics one sees that "Chinggis's [i.e., Genghis Khan's] Arts of War were based on five key elements: speed, suddenness, ferocity, variety of tactics, and iron discipline." A typical tactic, Dalintai writes, was the Crow Soldiers and Scattered Stars Tactic, also known as the Ocean Waves or Lake Tactic, where waves of warriors advanced toward the enemy, fired their arrows and withdrew, replaced by a new wave of warriors, to the back of the line to start over:

²⁶⁵ Wang Kuo-wei 390; Saunders 63-64; Weatherford xvii, 94.

When facing the enemy, the army would split into small groups consisting of three to five soldiers to avoid being surrounded. When the enemy regrouped, the Mongols too regrouped. They were to appear suddenly, like something dropping from the sky, and disappear like lightning. [Each] attack would be signaled by a shout or the crack of a whip. One hundred cavalymen could surround one thousand enemy soldiers and one thousand cavalymen could control a front thirty-three miles long in order to attack the enemy at the right place and the right moment.²⁶⁶

In addition to speed and surprise, one persistent novel theme of Mongol attacks was multiple front and encirclement techniques. The Mongols generally created elaborate schemes to attack their enemy from all directions at once, or at least several simultaneous directions, to disconcert it. The multiple front element was present in most Mongol attacks and at any level, whether against an entire country or simply an enemy camp. As Carpini explains, "columns of stronger men they [i.e., the Mongols] dispatch far off to the right and the left so that they are not seen by the enemy and in this way they surround them and close in and so the fighting begins from all sides." As a result, the often stronger enemy was forced to divide up its own forces and thus mitigate its power. The tactic was highly problematic for all Mongol enemies and complicated their defense preparations tremendously because they could not tell where the Mongols would appear next. In addition, it was a major deterrent to cooperation between Mongol enemies, because one prince or city could not take the chance of moving the army out of their home base to help someone else given the high uncertainty of the next target's location. In large-scale attacks, the Mongol armies never traveled and attacked as one body as did most armies. The Mongol army was always separated into several wings when launching a campaign, generally at least a right, center, and left wing, which sometimes traveled and attacked the target

²⁶⁶ Dalintai Zhu, *Menggu Bingxue Yanjiu: Jian Lun Chengjisihan Yong Bing Zhi Mi* (Beijing: Di Ban, 1990), transl. by Urunge Onon (*The Mongol Art of War: the Strategy of Genghis Khan*) and published as an appendix in *The Secret History of the Mongols* (Richmond, UK: Curzon Press, 2001), 281-287.

hundreds of miles apart to open multiple fronts. Spreading out also provided the Mongols with larger hunting grounds and pastures for their mounts. While it handicapped the Mongols' enemies, who had not planned to divide their troops, the Mongols had no coordination problems thanks to their mobile couriers and efficient communication system, which allowed them to be in constant contact despite the distance.²⁶⁷

For example, both the Chinese annals and the *Secret History* report that Genghis Khan sent three armies that were as far as two hundred miles apart and attacked the Jin simultaneously in 1213: the first army, under the command of his sons Jochi, Jaghatai, and Ogodei, was sent westward to the Jin territory north of the Huang-Ho (Yellow) River; the second army, under the command of his brother Khasar and general Jebe was sent eastward toward Manchuria and the sea; the central army, led by Genghis Khan and his fourth son Tolui went straight south through the fortified Juyong Pass toward the Jin capital of Zhongdu. All three armies launched parallel, well-synchronized attacks in 1213-1214.²⁶⁸ In 1219 again, for the invasion of Khwarezm, Genghis Khan opted for a clever three-front attack. He sent two armies to Khwarezm on the most direct routes from Mongolia, one led by Jaghatai and Ogodei toward Otrar and the other led by Jochi toward Khojent, while he and Tolui secretly led another army a much longer way—some 2,000 miles through deserts, mountains, and steppes—around Khwarezm to attack Bukhara from behind the enemy lines, where they would not be expected at all. To avoid detection and keep their arrival secret, Genghis Khan's warriors befriended nomads they encountered along the way to camouflage their progress, and purposefully passed through deserts and mountains to keep away from population centers. When arriving in more settled areas, the secret army slowed their pace and disguised themselves as merchants. They were so discreet, despite being a whole army,

²⁶⁷ Carpini 36; Weatherford 87, 145.

²⁶⁸ Ssu-ma Kang et. al. 58; *Secret History* 234.

that they were able to casually reach the gates of Bukhara before anyone realized their identity. The same occurred again during the European campaigns. Three Mongol armies attacked Eastern Europe simultaneously in the late 1230s: Subudai took the Southern route along the Danube, while Batu followed the central route through the Carpathians and a diversionary force led by Jaghatai's son Baidar, Ogodei's son Kadan, and Batu's brother Orda took the Northern route to Poland up the Vistula River, where they defeated the armies of Henry the Pious at Liegnitz. After regrouping and defeating the Hungarians at Mohi a few days later, the armies again separated and raided the rest of Hungary and Croatia, as well as Austria, Dalmatia, and Bohemia, then eventually met on the Danube in the spring of 1242, having been informed of Ogodei's death, and headed home.²⁶⁹

Similarly, Rashid al-Din writes that Hulegu attacked the Assassins from three different directions. "The right wing [of the Mongol army], under the orders of Buka Timur and Kuka Ilkan, took the road of Mazandaran [Northeast of modern-day Tehran]; the left wing, under the command of Tekudar Ogul and Kitubuka Noyan, followed the road from Khowar and Semnan [Afghanistan]. Hulegu Khan headed the central wing of the army," Rashid al-Din explains. Hulegu chose the same tactic and encircled Baghdad a few years later. Leading the center aisle of the army, he attacked first from the Northeast, arriving through the Iranian mountains. While the Caliph's army was thus occupied in the East, one of Hulegu's best commanders, Baiju, attacked with the right wing of the army from Rum in the West and Batu's men attacked from the North, forcing the panicked Caliph's army to turn around to face the multiple attackers at once.²⁷⁰

Even on a smaller scale, the multiple front and encirclement tactic worked well. Al-Nasawi witnessed the Mongols' final attack on his leader, Jelal ad-Din, whom they defeated at

²⁶⁹ Weatherford 3-4, 144-145; Saunders 57, 85-86; Turnbull 49.

²⁷⁰ Rashid al-Din 191-193, 263-265, 281, 327-329.

his camp near the city of Amid (present-day Diyarbakir, in Southeastern Turkey). The Mongols attacked at daybreak, storming onto the camp from all sides. "The Tatars [i.e., Mongols] had spread in every direction, just like these proverbs that fly everywhere from ear to ear," Al-Nasawi writes. Jelal ad-Din's army, thoroughly surprised and confused, was easily routed.²⁷¹ In reality, in its small-scale version, the Mongols' multiple attack tactic evokes modern guerilla or insurgency warfare. Dalintai stresses that the Mongols frequently "lur[ed their enemy] into ambushes." Genghis Khan and his successors, often outnumbered by the enemy, tried to avoid full-blown, frontal battles, and preferred instead to send small, dispersed squads of warriors to attack the enemy unpredictably in multiple localized skirmishes, immediately withdrawing before the enemy could recover and respond in numbers. In the Moving Bush Tactic, for example, little squads of no more than ten Mongol warriors were sent to attack with speed and from every possible direction, then fled in all directions, inflicting casualties upon the enemy without taking much risk themselves. This unusual kind of tactic, besides being highly effective, was also another novelty brought by the Mongols; the only other civilizations known to engage in guerilla warfare at the time were the Koreans and the Javanese. The Mongols' use of insurgency tactics was probably a direct result of the flexible and mobile organization of the army—instead of traveling, hunting, and camping at night in large, easy to spot groups, the Mongols mostly broke into smaller groups by decimal unit, catching up with their leaders regularly to take orders and make reports.²⁷²

In addition to speed and surprise and multiple front attacks, a third innovative battle tactic recurrently used by the Mongols was the fake withdrawal or luring technique, which was generally very successful even against strong, well-prepared armies. It consisted of feigning

²⁷¹ Al-Nasawi 405-409, own translation.

²⁷² Dalintai 283; Turnbull 18-19; Weatherford 61-62, 88-89.

retreat after the beginning of combat, as if to indicate that the Mongols were weakening or were abandoning the battle. This usually gave the enemy a false feeling of superiority and self-confidence, sufficient to draw the enemy out of its position and into a chase behind the fleeing Mongol forces. As soon as the chase began the Mongols were in control of the game thanks to their fast, sturdy horses, because they could easily tire out their pursuant over long distances. When they judged the pursuant to be sufficiently exhausted, they stopped at an advantageous location of their choice, whirled around and attacked the pursuant or waited in ambush for the enemy to arrive and fall over it. The worn out enemy stood no chance. Marco Polo gives an accurate description of the false withdrawal tactic:

[The Mongols] occasionally pretend to fly [flee], and during their flight shoot ... arrows backwards at their pursuers, killing men and horses, as if they were combating face to face. In this sort of warfare the adversary imagines he has gained a victory, when in fact he has lost the battle; for the Tartars [i.e., the Mongols] observing the mischief they have done him, wheel about, and renewing the fight, overpower his remaining troops, and make them prisoners in spite of their utmost exertion.²⁷³

The Mongol experimented with the tactic of false withdrawal in their very first large-scale campaign, against XiXia, when they were attacked by the Tangut forces in the mountains on their way to the capital of Yinchuan. Fake withdrawals were then widely used against the Jin. The *Secret History* gives the example of a 1211 battle in Northern Jin territory. Genghis Khan had sent his commander Jebe ahead to attack the Jin and rapidly retreat and lure them into following him. The Jin fell into the trap and followed Jebe's men, unaware that his forces constituted only a fraction of the Mongol army and that Genghis Khan and the bulk of the forces were waiting behind. Eventually Jebe stopped, turned around to face the enemy while Genghis

²⁷³ Marco Polo 93.

Khan closed in from the rear as the Jin were trying to escape Jebe's attack. The tactic led to a major Jin defeat. The Chinese annals describe a similar battle between Mongol commander Mukhali and the Jin by a city Southwest of Zhongdu. The Jin were overconfident because of their greater numbers—ca. 30,000 v. 3,000 Mongols according to the annals. Aware of his inferiority in numbers, Mukhali knew he had to use a trick to win. He sent the bulk of his army into a narrow mountain pass, and a small contingent to separately go attack the Jin. After attacking, the contingent feigned not being able to resist the Jin and withdrew in haste, abandoning flags and drums to make the retreat more realistic. The Jin fell for the trick and pursued them, and the Mongols led them straight into the mountain pass, where the rest of the army fell upon them. "Their surprise was extreme," the annals conclude. 7,000 Jin soldiers were slaughtered; the rest fled.²⁷⁴

The Mongols further put the fake withdrawal technique into practice against Khwarezm. Al-Nasawi, who witnessed Jelal ad-Din's struggle against the Mongol army, gives a detailed account of a battle by Isfahan around 1227-1228. The Mongols took the first blow, he writes, and then retreated, letting Jelal ad-Din's army chase them. Unbeknownst to the Muslims, as they fled the Mongols also left a battalion in ambush hidden behind a hill, which fell on Jelal ad-Din's pursuing army and destroyed it, owing to the confusion created by the surprise attack and the reversal of roles. Jelal ad-Din's army suffered a full defeat, though the leader and parts of the army managed to flee.²⁷⁵

There are several reports of the Mongols using the fake withdrawal tactic during their first invasion of Europe. When Jebe and Subudai led two small subsets of Genghis Khan's army toward Europe in pursuit of the Shah of Khwarezm in 1221, they first arrived in Georgia, which

²⁷⁴ *Secret History* 234; Ssu-ma Kang et. al. 99-100, own translation; Turnbull 14-15.

²⁷⁵ Al-Nasawi 224-231.

had a large, well-trained professional army. Jebe's men attacked first, then feigned to retreat in a panic, so the Georgians enthusiastically chased them. Just as the Georgian horses were starting to slow down, exhausted by their heavy burden of armors and weapons, Jebe's men led them straight to where Subudai's detachment was waiting in ambush. While Subudai was attacking the Georgians, Jebe's men quickly changed horses and went back into the battle, leaving the Georgians no chance. Ibn al-Athir and the *Chronicle of Novgorod* both relate how the Kipchaks and their Russian allies also failed to see the trick of fake withdrawal. When the Mongols entered Kipchak territory in 1223, the Russians and Kipchaks prepared to fight them, but then heard rumors that the Mongols were pulling back and eagerly went in pursuit, Ibn al-Athir writes, "thinking that they had withdrawn out of fear of them and from being too weak to fight them." They pursued the Mongols for over a week. Then, however, "the Tatars [i.e., Mongols] turned on the Rūs and the Qipjaq, who, before they realized it, were confronted by them in an unready state, because they had come to feel safe from the Tatars [i.e., Mongols] and sensed that they had the upper hand over them." The Kipchak and Russian princes were totally destroyed at the battle of Kalka. "They had not had time to form into order against them [the Mongols]; and they were all thrown into confusion and there was a terrible and savage slaughter," the *Chronicle of Novgorod* concludes.²⁷⁶

The Mongols resumed the same type of attacks during the second invasion of Europe. 13th century Polish chronicler Jan Dlugosz relates the story of the 1241 Battle of Liegnitz, where the Mongols feigned to suffer from the attack of the European cavalry led by Duke Henry the Pious and fled. The overconfident Europeans went in pursuit. When their horses, carrying heavy-armored warriors, grew tired, the Mongols hid and ambushed them, while at the same time burning reeds that produced a foul-smelling cloud of smoke. Taking advantage of the low

²⁷⁶ Ibn al-Athir 223; *Novgorod* 65-66; Weatherford 138.

visibility, the Mongols sent a galloping warrior in enemy lines screaming "Flee! Flee!" in Polish, and attacked in the midst of the resulting chaos, destroying the allied army.²⁷⁷ Torre Maggiore reports a similar tactic at the Battle of Mohi in Hungary a few days later. After King Bela's troops attacked, the Mongols pretended to withdraw in fright of the Hungarian army, which followed them. Although King Bela warned his men to be ready for anything, Torre Maggiore writes, "the Hungarians ... rested in the assurance of their large numbers." That night Mongols and Hungarians each camped on one side of the Sajo River and the confident Hungarians merely guarded the bridge over the river. During the night the Mongols silently waded across the river, circled around the Hungarian camp, and attack the Hungarians from behind at daybreak, wreaking havoc. Though some Hungarians were able to flee, most perished in the battle.²⁷⁸

Finally, the Mongols also used several variants of the always successful false withdrawal or luring tactic. The most common variant was to apply this battle tactic to siege warfare. The *Secret History* tells of the siege of Dongchang, at the beginning of the Jin campaign. Mongol commander Jebe, unable to take the city, withdrew his forces for six days and rode away from the city "[so that the enemy] no longer paid attention to him." The inhabitants of Dongchang were persuaded that the Mongols had withdrawn for good and were therefore totally off-guard when Jebe attacked again, by night. The Mongols easily took the city. Jebe used the same trick to capture the inapproachable Juyong Guan fortress perched on the mountain pass that protected access to the Jin capital of Zhongdu. During the siege of Liaoyang, similarly, the Mongols met with fierce resistance and were unable to penetrate the fortified city. They pretended to leave in a hurry, leaving material behind to lure the inhabitants out to collect the goods for themselves. The

²⁷⁷ Jan Dlugosz, *Annales Seu Cronici Incliti Regni Poloniae* (ca. 1480) (Annals or Chronicles of the Famous Kingdom of Poland), transl. into English (*The Annals of Jan Dlugosz*) and edited by Maurice Michael (IM Publications, 1997).

²⁷⁸ Torre Maggiore 161-164.

trick worked and while the people of Liaoyang were out gathering the Mongols' abandoned material, with carts encumbering the temporarily opened gates, the Mongols, who had hid in ambush, stormed in. Torre Maggiore describes a similar tactic used in Hungary, which he witnessed during the siege of the city of Großwardein, while he was hiding from the Mongols in the nearby woods. Hungarians soldiers and inhabitants of the city had taken refuge in the city's fortress, and the Mongols were unable to dislodge them. The Mongols left, and unbeknownst to the people in the fortress, camped five walking hours away and waited for a few days. "When they [the Mongols] did not appear for several days, the inhabitants believed that they had left the region and they left the fortress and moved back into their houses," Torre Maggiore saw. The Mongols attacked a few days later and captured the weakened fortress within a day.²⁷⁹

Torre Maggiore mentions two other variants of the fake withdrawal tactic in Hungary that proved highly successful. He saw how the Mongols used a similar tactic to take over a small fortified island near which he had hidden. The island was accessible by land only from one side, through a small path protected by heavy gates. The Mongols went to the other side of the island and feigned an attack through the water, so that everyone on the island hurried to that side to defend the island. The Mongols, of course, had planned the real attack through the path and gate, which had become unguarded due to their diversion, and they took the island without problem. Torre Maggiore observed another clever fake withdrawal variant by the frozen Donau River, that the Mongols sought to cross but were apprehending because of the weight of their horses. They stayed on the banks for a while, and people on the other side, seeing that the Mongols were not crossing, thought they were safe. Eventually the Mongols left altogether, abandoning a few horses behind as bait. People on the other side of the Donau hurried across the frozen river to get the horses. The Mongols, waiting in ambush, saw that the ice was thick enough and hurried

²⁷⁹ *Secret History* 234-235; Torre Maggiore 171; Phillips 55-56; Weatherford 95.

across the ice on their horses to continue the invasion.²⁸⁰ The Mongols thus came up with uniquely innovative battle tactics—primarily combining speed and surprise, multiple-front attacks, and fake withdrawals—that outfoxed their opponents and repeatedly gave them an edge in battle.

6. Innovative Siege Techniques

The Mongols' tactical prowess and imagination extended beyond the battlefield to encompass new and unusual siege techniques. Although the Mongols engaged in numerous open-field battles and at first knew very little about siege warfare, they soon preferred focusing on taking over cities rather than defeating armies. The counter-value strategy made more sense to the Mongols than counter-force, not only because of their smaller numbers, but also because cities were more vulnerable to their psychological warfare techniques. Thus, when initiating a campaign, the Mongols often targeted a country's cities first, hoping to discourage the population and force it into surrendering to diminish support for the army and central government. This paved the way for easier battles against the country's army. In addition, cities were often commercial and financial cores and many kingdoms could not survive and resist the Mongols long without them. So, for example, in the Khwarezmian campaign, Genghis Khan immediately targeted the major cities—Samarkand, Bukhara, Urgench, Balkh and Merv—before even combating the Khwarezmian army in the field. As Rashid al-Din notes, the cities of Khwarezm needed to be taken first because they constituted the country's power strongholds, many with fortifications, and were thus most dangerous. He credits Mongke for ordering his commanders: "Beginning with Kohistan and Khurasan [districts of Khwarezm, located in present-day Pakistan and Iran/Afghanistan, respectively], destroy to the ground all the citadels and fortresses." Within

²⁸⁰ Torre Maggiore 173-175, 180.

the cities to be taken, the Mongols usually proceeded hierarchically according to a net strategy, taking first the second most important cities and gradually closing in from all sides on the capital or strongest city after support from other cities had been eliminated.²⁸¹

As a result of their focus on counter-value, the Mongols not only had to catch up with the latest siege warfare technology they were missing, they also needed to create new siege tactics if they wanted to surpass their targets and succeed. According to many witnesses, one element that set the Mongols apart was their unusual patience. They did not easily get discouraged by failure. Often they were able to take a town simply by persevering in their siege and outlasting the opponents. Various witnesses say that if they were not immediately successful in a siege, the Mongols would simply wait it out, no matter how long it took, until the population starved. There are several reports of cities that resorted to cannibalism before eventually surrendering to the Mongols. Juzjani writes about the besieged fortress of Ashiyar, in present-day Afghanistan, where "every person used to keep his killed and dead for curing and eating." Juzjani then gives a detailed account of the commerce of human flesh within the fortress. After fifteen months and ten days of Mongol siege, when the fortress finally surrendered, only thirty men had survived. Akanc' similarly describes the siege of the heavily fortified holy city of Marut'a in the Caliph's land, which lasted over three years. The city also ran out of food, Akanc' writes, and "the patient Tat'ar [i.e., Mongols] laid siege till they [the inhabitants] began to eat one another from hunger," after which they finally surrendered. A similar fate faced the inhabitants of the Jin capital of Zhongdu, which had formidable fortifications—walls made of packed clay, topped with crenellated bricks, eighteen miles long and forty feet high, with twelve gates, nine hundred towers and three lines of moats. After several failed attempts at taking the city in 1214, the Mongols simply decided to wait it out passively and let the inhabitants starve. Cannibalism was

²⁸¹ Rashid al-Din 143, 151; Phillips 62; Weatherford 187.

reported by the summer of 1215. In June, the Jin commander abandoned the city and fled, and the inhabitants eventually surrendered.²⁸²

But beyond perseverance, it was the Mongols' inventiveness that gave them the edge over their opponents. Besieged cities, if they had enough provisions, could last a very long time because they could easily fend off any oncoming opponent by shooting arrows or sending other projectiles like rocks, bombs, or molten iron from their walls. As a result, fortified cities were difficult to approach without suffering heavy casualties. The Mongols thus came up with unusual defensive tactics to shield their attackers. One such tactic was to erect their own walls and fortification all around the city, to be able to use their artillery and shell the city with projectiles without being hurt themselves. According to Carpini, they did this frequently. The Mongols often built fences around besieged cities, he writes, "so that no one can enter or leave" and in order to bombard but "not ... suffer any injury from the missiles of the enemy" in the city.²⁸³ The *Chronicle of Novgorod* reports that Batu had several such walls built during his 1238 Russian campaign. In Vladimir, "the lawless Ismaelites [i.e., Mongols, all non-Christians were considered the same] ... surrounded the town in force, and fenced it all round with a fence." Approaching the city of Torzhok, Northwest of Moscow, "they fenced it all round with a fence as they had taken other towns," the *Chronicle* mentions.²⁸⁴ According to Ibn al-Athir, the Mongols used an odd variant of the fence tactic to take the Khwarezmian fortress of Mansūrkhūh (in present-day Iran) in 1221, after besieging it for ten months without success. Genghis Khan ordered the Mongols and prisoners to cut down trees and layer timber and earth some distance away from the fortress until it formed a mount of the same height as the fortress. Then the Mongols climbed on it and assembled a trebuchet at the top, which allowed them to shoot projectiles straight into the

²⁸² Juzjani 1076-1077; Akane' 67; Turnbull 31.

²⁸³ Carpini 37.

²⁸⁴ *Novgorod* 82-83.

fortress rather than try to clear the walls from the ground. They had soon conquered the fortress.²⁸⁵

Another tactic the Mongols frequently used when facing a recalcitrant city was to secretly dig an underground tunnel to gain access. As Carpini explains, "should they [the Mongols] not be able to ... [take the city], they undermine the city and armed men enter it from underground; once inside, some of them start fires to burn the fortress while the rest fight the inhabitants." This was another trick used against the Jin capital of Zhongdu after it offered further resistance. "The Mongols made a big subterranean passage from the army to the middle of the city and, suddenly making an opening in the ground—the inhabitants being all unawares—they leapt out into the middle of the city and fought with the men of the place," winning easily. They reportedly also used the tunnel tactic to defeat the XiXia fortress of Shazhou during the second invasion against Tangut in 1224.²⁸⁶

The Mongols' most famous and most widely used innovative siege tactic was to flood the besieged city. Carpini explains the tactic: "If the [the Mongols] are ... unsuccessful [in a siege] and the city or fort has a river, they dam it or alter its course and submerge the fortress if possible."²⁸⁷ Perhaps because of the Mongols' discomfort and inexperience with water, though, their first attempts at using the flooding tactic resulted in catastrophes and the Mongols had to learn to master dams and river diversions before their new tactic became effective. The first time they put the flooding tactic into practice was likely against the XiXia capital of Yinchuan in 1208, a city located on the Huang-Ho (Yellow) River with a system of canals irrigated by the river. This was probably also the first long siege the Mongols were confronted with. Although they were new at siege craft altogether, they did not hesitate to experiment with innovative ideas.

²⁸⁵ Ibn al-Athir 225.

²⁸⁶ Carpini 21, 27; Turnbull 31.

²⁸⁷ Carpini 37.

Genghis Khan, seeing that the Yellow River had risen with the fall precipitations, had a large dyke constructed upstream, so that the river and canals overflowed and flooded the city's surroundings. But just as the city walls were about to burst, the dyke broke—it is unclear whether it broke because of poor construction or as a result of Tangut sabotage—and water flooded the Mongol camp instead of the city. The next attempt, in 1209 against a Jin city also on the Yellow River, on the way to Zhongdu, was no more successful. The Chinese annals report that Genghis Khan "drew out the water of the Huang-Ho to divert it and send it toward the city; but the water broke their levies and rushed furiously onto his camp, and he was forced to suspend the siege."²⁸⁸

The Mongols quickly learnt from their mistakes, though. They successfully flooded Samarkand through its canal ahead of their planned invasion in 1220. A few months later, when the siege of Urgench did not bear any fruit, they built a dam on the Oxus River (now called the Amu Darya) that flowed near the city, diverted the river's course, and flooded the city. Ibn al-Athir writes that the result was dramatic. "They [the Mongols] opened the dam which kept the waters of the Oxus away from the city, so it was completely inundated and the buildings collapsed ... Not one of the populace survived," Ibn al-Athir notes. The Mongols used the same tactic against the city of Ornas, on the Don River, near the Sea of Azov in Northern Crimea. Carpini writes that "the Tartars [i.e., the Mongols], unable to conquer the city by other means, threw a dam across the river, which ran through the town, and submerged it with its inhabitants and property." Juzjani and Rashid al-Din both report that Hulegu used a similar tactic, this time not against a city but against a fortified encampment, to destroy the Caliph's army's camp in the countryside outside of Baghdad in 1258. Taking advantage of the fact that Caliph's army had built their camp on low ground, in the vicinity of the Euphrates River, Hulegu sent his warriors

²⁸⁸ Ssu-ma Kang et. al. 43, own translation; Weatherford 85; Turnbull 14-15.

to divert the river during the night, right onto the Muslims' camp. The flood not only destroyed their camp and drowned thousands of their soldiers, but by submerging vast amounts of lands, it also cut the army off and isolated it from Baghdad so that retreat became impossible. The Caliph's army was forced to fight a messy battle against the Mongols in mud and water, which the Mongols won. After that, the Mongols circled the flood to reach the now undefended city.²⁸⁹

Thus, in addition to their adaptability that greatly helped them catch up with new weaponry and technology, the Mongols demonstrated unusual innovative skills in military strategy and tactics. But just like there were few attempts to outdo the Mongols' progress in weaponry, there were even fewer attempts by their opponents to outwit their strategic and tactical innovations. Nothing was undertaken to diminish the edge that the Mongols developed thanks to their innovative skills and adaptability. It is thus not surprising, given their unopposed military progress, that the Mongols grew to hegemony with extreme ease. A small handful of heroic military leaders tried to adapt to the Mongols' technical and strategic advances, but they were isolated and so few that their efforts did not make a difference. One of the rare leaders who tried to counter the Mongols' innovative tactics were the Koreans Pak So and his lieutenants Kim Chugon and Kim Kyongson, who held firm against Sartaq's army at Kuju, starting in 1231, in one of the longest and most famous sieges of the Mongol epos. The Mongols built their own towers to reach over the walls of Kuju and dug tunnels under the walls, but the Koreans pushed them back, bombarding the assailants with trebuchet-launched molten iron, burning down the siege towers with projectiles made of burning straw bundles, and blocking the tunnel diggers. When the Mongols built ladders to climb over the walls, the Koreans smashed the ladders back onto the Mongols. But eventually Pak So remained the sole resistance of Korea, and his efforts

²⁸⁹ Ibn al-Athir 228; Carpini 29; Juzjani 1241-1242; Rashid al-Din 281; Saunders 110; Phillips 62; Weatherford 124, 182-182.

were sabotaged by the wary central government that negotiated a settlement with the Mongols behind his back. The Mongols were so impressed with the quick-thinking Pak So that they later saved him from being executed by the Korean King for refusing to surrender despite his orders.²⁹⁰

It is surprising that no one reacted to the Mongols' military prowess because news of their innovative military strategies and tactics spread fast. In fact, Carpini devotes a whole chapter of his report to them, entitled "How to Wage War Against the Tartars," and gives a detailed account of the Mongols' new attack techniques. He describes the Mongols' decimal army organization and advises Mongol opponents to adopt the same system, with a hierarchy based on merit, not birth, and that rewards loyalty and severely punishes the mistakes of all. He stresses the importance of fighting the Mongols with a light and well-trained cavalry and no infantry, with weapons similar to the Mongols', such as lances with hooks to prop riders off their saddles. He offers precise guidance on appropriate weaponry. "Whoever wishes to fight against the Tartars [i.e., Mongols] ought to have the following arms," he writes, preceding a thorough list of weapons. He further recommends specific strategies and tactics to match or counter the Mongols' many tricks: travel and fight in large open fields to avoid ambushes, divide the army into small groups for attacks, always seek to surround the enemy, avoid pursuits because they often lead to traps, use reinforcements to counter the fact that the Mongols and their horses do not tire easily, live in constant awareness of all surroundings, if fleeing is necessary burn all hay so Mongol horses do not find fodder, build fortifications out of reach of siege engines and in places where they cannot be flooded etc. Any country following his advice would have been very well-equipped to effectively balance the Mongols. Carpini knew many of the Mongols' innovations and recognized that in order to win against the Mongols, opponents needed to match or counter

²⁹⁰ Turnbull 37-38.

these innovations. His report provides a clear statement of the causes of Mongol military superiority and the remedies. Yet, no one thought of copying or offsetting any of the Mongols' innovative—and successful—military traits, presumably because they stemmed from and overlapped with a culture and lifestyle so different from that of most potential balancers' that those balancers could not fathom to emulate them.²⁹¹ Unsurprisingly, the same was true of the Mongols' non-military innovations.

2. Unique Non-Military Skills and Innovations (IV6)

The Mongols came up with a remarkable array of innovations in non-military arenas as well, which eased their rise to hegemony by diminishing resistance against their rule. Possessing a powerful military does not by itself guarantee a successful empire. Once territory and populations have been conquered, a rising hegemon must be able to maintain control to become a hegemon. The organization of peace to sustain conquest and transform military victory into political rule is traditionally much more difficult than the military battles. From the onset this might have seemed like an overwhelming task for the Mongols, who as tribal nomads had no prior experience of large-scale economic, administrative, and political organization like the sedentary, civilized societies they took over. Even though—or perhaps because—they started with a blank slate, the Mongols' state building achievement was unprecedented. From a myriad of small, segregated, hostile kingdoms and dominions they managed to peacefully transcend and combine the manifold interests, ethnicities, religions, and local practices and build one large, fluid empire that not only worked efficiently, but also kept most subjects satisfied. In order to succeed in such an enterprise, the Mongols instigated a revolution in state organization—strong,

²⁹¹ Carpini 43-49.

working institutions were necessary to hold their patchwork empire together. Geography, in the Mongols' case, did not generate any advantages (IV6.1). The Mongols' home territory was the barren steppes, which offered few resources for an aspiring hegemon and might even have undermined the Mongols' rise if it had not been for their unique nomadic resourcefulness, and its location was far removed from all major centers of civilization. If anything, since the Mongol heartland was so peripheral, geography further enhanced the need for strong institutions and means of communication and control because of the distance and diversity of the territory the Mongols set out to control. Thus, to keep control over their vast territory, the Mongols engaged in a massive state building effort that focused on the economy (IV6.2), as well as the political, social, and administrative organization (IV6.3) of their empire.

a. Economic Advantage (IV6.2)

Well-aware that a large empire can function and endure only if the population is satisfied and does not rebel, and avidly seeking to bridge the divisions between their subjects, Genghis Khan and his successors developed a flourishing trade system within the empire and beyond, under the assumption that enhancing trade would stimulate economic gain and drive up standards of living throughout the growing empire. Of course, this was not disinterested. Besides enhancing their popularity, greater economic exchange was also a means of increasing revenue and would make the Mongol leaders richer, more powerful and able to surpass all competitors. The trading system reached its goals because it was coupled with an array of innovative measures —support for fair trading practices, uniform customs and tax system, infrastructure expansion to bolster commerce, circulation of new, universal monetary vehicles of exchange, and central support to local communities to join the trading system.

The international trading system that Genghis Khan and his successors set in place was unprecedented in scope. The Mongols systematically opened the trade routes between Asia, the Middle East, and Eastern Europe for the first time, most notably the Silk Route, which had persistently been interrupted by unending local conflicts. Now that these vast areas belonged to the same political unit, the previous barriers to trade mostly disappeared. The Mongols signed commerce treaties with their neighbors and dominions to guarantee the free flow of goods throughout the empire and beyond. Encouraging trade was a clever strategy for the consolidation of the empire; because of the wealth generated by the flourishing trade, the lands taken over by the Mongols had a strong interest in remaining vassals of the Mongols and the Mongols' neighbors sought to remain in good terms with the Mongols, thus diminishing the dangers of revolts and balancing efforts. The Mongol leaders further encouraged trade by spurring the Mongols' own demand for foreign goods, which had traditionally been low because of their nomadic, self-reliant origins, but grew exponentially as the Mongols transformed from conquering barbarians into empire-builders. Since the Mongols were not merchants themselves, they employed European, Muslim, Hindu, and Chinese merchants at their service, thus gathering more supporters. This is how the Polo family, originally Venitian merchants, became involved with the Great Khans and traveled to Karakorum, for example. The Mongol leaders built several cities, including their capital of Karakorum, into major international trading hubs. Karakorum was bustling with merchants and markets, despite its remote location in Mongolia out of the way of the major trading routes, thanks to focused incentives. To reward merchants for bringing goods all the way to the faraway capital and induce them to come back, for instance, Ogodei reportedly offered to pay prices higher than market value for goods and routinely granted

merchants a ten percent bonus. He also financially sponsored the merchants' caravans to defray the cost of transportation.²⁹²

1. Fair Trading Practices

The Mongols' trading system was successful because they took new, special measures to remove the region's traditional trading hurdles. One widespread impediment to commerce at the time was rampant corruption and collusion to control prices. Juvaini testifies that Genghis Khan learned about trade practices and market pricing and encouraged fair values and competition throughout the empire, even in distant Mongolia. For example, Juvaini reports an incident in which a foreign merchant purposefully attempted to sell his goods far above market value to draw a wider profit. Genghis Khan punished the man, after having compared his goods with similar items bought from other merchants. To anchor the rules of fair competition and ensure their even application throughout the empire, Genghis Khan not only led by his good example but also codified the rules in the Yasa, the Mongols' universal legal code, in effect providing one of the first examples of antitrust law. Under Kubilai the merchants were further encouraged to organize in guilds, by regional specialization (silk, rice, sugar, etc.), a supplemental means of professional regulation and supervision.²⁹³

2. Uniform Customs and Tax system

A second, key improvement of the Mongols was to tear down the multitude of local tax systems that hampered trade—which often amounted to quasi-hostage taking, requiring the purchase of multiple passage rights that made trade so expensive merchants would sometimes

²⁹² Weatherford 101, 106, 136.

²⁹³ Juvaini 78; Weatherford 136.

sell at a loss—and replace them with one of the first instances of free-trade zone, rooted on a much less restrictive universal customs and tax system. Chinese advisor Yelü Chucai was in charge of this reform, which singlehandedly multiplied the volume of trade by dismantling the segregation between the various cities scattered along the Silk Route.²⁹⁴

3. Infrastructure Expansion

A third contribution of the Mongols was to build up the necessary infrastructure to buttress the flourishing commercial routes. The first advance in infrastructure concerned the trade roads themselves. The Mongols significantly expanded the number of roads to facilitate the passage of troops, and later on these roads became commercial arteries. On the way to war the army built a whole network of roads, along with bridges, tunnels, and mountain passes. Taoist scholar Ch'ang Ch'ung, who traveled from the Shandong province of China all the way to Samarkand and Afghanistan between 1219 and 1224 to meet with Genghis Khan, praises the quality of the roads and bridges, many of which he says were originally built by Ogodei for military purposes.²⁹⁵ Rachid ad-Din also describes how, to prepare Hulegu's campaign in Persian and Kubilai's campaign against the Song, Mongke ordered massive road and bridge constructions: "Before the army's departure, ... the Mongols threw solid bridges on all deep rivers and all streams that had rapids, covering all the areas where troops had to pass."²⁹⁶

Even more importantly, the Mongols revolutionized road security. In the Middle Ages, travel in general, and trade in particular, suffered greatly from roaming gangs of bandits that thrived on the roads because of the frequent absence of strong, centralized institutions of state coercion. Genghis Khan first used the army to rid the roads of bandits and make them safer,

²⁹⁴ Saunders 125.

²⁹⁵ Chih-Ch'ang Li 76.

²⁹⁶ Rashid al-Din 135.

assigning permanent detachments to patrol the roads. Rashid al-Din emphasizes that the Mongols had very little tolerance for such criminals. He relates that Mongke gave special instructions to Hulegu to rid the roads of bandits before sending him off on the campaign against Persia and the Sultanate, a region particularly plagued by bandits. "Exterminate [those people] ... who are constantly infesting the roads," Mongke reportedly ordered. Rashid al-Din later describes how Hulegu carried out the order and repeatedly destroyed the bandits. Upon marching toward Syria, for example, Hulegu "entered the territory of Khelat and the Hakkâr mountains, which was the hiding place of the Kurdish brigands; those among them that the army encountered were pitilessly decapitated."²⁹⁷ Safeguarding the roads for commerce became such an important task that Genghis Khan even appointed special guards, the *qaraqchis*, to ensure public safety on the highway, Juvaini writes.²⁹⁸ The Mongols further improved the road system by establishing a systematic road maintenance system. According to Marco Polo, Kubilai engaged officers "whose duty it is to see that ... the roads [are] constantly kept in good order." As part of the same program, Marco Polo continues, Kubilai had trees planted at regular intervals on both sides of each major road of the empire. "They serve—beside the advantage of their shade in summer—to point out the road—[especially] when the ground is covered with snow," Marco Polo points out. Wherever trees could not be planted, as in deserts or mountains, Kubilai had stones and columns erected "as marks for guidance," contributing to making the roads safer and more amenable to commerce.²⁹⁹

But improvements in the road system were not the only infrastructure innovations that enhanced trade. The Mongol leaders ordered the construction of storage buildings and warehouses, including public granaries, to hold reserves for disaster relief in case of famine or

²⁹⁷ Ibid. 143, 329.

²⁹⁸ Juvaini 78.

²⁹⁹ Marco Polo 166.

bad harvest. The *Secret History* also reports that Ogodei developed a new irrigation system. He had wells dug in arid locations to provide these regions with water. "I provided the people [of] the nation with a sufficiency of water and grass," Ogodei purportedly said. Under Kubilai, the Mongols completed the Imperial Grand Canal in China, a massive public works and hydraulic engineering project linking the capital of Zhongdu (present-day Beijing) and the economically central Yangtze River basin and started some 700 years earlier under the Sui Dynasty. The Grand Canal, which is still in use today, considerably improved and accelerated the transportation of goods, and propelled commerce between the North, wheat-producing regions of China, and its South, rice-producing regions. Kubilai launched other vast waterway projects, constructing canals, dams, and reservoirs, and sending an expedition to trace the origins of the Yellow River. Travelers such as Marco Polo and 14th century Moroccan scholar Ibn Battuta describe massive improvements in navigation, with the development of a commercial navy, the establishment of sea routes, nautical charts, and bustling port cities. As a result the sea was not a hurdle to commerce anymore, and exchanges soared between the Mongols and India, Malaya, Java, Ceylon, and other islands.³⁰⁰

Moreover, the yam trading post system, originally designed to facilitate military communication, further developed and improved commercial exchanges dramatically by providing a vehicle of rapid communication in peacetime too, where it doubled as the first international postal service. Each Great Khan expanded the yam system as the empire grew. As the war routes gradually transformed into commercial routes, the Yam posts began offering provisions, transport animals, and even guides for foreign merchants and envoys like Carpini or Rubruck. Carpini notes the ease and speed of travel through Mongol lands thanks to the convenience of the yams. "We had fresh horses three or four times almost every day," Carpini

³⁰⁰ *Secret History* 277; Saunders 124-125; Weatherford 220.

explains, which enabled his mission to ride all day "from morn till night, and very often during the night." Rubruck likewise writes that his party was "covering almost each day, as far as I can judge, the distance it is from Paris to Orleans [82 miles], and some days more ...; for sometimes we changed horses twice or three times a day." Similarly, merchants and goods were able to circulate in record times.³⁰¹

4. Universal Monetary Instruments

A fourth contribution of the Mongols that favored their economic boom was the circulation of new, standard instruments of monetary exchange throughout the empire and its dominions. Besides creating a uniform system of weights and measures and generalizing the use of Mongol silver coins, the Sukhe, to simplify commercial transactions, the Mongols also printed and circulated paper currency, called the Chao, for the first time on such a large scale. Although the Song Dynasty in China circulated the world's first paper money in the 10th century, its notes were used only locally, and the Mongols were the first to make paper money the predominant currency of an entire country. Genghis Khan authorized paper money in 1227, initially backed by precious metals and silk. An impressed Rubruck describes the Chao as "... a piece of paper made out of cotton, a handbreadth in width and length, and on this they stamp lines" marking a seal.³⁰²

While it took some time to for merchants and the population to familiarize themselves with the new concept and trust the pieces of paper as a much as a metal coins, the use of paper money dramatically increased under Ogodei, rendering trade faster and safer. In order to encourage the use of the Chao, Marco Polo points out, the Great Khans themselves bought all their goods with it. In addition, if a merchant came from a country outside the empire that did not

³⁰¹ Carpini 55; Rubruck 132.

³⁰² Rubruck 171; Phillips 29.

accept the Mongol paper currency, then they were required, before leaving the empire, to "invest the amount in other articles of merchandise suited to their own markets," thus stimulating Mongol exports, Marco Polo writes. Eventually, all army salaries were paid in paper money to further enhance its circulation, and beginning with Mongke, taxes were also collected in monetary form instead of goods, using both the Chao and the metal coins, also making tax collection safer and more efficient.³⁰³

Under Guyuk, the Chao had become so popular that it threatened to get out of hand because too much was printed and as a result, Guyuk incurred large debts. His successor Mongke, realizing the problem and the need to better control paper money and its issuance, took measures to sanitize and strengthen the Mongol financial system. He created a Department of Monetary Affairs in 1253 to centralize and supervise the issuance of the Chao and prevent inflation. To further stabilize prices and make trade less unpredictable, he also created what was perhaps the first ever exchange rate regime by pegging all currencies used in the empire, including the Chao, to the Sukhe. As in the next decades paper money became the most common form of currency in the Mongol empire, particularly in China, Kubilai continued Mongke's effort to safeguard the system by creating additional institutions to regulate its use. Among other things, Kubilai centralized the issuance of the Chao to one mint facility located in Zhongdu (then renamed Khan-Balik), where the Mongol-Yuan capital was moved to in 1274. Kubilai also established strict rules for printing the bills, requiring that each bill be authenticated with the names and seals of several specially appointed officers, and stamped with the Great Khan's seal. Finally, Kubilai issued a law severely reprimanding any misuse of the bills. For example, Marco Polo writes, "the act of counterfeiting is punished as a capital offense." The universality and security of this new currency greatly simplified and promoted commerce, as Marco Polo

³⁰³ Marco Polo 156-158.

concludes: "This paper currency is circulated in every part of the Great Khan's dominions ... [and] with it ... every article may be procured."³⁰⁴

Besides paper money, the Mongols also devised other revolutionary financial tools to stimulate economic growth. One of their most stunning novelties was an early version of a combination passport and credit card, called Paiza. Issued during Kubilai's reign, the Paiza was a small tablet made of gold, silver, or wood—depending on the importance of the holder—that was worn on the belt or around the neck and bought the carrier protection, shelter, and transportation while traveling, as well as an exemption from local taxes.³⁰⁵ Another innovation consisted of financial partnerships between merchants and princes or other financially endowed individuals, where princes invested capital in the merchant's venture, receiving certain privileges such as tax exemption in exchange. The Mongols generally sought to spur entrepreneurship and for this purpose encouraged safe credit and lending practices, and even enacted a law allowing for declarations of bankruptcy to eradicate debts, but with the safety net that beyond two declarations, an individual would face the death penalty.³⁰⁶ The Mongols thus provided unique financial and monetary instruments that worked uniformly throughout the empire to enhance commerce and bolster economic growth, propelling the Mongol beyond all its competitors.

5. Central Support of Local Communities

A fifth Mongol reform that contributed to enhance the economic potential of the empire and solidify the Mongols' rule was the subsidization of local communities that suffered in the invasions. Genghis Khan's advisors outlined the economic potential of conquered lands and

³⁰⁴ Ibid.; Phillips 79; Weatherford 176.

³⁰⁵ Sheila Blair, "East Meets West under the Mongols," *The Silk Road Journal* 3.2 (Dec. 2005), available online at <http://www.silk-road.com/newsletter/vol3num2/>.

³⁰⁶ Saunders 68; Weatherford 205, 221.

emphasized to him the additional value and revenue that could be gained by propping up local communities destroyed during the invasions and enabling them to bear economic fruits. As a result, instead of concentrating the wealth earned in the conquests in the hands of the rulers, as did most invaders at the time, Genghis Khan and his successors redistributed parts of the spoils of conquests to local communities as reconstruction aid, in order to encourage economic recovery and bolster support for the Mongol and dissipate the resentment created by the invasions. According to all accounts, Hulegu was the Mongol general most involved in the reconstruction program. Rashid al-Din mentions that Mongke gave Hulegu clear instructions to that effect, before sending him on the campaign to Persia and westwards. "Make sure to repopulate the countries that war will devastate," Mongke reportedly ordered.³⁰⁷ Hulegu carried out the order. Akanc' maintains that after completing the invasions of parts of Eastern Europe and the Caucasus:

He [Hulegu] began rebuilding the devastated places, and from each inhabited village he selected householders, one from the small, and two or three from the large villages, and ... sent them to all the destroyed places to undertake rebuilding. They paid no taxes at all, but gave only bread and brother to the Tat'ar [i.e., Mongol] travelers.³⁰⁸

Rashid al-Din gives a similar testimony of Hulegu's efforts to rebuild Persia. The city of Khabushan, currently in Eastern Iran, was reduced to shambles and had remained inhabited after the first Mongol invasion. "Hulagu Khan gave order to rebuild the city, and to defray, at the Treasury's expense, all necessary costs, so that the population would not have to bear any expense," Rashid al-Din writes. Hulegu had ground water pipes built throughout the city to provide the population with fresh, drinkable water; he had artisans' workshops rebuilt, as well as

³⁰⁷ Rashid al-Din 143-145, own translation.

³⁰⁸ Akanc' 77

mosques and gardens. This encouraged the inhabitants to come back to Khabushan and restore their own houses and businesses.³⁰⁹ Similarly, Juvaini says, after having repeatedly plundered and done much damage to the Transoxania region around Bukhara and Samarkand, "the Mongols pacified the survivors and proceeded with work of reconstruction, so that at the present time, i.e. in 658/1259-1260, the prosperity and well-being of these districts have in some cases attained their original level and in others have closely approached it." Juvaini notes that although the Mongols made great efforts in reconstruction and were highly successful in the areas described above, other areas like Khorasan and parts of Iraq were so badly damaged that they remained a ghost zone, with most of the survivors simply moving away. Overall the reconstruction policy greatly helped them to transform from feared conquerors to respected sovereigns.³¹⁰

6. Foreign Influence

Just like with military progress, foreigners played a crucial role in the economic improvements undertaken by the Mongols. One figure in particular, Chinese advisor Yelü Chucai, was behind to the reconstruction program, as well the financial reforms, the tax and customs overhaul, and most of the infrastructural enhancements the Mongols undertook. He singlehandedly educated the Mongol leaders, starting with Genghis Khan who hired him in 1218 and until his death in 1243, to the wisdom and value of encouraging growth and commerce to consolidate the empire domestically and internationally and gather the support of the population. When Genghis Khan planned to raze the conquered Jin territory in 1227 and transform it from farmland to grazing land for the Mongol cattle, for example, the Chinese annals credit Chucai for

³⁰⁹ Rashid al-Din 185, own translation.

³¹⁰ Juvaini 96-97.

stressing the income potentially generated by these farmlands and convincing the Great Khan that farming and trading would be more profitable, coupled with a departmental subdivision, a system of taxation and tolls for merchandise, and tribunals to administer the departments and collect the taxes. Genghis Khan accepted, and the reforms were later implemented by Ogodei in 1230. The Mongols thus learned much from foreign advisors, not solely in the military arena but also in the domain of creating wealth. As a result they promoted easy travel and immigration rules for foreign scholars, merchants, artisans, and craftsmen to come to Mongolia and share their knowledge and provide new products. They granted immunity not only to diplomats and envoys, but also to merchants and sellers of all kinds, guaranteeing their safe conduct in the Yasa even for those coming from hostile regions, Juvaini explains. The Mongols thus imported numerous advisors and merchants to work for them, but also Persian and Chinese architects, who built the capital of Karakorum, Silesian miners, European metalworkers, including Parisian goldsmith Guillaume Boucher and fifty of his assistants, who brought new goods to the Mongol empire and further fueled its economic dominance.³¹¹

Overall, the Mongols' innovative, unitary, growth-oriented economic system worked well to cement the growing empire by providing strong rewards for belonging to the Mongol sphere of influence and incentives to transcend local political antagonisms. The Mongols' multiple economic and financial reforms ultimately propelled their power and influence beyond the reach of any potential competitor or balancer, and helped them transition from a period of wars and conquests into a new era of Pax Mongolica—exchange, peace and security.³¹² Parallel to their economic consolidation, the Mongols also built up their political, administrative, and social institutions.

³¹¹ Ssu-ma Kang 124; Juvaini 78; Al-Nasawi 156; Ibn al-Athir 226; Weatherford 153, 170; Saunders 57, 60.

³¹² Saunders 68-69; 124-125.

b. Political, Administrative, and Social Organization (IV6.3 and IV6.4)

The Mongols' second major non-military enterprise to consolidate the empire was to engage in major, revolutionary state building. Starting with Genghis Khan, the Mongol leaders constructed one of the most progressive and efficient political, administrative, and social arrangements of the time that enabled them to successfully govern a vast and fundamentally diverse territory without succumbing to the many conflicts that had previously torn the same areas apart. The Mongols' novel state organization relied on ten innovative pillars: strong, sensible, and semi-democratic central leadership; local autonomy with partial Mongol oversight; a universal legal and constitutional body of rules equally applicable to all; flat-rate, limited taxation; egalitarian, transparent, and expert administrative practices; record-keeping and census; promotion of cosmopolitanism and cultural diversity; religious tolerance; repelling of aristocratic statuses and traditional privileges and support for new, egalitarian loyalties; and finally, gender equality. These pillars hint at a surprisingly modern society for the 13th century and may constitute one of the key reasons why the Mongols were able to rise to hegemony virtually unopposed.

1. Semi-Democratic Central Leadership

The Mongols' first major reform in political organization resided at the level of central leadership. While Mongol government remained essentially a hereditary, autocratic rule, the Mongols introduced a dose of democratic choice and debate in the political equation through the institution of the kuriltai. The kuriltai, or tribal council, was a reunion open to all Mongols to vote on important political matters, such as starting a campaign or electing a new Khan, that took

place at a designated location. An ancient democratic practice and precursor of the referendum, the kuriltai ratified a political decision by the number of attendees. The Mongol people simply cast their votes for the measure by traveling to and attending the kuriltai. Not attending the kuriltai was a vote against the measure. As a result, kuriltais, especially for dramatic decisions like the election of a Great Khan, were a massive affair, with hundreds of thousands of attendees and tents spread over miles and miles, as Carpini, who was present at the kuriltai that elected Guyuk as Great Khan in 1246, describes it. Because they were such large meetings, they were planned long ahead and the issue was publicized, especially since the Mongol leaders encouraged large public discussions of the issue prior to the vote. For example, Genghis Khan summoned a kuriltai in 1211 at his home base by the Kherlen River to decide whether or not to go to war against the Jin. The kuriltai gathered a massive attendance, and the decision was ratified. Thus, a measure passed if a majority of the population attended the kuriltai. A successful kuriltai also required the presence of senior tribe members and major military leaders, which explains why Batu had to abandon the European campaign and return to Mongolia at the death of Ogodei, and Hulegu was forced to withdraw from Syria after Mongke's death. Of course, the decision and count of the attendees remained to the discretion of the ruling Khan or his family, but disregarding the results of a kuriltai was highly frowned upon and weakened the legitimacy of the measure or leader elected.³¹³

Genghis Khan used this semi-democratic practice since the beginning of his leadership, and it is at such a kuriltai that he was proclaimed Great Khan in 1206. Consequently, in the Mongol empire, all Great Khans, chosen among the closest male relatives of the deceased Khan, must be elected by a kuriltai. Regional khans were elected by similar, local kuriltais. Carpini points out that Genghis Khan codified this rule in the Yasa, the equivalent of the Mongols'

³¹³ Weatherford 83.

constitution, making it a very strict rule not to be broken under penalty of death. The Mongols did not hesitate to uphold the rule. Carpini watched as "before Cuyuc Chan's [Guyuk Khan's] election, in accordance with this law one of the princes, a nephew of Chingis Chan [in reality if was Temuge, brother of Genghis Khan and great-uncle of Guyuk] paid the death penalty, for he wanted to rule without an election."³¹⁴ Because he was elected and as such, faced possible competition for the office, a Khan also had to demonstrate his leadership skills. Thus, the most able man in the deceased Great Khan's family, and not just his immediate heir, was chosen to succeed him. Just as within the army, the rule of merit also applied to the highest political and military offices. For example, Guyuk was succeeded by his cousin, not his son, and likewise, Mongke was succeeded by his brother. There was no automatic succession like in feudal Europe, which often brought about incompetent leadership since ability never entered into consideration. The law making the khanate a semi-elective position thus brought higher quality to the Mongols' central leadership.³¹⁵

More able leaders also meant more benevolent, successful leaders. Foreign observers like Carpini, Rubruck, or Marco Polo could not suppress a certain awe toward the Mongol khans and all point out that despite being ruthless campaigners, the Mongol leaders were also virtuous rulers in their empire. Genghis Khan eventually became the object of a veritable cult of personality, and is to this day a symbol of Mongol unity because of his reported righteous and fair leadership. Both the Yuan history of the Mongols, based on the Khan family's *Golden Book*, and the *Precious Summary* emphasize his egalitarian character, stemming from a rough childhood where he suffered from disdain and rejection by clan leaders and upper-class clan members, and from his ascent to power starting from nothing. It is his military heroism and his

³¹⁴ Carpini 25.

³¹⁵ Saunders 73.

sense of justice and modest roots that attracted many Mongol tribes to freely support his leadership. The Yuan history observes, for example, that:

The members of the Taiçi'u[t] tribe [i.e., Tayichiud] could not bear that their leader did not follow any laws. They talked amongst themselves: "the Prince [Temujin] dresses people with his own garments; he hoists them up on his own horse; to raise the people's spirits and bring peace to the kingdom, he is the one assuredly." Therefore they submitted to his leadership.³¹⁶

Many other tribes similarly joined Genghis Khan voluntarily, out of conviction, because he had convinced them of his better leadership.³¹⁷ Carpini and other European witnesses also testify that it is because of their leaders' insistence on justice, equality, and providing the good example that "these men [i.e., the Mongols] ... are more obedient to their masters than any other men in the world ...; they show great respect to them nor do they lightly lie to them."³¹⁸ The insistence on ability, merit, and benevolence in leadership, reinforced by the introduction of debate and semi-democratic practices in the political process, resulted in a strong, yet non-arbitrary central regime that enhanced popular support for the Mongols and dissuaded rebellion in the empire.

2. Local Autonomy with Mongol Oversight

A second pillar that similarly increased adherence to the Mongol rule consisted in a novel principle of administrative decentralization. Ruling over the largest territory every to be under the same political entity was a daunting task for the Mongols, in particular ruling it in a way that would foster allegiance to the Mongols. Firm centralized control was out of the question not only because it was infeasible over such a distance and with such a diversity of local practices and

³¹⁶ Wang Kuo-wei 148-149.

³¹⁷ Sagan Setsen 61-89.

³¹⁸ Carpini 14.

forms of government, but also because it would cultivate resentment and trigger resistance. Full autonomy was unthinkable because the annexed lands would not stay in the empire and follow the Mongol edicts on their own, and would likely just revert to their old habits and local conflicts. The Mongols' solution was thus a semi-autonomous federal administration. Their general rule was to combine two levels of government in the regions, local and Mongol, by retaining and legitimizing the local power arrangements, whatever they were, as long as they were recognized and approved by the Mongols, to be by default responsible for the everyday management of the lands, and superimposing a layer of Mongol control to them by appointing a Mongol, or at least a loyal, governor who would make sure that the local rules did not violate the more general Mongol rules contained in the Yasa and determined by the Great Khan. Thus, the Mongols and the Great Khans were able to maintain the ultimate authority while at the same time avoiding excessive intrusion upon local customs and traditions and preserving the diverse political and ethnic structures of the empire. Genghis Khan immediately settled on this solution after taking the conquest of XiXia. Local kingdoms were left to administer themselves as they meant, as long as they respected Mongol directives and paid their annual tribute tax. Any hint of rebellion, however, was punished by renewed invasion and full Mongol takeover, as happened in XiXia in 1226, a vassal which suddenly refused to furnish troops and tribute to Genghis Khan. All provinces that were under the direct command of the Great Khan or his family, like Mongolia and the Jin territory, were subject to the Mongols' nomadic, decimal organization of households, which soon covered the entire Northern steppes.³¹⁹

During the first decade of the conquests, Genghis Khan already left the Uighurs, Tanguts, Khitans, and Jurches invaded in Northern China to administer their own lands, and withdrew his troops from their territory. But while granting local lands latitude in governing themselves,

³¹⁹ Phillips 76.

Genghis Khan did not tolerate overt rebellion. When the Jurched government refused to submit and the Mongols had to pursue them all the way to Kaifeng, Genghis Khan decided to incorporate the Jin land into his area of direct control, not just as a locally administered region as he did with other conquered land. The *Secret History* mentions another instance where a conquered territory, left on its own, rebelled, and was punished by the Mongols. The Sarta'ul tribe of Siberia refused to pay tribute and killed the Mongol envoys who had come to collect it in 1216 or 1217. Genghis Khan sent a retaliatory force to put them in compliance. Thanks to the preservation of local government structures, however, rebellion remained a relatively rare occurrence. Even in the lands under direct rule, Mongol leaders still left vast rein to local political customs and traditions.³²⁰

And even when Kubilai became the Yuan emperor of China after conquering the Song, he combined Chinese and Mongol political and administrative practices but granted his subjects a large does of autonomy, just keeping provincial outposts for his central offices to oversee the local administration. For example, since China was a farming country composed mostly of sedentary peasants, he recognized that the peasants could not be incorporated into the mobile Mongol warrior household system, so he adapted it and designed a new decimal organization specific to the peasantry. The peasant were grouped in units of fifty households called *she*, which were virtually autonomous, with their own leadership, authority to manage their lives and regulate their farming practices, natural resources, and food reserves in case of famine. Kubilai thus adapted the Mongol system to uphold local circumstances and traditions. In addition, he appointed pacification-commissioners to reconcile the new Mongol rules with the local Han traditions, which led the Mongols to sometimes grant local concessions and exceptions to the

³²⁰ *Secret History* 241; Saunders 77; Weatherford 96-98, 102-103.

genera Mongol rules, in some cases for example extending some social privileges to the higher classes, which were banned under the Yasa but had been widespread under the Song.³²¹

Beyond the territory under direct control of the Great Khan and his family, the semi-autonomous federal organization was very apparent and worked well, for example in Persia. The Chinese annals report that as soon as Genghis Khan returned to Mongolia in 1225, after the first stretch of the Khwarezmian campaign, he immediately named governors for each city he had conquered, "in order to retain the conquered lands," but otherwise let the locals govern themselves. When Ogodei conquered Eastern Iran in the early 1230s, he similarly installed Mongols as heads of each district but appointed local Muslims as emirs of each town. When the region settled down toward the end of the decade and seemed to accept the arrangement, Ogodei even appointed a local Muslim, Mahmud Yalavach, as governor of the whole region of Transoxania. This semi-autonomous political organization continued under Hulegu and his successors, who set up the later Ilkhan khanate in the region. Just like the Great Khan, the Ilkhans remained in charge of the military and of enacting general laws but appointed local vizirs to handle the day-to-day administration, who were left with vast authority, including police powers, tax collection, postal services, and infrastructural tasks. This enabled the Mongol Ilkhans, who were originally not Muslim, to govern a Muslim land with few problems. Ilkhan ruler Abaka, for example, who succeeded Hulegu in 1265, employed locals administrators such as the historian Juvaini and his family. Hulegu appointed Juvaini as governor of Baghdad, a position he retained under Abaka, while his brother Shams al-Din became Abaka's minister of finance.³²²

³²¹ Phillips 105-106, 110; Weatherford 197, 206; Saunders 126.

³²² Ssu-ma Kang 116; Phillips 72, 119; Saunders 74, 130.

In the Caucasus, after invading Crimea, Armenia, Georgia, and Seljuk outposts, the Mongols also left after having appointed governors to oversee the local chieftains. The preexisting administrative and political structures were kept in place. For example, Akanc' writes, "they [the Mongols] left a governor ... in the land of Rum, while they themselves went ... to the eastern country, to the habitat and to their royal tent." This vast regional autonomy ensured that their vassals remained loyal and did not rebel, since the Mongols did not impose an oppressive yoke on their vassals but rather lent them vast degrees of freedom. Akanc' mentions along those lines that "King Hetum [of Armenia] rejoiced greatly" at this power arrangement. Because he remained in power and could largely administer his kingdom himself, his support of the Mongols was guaranteed. The Mongols even refused to take part in local political quarrels, to avoid antagonizing any parties, and instead always sought Solomonesque dispute resolutions. For example, when two claimants for the throne of Georgia, both the legitimate and illegitimate sons of the deceased king, failed to find a mutual agreement over who would rule the country and Guyuk was asked to mediate the issue, the Great Khan simply named them co-kings and ordered them to rule jointly.³²³

The same principle of semi-autonomous local administration was upheld further West in the Russian principalities and Eastern Europe too. The *Secret History* explains that around the time of the first incursions into Russia and Eastern Europe in 1220, "Chinggis Qahan [i.e., Genghis Khan] issued a decree placing resident commanders in all the various cities" he had conquered there. The *Secret History* goes on to list the names of the various local leaders Genghis Khan approved to head their cities. Genghis Khan also always appointed a Mongol governor to supervise the local leaders. "[The local leaders] told Chinggis Qahan [about] the customs and laws of the cit[ies]. Knowing that they had mastered the laws and customs, [the

³²³ Akanc' 45, 47; Saunders 98, 158.

Great Khan] put [them] ... in charge, together with our ... commanders of various cities," the *Secret History* reports.³²⁴

After Batu settled down in the Russian steppes following the second European invasion and formed the Golden Horde, he continued to rule his territory by relying on local officials for everyday administrative tasks. He left all local princes and rulers in office after having confirmed and approved their appointments. Observers thus describe how local princes and governors traveled to Batu's camp at Saray to formally accept the Khan's superiority and be established as the successor when another official would debase or be removed from office, in a ceremony that included Mongol purification rituals like walking between two fires and bowing before idols of Genghis Khan. Batu thus granted Yaroslav Vsevolodovich the title of Grand Prince of Vladimir and Kiev in 1243 after the death of his brother Yuri. But Carpini describes how Prince Mikhail of Chernigov, refusing to bow to the idol of Genghis Khan because of his Christian faith, was not confirmed in his position and instead was beheaded for his inflexibility. The *Chronicle of Novgorod* points out the Golden Horde long continued the same practice, confirming local appointments and ensuring local leaders' loyalty but otherwise giving them a vast range to rule their own lands. In 1322, for example, the *Chronicle* says, "Knyaz [i.e., Prince] Dmitri Mikhailovich went to the Horde, and obtained the title of Veliki Knyaz [i.e., Great Prince]."³²⁵ The semi-autonomous regional decentralization was a brilliant tool of empire building. By applying this principle throughout their conquered territories, the Mongols were able to strike a delicate balance between forging the political unity of the empire and preserving their authority over its faraway regions, while still tolerating sufficient local diversity to retain the support and loyalty of local leaders and populations.

³²⁴ *Secret History* 254.

³²⁵ Carpini 10; *Novgorod* 121-122; Phillips 127; Turnbull 56.

3. Universal Legal Rules

A third, particularly innovative pillar of Mongol state building was an egalitarian legal system. The Mongol legal system was based on the Yasa, a broad-reaching constitutional code enacted by Genghis Khan as early as 1206 that contained the laws and customs that underscored all aspects of Mongol life and was scrupulously followed. Although the Yasa was not the world's first constitution, since the Roman Republic had its own some 1700 years prior, the Yasa was unique for its universal application throughout the Mongol-controlled territories and for the revolutionary measures it contained. The Yasa was in fact fundamentally different from other supreme laws of the time: it was not divinely revealed, nor did it result from ancient rules, but it was derived from the customs of the tribes. Genghis Khan did not keep all the customs of the Mongol tribes, though, but just selected the ones he judged most practical. For example, he abolished social privileges typically granted to white-boned, or aristocratic, tribes members. For Genghis Khan, the Yasa was a life's work, and he repeatedly polished it and improved it. He intended the Yasa to outlive him and become the cement of his new empire, which it did. All of his successors faithfully followed all the rules of the Yasa.³²⁶ For example, Rashid al-Din remarks that Mongke instructed his brother Hulegu, before sending him off to the campaign against Persia, Syria, and the Caliphate, to "ensure that the ways, customs and laws of Tchinghiz-khan [i.e., Genghis Khan] be applied in all points, in mass as in detail." The few changes that Genghis Khan's descendents made to the Yasa just reinforced Genghis Khan's objectives. Besides the Great Khan, no one could alter or update the Yasa, and just like a modern constitution, all subordinate laws had to comply with the Yasa. Thus, local rulers and khans were

³²⁶ It was even incorporated in modified forms by some of the Mongols' neighbors and fiends, for example the Mamluk empire.

allowed to write their own decrees as long as they did not contradict the Yasa. Every high-official and military commander in the empire carried copies of the legal document, written in Uighur script on parchments, to consult when necessary.³²⁷

The Yasa superimposed modern notions and old, quasi-superstitious but practical rules that preserved the Mongol lifestyle and bridged the communities they conquered. Some rules clearly stemmed from Genghis Khan's own childhood experiences. For example, the Yasa outlawed the kidnapping of women—his own wife was kidnapped by a rival tribe, the selling of women into marriage—he 'purchased' his wife by working for her parents as a child, and the abduction and enslavement of any Mongol—he was abducted and enslaved as a child. Similarly, the Yasa severely punished the theft of animals and the non-return of lost animals to their owners—young Temujin and his family almost perished because their horses were stolen, and prescribed that all Mongol children, including the children of concubines, were equally legitimate—another issue that had clouded Temujin's youth. Other rules were more simply inspired by health or safety concerns, or were included to avoid internal struggles and protect the community. For example, the Yasa contained a full criminal code with lists of offenses and punishments. It also included a strict hunting code to preserve wild game supplies. It forbade the use of a knife in an open fire, even to cut or lift meat. It banned adultery outside of the household. It also proscribed urinating in a tent or in running water, and bathing in rivers and streams, obviously to avoid contaminating the drinking water—rules that now appear commonsensical but whose lack of observation in feudal Europe or China at the time led whole towns to be crippled by disease. Other rules of the Yasa buttressed the Mongol empire's military, economic, and political structure, as mentioned before, such as the punishment by death of spying and desertion, the regulation of bankruptcy and lending practices, safe-conduct for all

³²⁷ Rashid al-Din 141; Saunders 69-70.

envoys and merchants, tax exemptions for professions that fulfilled a fundamental public service like undertakers, lawyers, teachers, scholars, and doctors, and of course the rule of the kuriltai to access the khanate, under penalty of death to preclude wars of succession.³²⁸

But the Yasa's originality lies in the revolutionary passages of the document. The Yasa contained one of the first religious freedom laws, protecting the religious diversity of the empire by guaranteeing the right to adhere to and practice any religion without additional taxes or worship fees, and providing a full tax exemption to all members of the clergy, regardless of religious affiliation (see eighth pillar). Another unprecedented addition of the Yasa was its defense of private property and ownership rights, a significant advance in comparison to the feudal societies of the time where most peasants were attached to land they did not own and where themselves the property of their suzerain. Juvaini describes the property clause in detail. The Yasa allowed everyone under Mongol rule to own land if desired, regardless of social rank, occupation or ethnic origin, and also defined clear rules of succession protecting individuals from state or third party encroachment. When a man died, whether he was a high-ranking official or a commoner, Juvaini points out, the Yasa ensured that Mongols authorities "do not interfere with the estate he leaves, be it much or little, nor may anyone else tamper with it." The deceased were free to legate property to their heirs, or if they had no heirs to anyone they wished, including an apprentice or a servant. "On no account is the property of a dead man admitted to the treasury [i.e., confiscated by the state]," according to the Yasa.³²⁹

Kubilai inserted a few reforms into the Yasa that were just as revolutionary. Besides reinforcing the property and landownership rights, he modernized the criminal code to make it less harsh and more human, introducing an amnesty law for repentance, fines to replace most

³²⁸ Phillips 35, Saunders 69-70.

³²⁹ Juvaini 34; Weatherford 69-71.

corporal punishments, limitations to torture, and even a system of parole for the early release of criminals. But his most innovative legal reform was to introduce the framework for a modern welfare state. Marco Polo describes the new measures at great length. First, Kubilai established a system of disaster relief, based on special commissioners that were dispatched every year to evaluate damages resulting from bad crops and natural disasters in the various regions of the empire. "In such cases [the central government] not only refrains from extracting the usual tribute [i.e., tax] of that year, but furnishes them from the [central] granaries with so much corn as is necessary for their subsistence, as well as for sowing their lands," for free in extreme cases and otherwise at a discount price. Then, Kubilai instituted a system of unemployment and disability benefits. He created a special department within the central administration where individuals could submit claims when faced with "misfortunes" or hardships that rendered them poor or when they were otherwise unable to work as a result of "infirmity." "To a family in that situation [the department] gives what is necessary for their year's consumption," Marco Polo explains, with the possibility to reapply for the benefits the following year. Finally, Kubilai set up a system of official charity warehouses that provided clothing and other necessities for the poor that qualified and that was stocked by a rotating public work program, under which every artisan was required to volunteer a certain number of hours every week to make items for the warehouse. These revolutionary measures contained in the Yasa brought unprecedented levels of freedom and equality of opportunity to the populations living under the Mongols, and guaranteed them a less arbitrary government than they had experienced before, while surpassing everything that the Mongols' competitors offered. The Yasa laws thus played a major role in fostering loyalty and reinforcing Mongol rule throughout the empire. As Marco Polo concludes—although

this might reflect his own bias coming from the West—"by reason of this admirable and astonishing liberality which the Great Kahn exercises ... the people all adore him."³³⁰

The ultimate innovation of the Yasa was its universal applicability. Everyone living in the Mongol empire, rich or poor, high-ranking official or commoner, was equally subject to the Yasa and was held responsible for obeying the law. Accountability followed the decimal system: every leader of ten was responsible for the action of his men and households in front of the leader of a hundred, who was himself responsible in front of the leader of a thousand, etc. Genghis Khan, with his abhorrence of privilege and obsession with meritocracy, insisted on submitting himself and all the members of his family to the Yasa and did not hesitate to punish those close to him for disobeying the law. Genghis Khan also organized a system of hierarchical system of courts, with at its summit a supreme judge, in charge of upholding the law and ensuring its application. In the end, the *Precious Summary* concludes, after disturbing and destroying the whole world with their violent conquests, the Mongols, thanks to their elaborate state building efforts, "brought law and order to the people and set up a ... solid political and administrative system, and by governing ... with respect of local ways and customs, promoted the return of peace and welfare for all."³³¹

4. Limited Taxation

A fourth pillar of the Mongols' political organization was a simple, flat-rate taxation. Originally, in the first decades of the Mongol empire, the Mongol leaders relied on a traditional tribal tax system. The *Secret History* explains that the Mongols were expected to "contribute one three-year-old sheep each year from [their] herds to make [the Great Khan's] soup ... [and] one

³³⁰ Marco Polo 164-165, 168-169; Weatherford 202-203, 208.

³³¹ Sagan Setsen 115; Turnbull 17.

sheep from [every] hundred to the poor and needy within their own [military district]." While archaic, this tax system worked decently and was not excessively burdensome on the Northern steppes, but it became impractical to carry through once the empire expanded over far distances, particularly when it started including civilizations that did not raise sheep, and peasants, artisans, and merchants who did not possess any cattle at all. When the empire grew to include the former Jin dynasty and Khwarezm, Genghis Khan at first did not plan to extract regular taxes from his new territories, but simply to raid and plunder them whenever revenue was needed. Thanks to the intervention of Chinese advisor Yelü Chucai, who had dual intent to help the Mongols erect an efficient state structure and to save his own land of origin from destruction, Genghis Khan realized the wealth generating potential of his dominions if he set up a smart, non-suffocating tax system. A Mandarin scholar and governor under the Jin dynasty, Yelü Chucai had considerable experience with taxation issues because the Chinese had had elaborate taxation systems for centuries. The Chinese system, however, imposed a very heavy tax burden and was wrought with widespread corruption, which recurrently led to uprisings from the dissatisfied population.³³²

Seeking on the contrary to foster loyalty and avoid quelling the income potential of the Mongol lands, Genghis Khan and Yelü Chucai devised a much improved taxation system. As Carpini notes, they determined that all subjects of the Mongol empire "shall hand over a tenth part of everything, men as well as possessions." In other words, Akanc' explains, Mongol subjects were required to pay a flat-rate tax or tribute amounting to 10% of their wealth output, and every district was required to furnish troops for the Mongol army and participate in the conquests whenever asked. The tribute was collected, first in goods or coins and later in paper money as soon as it was available, by Mongol tax inspectors who were also in charge of the census. As for the troop requirement, the Mongols just incorporated their vassals' battalions into

³³² *Secret History* 274-276; Weatherford *xix*.

their ranks. Akanc' writes, for example, that the "great ... princes of Georgia and Albania ... gave freely all the tribute demanded ... [and] they themselves, according to their resources and ability, came with their cavalry with them [i.e., the Mongols] on raids, and took the unconquered towns and castles," and even went along in the campaign against the Caliphate.³³³

The flat, relatively low tribute rate must have appeared fair enough and not excessively burdensome to the population, because the few instances of tax rebellion all originated with local kings or warlords, and not from the grassroots like in China. In addition, the Mongols offered exemptions to stimulate the Mongol economy, covering hardships and certain professions that provided public services. Also with the help of Yelü Chucai, the Mongols set up a similarly simple and uniform customs system to replace the multiple customs collections with only one, flat sales tax divided in two categories: according to the Chinese annals, first-need items were taxed at 3%, while the Mongol government collected a 10% tax on luxury items and items not deemed necessary for everyday life, such as wine. Just like with the rest of their state-building efforts, the Mongols attempted to exercise restraint to keep their subjects loyal and satisfied and forestall revolts. Mongke's tax reforms are a clear indicator of these goals. One of the new measures he introduced was a waiver for the collection of arrears. In addition, a specific amount of the local revenues was reserved to be spent toward local development, including local army camps, yams, etc.³³⁴

5. Improved Administrative Practices

To bolster the effectiveness of the tax system and the population's allegiance to the imperial state-building efforts, a fifth, unique reform of the Mongol empire was the introduction

³³³ Carpini 38, 40; Akanc' 35, 65.

³³⁴ Ssu-ma Kang 136; Phillips 85.

of egalitarian, transparent, and expert administrative practices. Here again the influence of Mandarin advisor Yelü Chucai is obvious—he sought to infuse the Mongol government with the age-old Chinese administrative experience, while at the same transcending some of the problematic aspects of Chinese bureaucracy. The result was a administration much more modern and efficient than that of the Mongols' opponents.

Yelü Chucai's first move toward that goal was to establish a system of public service exams to enter administrative employment at any level. Not only did it ensure that administrative positions were filled with the best qualified personnel, but it also made them merit-based positions, which appealed to Genghis Khan. In fact, and this was a major improvement upon previous Chinese practices, the Mongol public service exams were open to former prisoners and slaves, which enabled the Mongols to free over four thousand educated foreigners to fill positions of officials and judges. To further enhance the quality of the Mongol administration, Chucai set up special schools to train future officials.³³⁵

Kubilai pushed Chucai's reforms a step further by adjusting the training of administrators and officials. While in China bureaucrats were traditionally trained and tested on subjects like art and poetry, Kubilai required that the teaching and exams focus on more practical topics. To raise levels of expertise, Kubilai also set minimum standards of knowledge for all public service-related professions, including doctors, lawyers, merchants, engineers and professors. Moreover, in an effort to diversify the supply of administrators and provide more equal access to the profession, he sent for scholars and experts from all over the empire and set new quotas for public service by origin—Northern China, Southern China, and non-Chinese—so that one groups would not dominate. In another key improvement from previous Chinese bureaucratic practices, Kubilai required that all ranks within the administration be based on achievement

³³⁵ Phillips 78-79.

alone, just like Genghis Khan had done with the army, thus encouraging social promotion for the lower classes and enlisting their loyalty. Such egalitarian practices largely helped spare the Mongols the numerous peasant uprising that had previously plagued China, and that would plague the country in the future.³³⁶

Chucaï's second move was to improve upon the other major pitfalls of the Chinese bureaucracy: arbitrary decisions and corruption. His novel measures included limiting the independent power of regional administrators and installing stricter oversight over the local administrations, and turning bribery, embezzlement and reckless spending of public funds into capital offenses. Genghis Khan and his successors did not hesitate to implement the harsh punishment regardless of the rank of the individual involved. For example, Abd al-Rahman, one of the chief tax collectors in the Chinese territories, was tried, convicted, and executed for fraud and extortion under Guyuk. To further enhance the transparency of administrative practices, Kubilai introduced an innovative system of kuriltais for administrative decision-making. Bureaucrats met daily in local councils to debate the various issues at hand, and decisions were valid only if the council had reached a consensus. Similar council were established to resolve administrative disputes, virtually stripping individual officials from their discretionary power. Finally, Kubilai standardized all public salaries, a key measure meant to put an end to the widespread practice of bribery to obtain public services in previous Chinese dynasties, where Mandarins traditionally did not receive an official salary. Genghis Khan had already systematized military pay in a similar way. As Carpini not, the distribution of tribute, gifts, and the spoils of war occurred in an orderly and equitable fashion, since after being carefully counted and recorded, "... these things were shared out among the Emperor and the chiefs ... [then] each

³³⁶ Weatherford 202-203.

chief divided his share among his men," trickling down following the decimal subdivision.³³⁷ By promoting administrative practices focused on expertise, equal treatment, and transparency, the Mongols laid sound and durable foundations for their empire that surpassed the achievements of their neighbors.

6. Census and Record-Keeping

A sixth pillar that tied together most of the previous reforms is the development of the census and record-keeping. Just like the previous reforms created centralized rules and authority, the addition of records and a census centralized information. Although census-taking was not a new practice in the region, its generalization and scope were unprecedented achievements of the Mongols. The census and records provided the Mongol leaders with an idea of the demographic and economic assets of their expanding dominions and their evolution, enabling them to consolidate their reforms and assert control over the vast empire. They were essential tools of government, Juvaini points out, because they were necessary for the successful design and operation of the tax system, the yams, and even the decimal army and household organization.³³⁸ In addition, as historian David Christian underlines, the census and records were "a key to successful mobilization for warfare," accounting for the human and material resources at hand for the conquest efforts. But before being able to conduct widespread censuses and write thorough records, the Mongols needed something even more vital: a writing system. As the leader of illiterate nomads with only a spoken language, Genghis Khan immediately understood the vital necessity of writing not only in forging an empire, but also in ensuring the legacy and endurance of Mongol history, culture, rules, and traditions. He therefore charged Tatar Tanga,

³³⁷ Carpini 64, Saunders 98, Weatherford 99, 204.

³³⁸ Juvaini 33-34.

the Naimans' Uighur script whom he captured upon defeating the Naiman tribe in 1208, to create a Mongol script based on the Uighur alphabet, the most closely related to the Mongols' Altaic language. The new script was thereafter used for all official business of the Mongols. Both censuses and record-keeping—recounting events and laws and edicts—became standard practice in Northern China from the 1230s onward, and in 1252, Mongke ordered the first empire-wide census, covering every Mongol land from Korea to Russia, recording the number of households, animals, and other assets throughout the empire.³³⁹

7. Promotion of Cosmopolitanism and Cultural Diversity

In addition to these major innovations in political and administrative organization, a few other pillars of Mongol state-building took the form of social reforms. One of these pillars, as previous sections suggest, was the promotion of cosmopolitanism and cultural diversity, a feature unique to the Mongol empire that boosted the Mongol rule's popularity within the empire and abroad. Perhaps because their nomadic lifestyle had precluded the development of many forms of art and sciences on the steppes, Genghis Khan and his successors became patrons of the arts and sciences throughout the empire, supporting the cultural rebirth of the regions and civilizations where the conquests had leveled artistic, humanistic, or architectural achievements, like in China or Persia. Despite their little knowledge of arts and sciences, the Mongol Great Khans all demonstrated a curiosity and openness for the cultural and scientific accomplishments of other civilizations and brought numerous foreign artists, artisans, and scholars to their camps and capital, transforming them into vast cosmopolitan centers. They imported skilled artisans from every land they conquered, as Carpini mentions, "tak[ing] all the best craftsmen and

³³⁹ David Christian, *A History of Russia, Central Asia, and Mongolia-Volume 1: Inner Eurasia from Prehistory to the Mongol Empire* (Oxford: Wiley Blackwell, 1998), 414; Saunders 67-68.

employ[ing] them in their own service." But they also attracted a number of scientists. For example, while in the Middle East in 1257-1258, Hulegu enlisted the services of famous scientists, including astronomer Nasir ad-Din Tusi and his sons, who formerly worked for the Assassins, and doctors Reis Eldaulah and Mouwaffek, to join Mongke's camp at Karakorum, Mongke planned to build an observatory there for Tusi, and Hulegu later built one for him in Maragha, the future Ilkhan's summer capital in Persia. Rashid al-Din writes that after assessing the scientists' morality and sincerity, "he [Hulegu] showed them great respect and favors. He had horses delivered to them so that they could bring their wives, their associates and the parents ... along with all the persons they were attached to." To further learn about the practices and advances of other cultures, the Mongols always kept ambassadors from all regions of the world at their courts and asked them to describe and explain the customs of their countries of origin.³⁴⁰

In addition, Genghis Khan, Ogodei, Mongke, and Kubilai, studying Yelü Chucai's teachings, understood the importance of civilization and sought an entourage of wise, scholarly men to educate them on the intellectual and cultural achievements of other societies. They hired Persian, Uighur, Chinese, Tangut, and Tibetan scribes, among others, to write dictionaries of their languages and Carpini and other foreign travelers reveal that the Mongol courts were never short on translators whenever they had an audience. Throughout the conquests, the Mongol commanders were instructed to look for such scholars and bring them back to the Great Khan. Rashid al-Din writes that for example, during the siege of Baghdad, Hulegu sent messages into the city, assuring that "the kadis, wise men, sheiks, and descendants of Ali ... are assured that we will spare their lives."³⁴¹ Juzjani tells a similar story, related to him by Malik i-Maraghani, an imam he met after the conquest of Khwarezm and who told him about his experience in the siege

³⁴⁰ Carpini 42-43; Rashid al-Din 217; Saunders 110, 126, 129.

³⁴¹ Rashid al-Din 283; Phillips 85.

of Herat. During the siege, Malik i-Maraghani recounted, he stood in armor atop the ramparts but lost his balance and fell forward into the Mongol army. Miraculously surviving the fall, he was expecting to be executed on the spot, but Tolui, who was in charge of the siege, upon learning that the prisoner was a learned man, spared his life and took him to Genghis Khan. "This person is a sagacious man, and a wise, and may be qualified for the service of Chingiz Khān," Tolui reportedly said. Genghis Khan employed him and showered him with "great favors," the imam explained, and even "used continually to inquired of me the traditions of the prophets, and concerning the sovereigns of the 'Ajam, and the kings of the past," before he fled the Mongol camp. The journey of Taoist master Ch'ang Ch'un at the request Genghis Khan further substantiates this learning and educative cultural interest of the Mongol leaders. During his stay at Genghis Khan's compound, Ch'ang Ch'un had regular conversations with the Great Khan, who listened and demonstrated avid interest in his teachings on religion, morality, and government, and even accepted the sage's criticism of the Mongols' ways and customs. Although Genghis Khan must have disagreed with many of Ch'ang Ch'un's teachings, his behavior suggests that he highly regarded the Taoist master and welcomed the opportunity to expand his knowledge.³⁴²

As a result of the Mongols' cultural openness, their camps became sought-after multi-ethnic cultural centers where individuals from all regions of the world mingled, protected by a strictly enforced absence of racial or national preference. Carpini describes a vast number of Eastern Europeans and Russians at Guyuk's encampment during his visit, some of whom had been there for decades, probably since the Mongols' first incursions into the West in the early 1220s. At Mongke's camp on the outskirts of Karakorum a few years later, Rubruck describes meeting Paquette, a woman from Metz, in Lorraine, who was captured by the Mongols in Hungary and now worked at the court for a Mongol lady of Christian faith. Paquette explained

³⁴² Juzjani 1038-1042; Chih-Ch'ang Li 47-51.

that she had married a Russian carpenter, and was treated very well by the Mongols because "he was skilled in making houses, which is a profitable craft among them [the Mongols]." Rubruck goes on to describe the Mongol capital of Karakorum as a diverse community including many foreigners, equipped with a Muslim quarter where the merchants usually gathered and operated the city's four markets—for grain, sheep and goats, oxen and carts, and horses, respectively—and a Chinese quarter where craftsmen and scribes usually lived. Modern excavations reveal that the city contained numerous elements of foreign art and architecture. The artifacts retrieved from Karakorum include Chinese-style ceramics and pottery and items with hieroglyphic inscriptions, a testimony to the Mongols' cultural and intellectual curiosity.³⁴³

Kublai continued the cultural revival initiated by the first Mongol khans. He was an avid appreciator of theatre and promoted its development. He also reportedly paid scholars to rewrite old manuscripts and create new historical records, encouraging them to write in dialects in addition to classical Chinese in order to increase the accessibility of knowledge, and he surrounded himself with scribes that specialized in all possible foreign languages. One of his most important achievements, though, was to promote equal access to education and build schools throughout the empire. He built the equivalent of a university in the new Mongol capital of Zhongdu (Khan-Balik, modern Beijing) in 1271, and, according to records, some 20,166 public elementary schools in China that provided universal basic education, including to the children of peasants, who had traditionally been kept illiterate and uneducated by previous Chinese dynasties. Kublai's educative system was unique for several reasons. First, it was inspired by practicality: the peasant children were sent to school locally, in each *she*, during the winter months so as not to overlap with farm work, and they were taught in their local dialect rather than in classical Chinese, which would have been virtually useless to their everyday life.

³⁴³ Rubruck 157, 184; Carpini 66; Phillips 96-97, 99.

Second, although it is unclear whether the system ultimately worked as planned, Kubilai's system constituted the first-ever outline of universal public education—the West did not develop a similar system for another five hundred year. Another radical endeavor of Kubilai in the cultural domain was to create a universal alphabet with which one could write all the languages of the world and that could serve as a new, revolutionary tool of communication to understand the many languages spoken in the empire and beyond. Kubilai charged Tibetan monk Phags-Pa with the task, and in 1269 he came up with a system of forty-one letters, which Kubilai made the empire's for all official business. But because he did not impose the new alphabet on the population, hoping it would be embraced freely as a simpler way to communicate, the new alphabet did not survive beyond his reign. In the end, while many rulers of the time remained cold and detached from the population, the Mongols' patronage of popular culture, arts, and sciences, and their openness to the diverse ethnicities of the empire and beyond earned them general popular support. Their cosmopolitanism and cultural interests was a major pillar of the Mongol state-building effort by easing their transition from a conquering warriors to imperial leaders. In stark contrast to their initial brutality, this increasing "restraint on the part of the Mongols produced positive returns," the Chinese annals conclude, that stabilized the empire.³⁴⁴

8. Religious Tolerance

The eighth pillar of Mongol state-building, religious tolerance, follows the same vein and similarly fostered popular loyalty and reinforced the cohesion and strength of the empire. Although the Mongols were shamanists, revering the gods of nature, in particular Tengri, the god of the sky and heaven, and their elders, they had no religious hierarchy or organized worship. Their only clergy were diviners, fortune-tellers, and magicians. As a result, the Great Khans were

³⁴⁴ Ssu-ma Kang 93, Weatherford 205-206.

naturally skeptical of rigid, organized cults and refused to be affiliated with any of the three major religions of the time that they encountered: Islam, Christianity, and Buddhism/Taoism. They also rejected the idea of imposing a given cult, and although they required visitors to partake to some Mongol customs such as the purification walk between two fires, most of these customs were not religious in nature. Genghis Khan's distance with organized religion began early on the steppes, with an incident that left him distrustful of the power of the clergy and intent on limiting it and preventing conflicts between political and religious authorities. The solution was to allow for the free competition of religious beliefs and organizations, which would limit the impact of the clergy. The incident, described in detail in the *Secret History*, occurred in 1206. The shaman Kokchu, who had help legitimate Temujin's transformation into the Great Khan spiritually and whom Genghis Khan considered a close ally, had become power thirsty himself and developed aspirations for political leadership, threatening to become a serious counterweight to Genghis Khan's authority. To assert his still nascent rule, Genghis Khan had Kokchu killed. Thereafter, Genghis Khan and his successors not only allowed, but strictly enforced, freedom of conscience and freedom of worship, granting full religious freedom to the numerous religions represented in the empire, large and small.³⁴⁵

This vast religious tolerance and the absence of an officially recognized or sponsored religion was a smart political move that significantly contributed to the solidification of the empire by accommodating the numerous, historically antagonistic religious traditions enclosed in the empire and avoid unleashing religious rivalries between them. As Juvaini notes:

Being the adherent of no religions and the follower of no creed, he [Genghis Khan] eschewed bigotry, and the preference of one faith to another, and the placing of some above the others; rather he honored and respected the learned and pious of every sect ...

³⁴⁵ *Secret History* 228-231; Saunders 53, 68.

And as he viewed the Moslems with the eye of respect, so also did he hold the Christians and idolaters in high esteem.³⁴⁶

Thus, the Great Khans from Genghis Khan to Kubilai upheld local religious traditions—each region of the empire could keep its own religious identity, as long as minorities were tolerated and accommodated. All clergy, Christian, Muslim, Jewish, Buddhist, Taoist, and shamanist, was granted tax exemption, ratifying their political equality, and missionaries of all faiths were allowed to proselytize freely as long as they did not employ force. Akanc' mentions, for example, how "King Hetum [of Armenia] ... rejoiced ... because of the document regarding the freeing from taxes of ... our monasteries."³⁴⁷ Genghis Khan employed high officials of all religions, and he even let his children and grandchildren adopt the religion of their choice; in fact, some perpetuated the shamanistic traditions of the Mongol nomads, while others became Christians, Buddhists, and Muslims. The most important social cement of the new empire, to Genghis Khan, was the prevalence of the Yasa, not of a particular religion. As a result of their broad-reaching religious tolerance, the Mongol khans enhanced the loyalty of their subjects and the admiration of foreign travelers. The Prince of Ordos concludes in the *Precious Summary*, that "by letting each faith worship according to its own prescriptions, the Great Khan [Kubilai] fomented peace and order throughout the empire, and he let the inhabitants enjoy their quietude and felicitousness, so that he became well-known everywhere as a worthy leader."³⁴⁸ Concurrently, the great monarchs of Western Europe displayed incomparable levels of religious intolerance. In most places, only one religion was allowed, and the persecution of religious minorities was rampant, which shocked Mongol envoys like Rabban Sauma.

³⁴⁶ Juvaini 25-26.

³⁴⁷ Akanc' 47.

³⁴⁸ Sagan Setsen 119, own translation. Note: Setsen honorifically uses the term "Tschakrawartin," i.e., Chakravarti, to describe Kubilai, which is the name given to the quasi-mythical ancient kings of India.

The Mongols, on the other hand, not only promoted religious tolerance but built a reputation as the guarantor of religious freedom and savior of religious minorities, and from early on religious minorities, aware of the Mongols' religious independence, even called on them for protection. When the Uighur (Muslim) minorities of Kara Khitai were persecuted by Kuchlug, who imposed prohibitions on their religious exercise, they turned to Genghis Khan for protection, prompting him to send Jebe to destroy Kuchlug. Although Genghis Khan had planned to oust Kuchlug anyway, he went further by imposing freedom of religion on the Kara Khitai after his victory. During the campaign he also took great care to spare the Uighur civilians and distinguish them from Kuchlug's supporters, who were plundered and killed. When during Mongol general Chormagan's campaign in Persia in 1235-1236 a great number of Christians were massacred, the Christian princes of the Caucasus did not hesitate to send envoys to Ogodei at Karakorum to report the issue. As a consequence, Ogodei ordered all Mongol generals to pay special attention to protecting the Christian minorities in the region against persecution. Hulegu carried out the order, for instance, by reportedly warning the Christian minorities before attacking Middle East cities, so that they could hide upon the attack and not become the victims of Muslim reprisals. Although such measures were not devoid of strategic motives, they served to enhance respect for the Mongols in many places as the protectors of minorities.³⁴⁹

The Mongols' openness to various religions was reflected in the uniquely diverse arrangement of the Khan's camps. From Genghis Khan to Kubilai, all Great Khans, as well as regional khans like Batu or Hulegu had special ministries at their camps for each religion. Visitors passing through the various camps describe seeing there Roman Catholic, Nestorian, Buddhist, and Taoist priests and monks, Confucian and Manichean representatives and Imams, and even Jewish and Indian clerics. All khans allowed for the building of churches and places of

³⁴⁹ Turnbull 77, 79; Phillips 58; Weatherford 103-105.

worship for the various faiths within their camps. For example, Rubruck happily discovered a Christian church set up in Mongke's camp outside of Karakorum, which he mentions was not just a shabby tent but had a richly decorated altar, embroideries of holy figures with pearls, and a cross set with precious stones. The capital of Karakorum was the quintessential embodiment of the Mongols' openness according to visitors' accounts. It was "probably the most religiously open and tolerant city in the world at that time" and the only place where "followers of so many different religions [could] worship side by side in peace," Weatherford remarks. When Rubruck visited Karakorum, the city held twelve pagan temples, serving each a different pagan tradition, two mosques, and one Christian church, which he describes as "quite large and beautiful, and the roof above it is all covered with silk interwoven with gold."³⁵⁰

The Great Khans' religious tolerance extended to their families. Interreligious unions were tolerated and frequent, and encouraged by the Great Khans' own practices. All four sons of Genghis Khan, as well as regional khans, took Christian wives, mostly women from the Merkit and Kerait tribes that were predominantly Nestorian Christian. Toregene Khatun, Ogodei's second and favorite wife and the mother of Guyuk, was originally from the Merkit tribe and was a Nestorian Christian. Guyuk's principal wife Oghul Qaimish was also a Merkit and a Christian. Tolui's first wife Sorkhokhtani Beki and the mother of Mongke, Hulegu, Kubilai, and Arik Boge, was the niece of Kerait Ong Khan Toghrul, whom Genghis Khan had defeated, and a Nestorian Christian. All three women had key political positions in the empire. Hulegu's favorite wife Dokuz Khatun, the mother of Abaka, was also a Kerait princess and an ardent Nestorian. Abaka himself wed a Christian wife, the natural daughter of Byzantine Emperor Palaeologus, Rabban Sauma mentions, and she became an important Christian leader among the Mongols. The Christian wives played a major role in encouraging the protection of Christians in the empire.

³⁵⁰ Rubruck 150-151, 178, 184; Weatherford 135.

Rashid al-Din notes, for example, that "Houlagou [i.e., Hulegu], to satisfy his princess, became a great benefactor of the Christians, so much that throughout the empire, new churches sprung up everywhere, and by the tent of Dukouz-Khatoun [i.e., Dokuz Khatun], a chapel was constantly kept where bells were rung."³⁵¹

As a result of their tolerance and interreligious marriages, it was not uncommon that the Great Khans attended religious services. Rubruck describes Mongke attending mass with his wife, though he reports that the Great Khan refused to attend Easter services because of the Mongols' abhorrence of corpses and death. For the same reason, Rubruck and other European travelers note, crosses used by the Mongol Christians were plain, with no effigy of the crucified Christ attached to them.³⁵² Akanc' also reports that Hulegu, upon taking the city of Jerusalem, "entered the Church of the Holy Resurrection and prostrated himself before the Holy Sepulchre."³⁵³ The absence of details about the khans attending other religions' services stems from the fact that most accounts of life at the Mongol camps are the work of Christian monks. One can only assume that the khans attended services indiscriminately. Juzjani, for example, praises Batu for being "friendly to the Musulmāns," adding that at his camp, Batu had an imam with a "regular congregation."³⁵⁴ In addition, Marco Polo reports that Kubilai indiscriminately attended services from all religions, and we can thus assume that other khans did the same. Describing Kubilai's participation in the Easter celebrations, Marco Polo writes that "this was his [Kubilai's] usual practice upon each of the Christian festivals, such as Easter and Christmas; and he observed the same at the festivals of the Saracens [i.e., Muslims], Jews, and idolaters."³⁵⁵

³⁵¹ Rabban Sauma 156-162; Rashid al-Din 93-95, own translation; Saunders 97, 129-130.

³⁵² Rubruck 161, 178.

³⁵³ Akanc' 81. Note: Although Akanc's account may be biased because of his own Christian faith, his abhorrence of the Mongols makes it unlikely that he would invent the episode and intentionally make the Mongols look good.

³⁵⁴ Juzjani 1171-1172.

³⁵⁵ Marco Polo 118.

Although their families included members of various religions and they occasionally attended religious services themselves, the Great Khans always stuck to their religiously independent position. While the regional khans tended to be more influenced by their surroundings and some eventually adopted their region's dominant faith—the Ilkhans and the Golden Horde became Muslim in the 14th century—during the unitary period of the empire, the Great Khans strictly enforced their own religious neutrality. In fact, to entertain the free competition of religious beliefs that constituted the basis of their religious policy, several of the Great Khans reportedly organized theological debates at their courts with scholars of the various faiths. The debates took place in front of an audience and a panel of judges from the religions involved, and often had several rounds. Rubruck mentions one such debate he participated to as a Christian representative while at Mongke's court. They focused solely on theological ideas, not politics, and interestingly, intellectual strategies sometimes led the participants to unexpected alliances—Rubruck and fellow Christians joined forces with the Muslims to win against the Buddhists. The debates reinforced the non-belligerent nature of faith in the empire; they were mostly good-natured and as in the case of Rubruck's debate, often ended, after many hours, before a winner could be declared because of the amount of fermented mare's milk absorbed by participants, judges, and audience alike. Juzjani describes similar debates held between ecclesiastics of various faiths in front of Great Khan Guyuk. Questioned by Rubruck after the debate about the Mongols' reasons for refusing to be converted to any particular religion and for respecting the freedom of each to worship their own way, Mongke explains:

Just as God gave different fingers to the hand so has He given different ways to the men. To you God had given the Scriptures and you Christians do not observe them. You do not find in the Scriptures that a man ought to disparage another, now do you?³⁵⁶

³⁵⁶ Rubruck 190-195; Juzjani 1160-1164.

This is a very powerful statement, especially reported by a fervent Christian missionary like Rubruck who fought heretics vigorously, that stands in stark contrast to the edicts and behaviors of all other kings and political leaders of the time. As Weatherford pointedly concludes, "while the clerics debated at Karakorum, their religious brethren were hacking at each other and burning one another alive in other parts of the world outside the Mongol Empire." Everywhere else religious intolerance prevailed—widespread torture of heretics by the Medieval Inquisitions, burning of thousands of Jewish manuscripts in France under Rubruck's sponsor Louis IX, repeated European crusades against Muslims and corresponding repression of Christians in the Middle East.³⁵⁷ Religious tolerance was therefore a unique trait of the Mongol rule. It was probably one of the most essential pillars of the Mongols' state building effort because it enabled them to transcend the religious diversity of their lands and prevent religious quarrels from damaging the unity of the empire, but it also garnered them the support and loyalty of the countless religious minorities enclosed in the empire that could worship freely only thanks to the Mongol rule.

9. Egalitarian Social Structure

A similarly influential pillar of Mongol state building—and a revolutionary measure at the time—was the reform of the social structures of society, from rigid, hierarchic classes toward new loyalties. Stemming from the lower, black-boned tribal lineage and not from the white-boned aristocracy of the steppes and having been constantly reminded of his lesser standing while growing up, Genghis Khan made the abolition of traditional social ranks and privileges one of his priorities. The reform was embodied in the decimal system. As soon as he became

³⁵⁷ Weatherford 173-175.

Great Khan, Genghis Khan abolished the organization by tribes and the tribal hierarchy and blended all social ranks into the new decimal order of households that artificially mixed kinship and lineages, tribes, religions, and ethnic origins and that subsisted even in peacetime. The military commanders, appointed by Genghis Khan on merit to head the decimal units, led their men in war and their families in peace, creating a new system of loyalty not based on traditional class ranks. The new structure was rigid, with each man and household prohibited from leaving the unit where they were assigned and required to do the work they were asked regardless of previous wealth or social standing. Genghis Khan believed that this rigidity would serve to blur social classes for good and, by creating new, "fictive" kinships and allegiances, lessen the risk of local warlords and dissent. The new close relations that replaced the old birth-right system were forged through multiple marriages, sworn brotherhood, and webs of friendship, Onon concludes.³⁵⁸

The Mongol social structure virtually erased all traces of nobility and imposed a strictly egalitarian society with a merit-based social ladder. Even within families, there was no favoritism, and the Mongols treated their different wives and children from different polygamous marriages equally. As Carpini explains, in the Mongol society "there is no difference between the son of a concubine and the son of a wife, but the father gives to each what he will." Even in the ruling family, Carpini adds, "the son of a concubine is a prince just as the son of a legitimate wife." This explains Guyuk's odd resolution of the Georgian succession crisis in 1248, where he ordered the legitimate and illegitimate sons of the deceased king to rule jointly.³⁵⁹ The insistence on social equality earned the Mongols the support of the average population, who was only too happy to see the aristocracy stripped of its privileges, which were rampant in most societies of

³⁵⁸ *Secret History* 8; Phillips 40-42; Weatherford 20-21, 44-45, 52, 58.

³⁵⁹ Carpini 38, 40-41.

the time. Throughout their conquests, the Mongols displayed special favor for men of high lineages, and on the contrary sought to bring them to their knees because they were generally the bearers of power and authority. The Mongols "never spare the noble and illustrious men," Carpini notes, and if such men are taken prisoner, "they can never afterwards escape from captivity either by entreaty or bribe," as was common practice everywhere else. Carpini, a traditional European himself, was outraged by the Mongols' treatment of aristocrats in the empire. "It is their object to wipe off the face of the earth all princes, nobles, knights, and men of gentle birth," he exclaims.³⁶⁰ He was perhaps not exaggerating. Juvaini echoes Carpini's statements, but finds the practice praiseworthy. "It is one of their [the Mongols'] laudable customs that they have closed the door of ceremony, and preoccupation with titles, and excessive aloofness and inaccessibility, which things are customary with the fortunate and the mighty." The Mongols rejected all titles of nobility. Their supreme leaders received the title of *khan*, which simply meant 'commander,' but everyone else was always addressed by their birth name, when present or absent, in writing or direct talk, thus in effect creating a society "making no difference between the sultan and the commoner," Juvaini concludes.³⁶¹

10. Gender Equality

A logical extension of the Mongols' rejection of ethnic, national, religious, and birth preferences, the tenth and last pillar of Mongol state-building was the promotion of gender equality to an unprecedented extent in the context of the Middle Ages. Although the Mongols were a paternalistic and polygamous society—a shocking practice to the European observers who visited them—they gave women a much larger role both in social and political life than any

³⁶⁰ Ibid 17, 44.

³⁶¹ Juvaini 26-27.

other civilization of the time. The Mongol women were such central actors in the everyday life of the Mongol camp that it could not function in their absence. As Marco Polo points out, "it is the women who attend to their [the Mongols'] trading concerns, who buy and sell, and provide everything necessary for their husbands and their families; the time of the men being entirely devoted to hunting, hawking, and matters that related to the military life." Rubruck concurs, observing that "it is the duty of the women to drive the carts, to load the horses onto them and to unload them, to milk the cows, to make the butter and grut, to dress the skins and to sew them ... The women also make the felt and cover the houses." They were thus vital not only to everyday life but also to the conduct of war since leather and felt constituted essential elements of the warrior's attire.³⁶²

In fact, according to all accounts, Mongol women were trained for hunting and war as riders and archers just like their male counterparts. An astounded Rubruck remarks that "all the women sit on the horses like men, astride." Carpini adds that "young girls and women ride and gallop on horseback with agility like the men. We even saw them carrying bows and arrows ... [They] are able to endure long stretches of riding ... All the women wear breeches and some of them shoot like the men." He goes on to remind his reader that the Mongol women sometimes followed the warriors on campaigns and served as decoys by standing on horseback near battlefields to artificially swell the ranks of the Mongol army. Ibn al-Athir goes further and cites some reports of Mongol women fighting in the army, something entirely unheard of. About the Mongol attack of the city of Safar, Azerbaijan, on March 30, 1221, Ibn al-Athir writes, "I was told that a Tatar [i.e., Mongol] woman entered a house and killed several of its inhabitants, who thought that she was a man. She put down her arms and armor and—there was a woman!"³⁶³

³⁶² Marco Polo 89, Rubruck 102.

³⁶³ Rubruck 103; Carpini 18, 36; Ibn al-Athir 216-217.

But the role of Mongol women did not stop at these crucial everyday tasks and military preparations; they were also central figures in political life. Their first political role was as advisors. The *Secret History* repeatedly mentions that Genghis Khan sought the guidance of his wife Borte and his mother Hoelun in certain decisions, and he took their advice seriously. Similarly, Rashid al-Din writes that Mongke, instructing Hulegu before sending him off to the Persian campaign, reportedly told him: "Never fail to consult Dokouz-Khatoun [Dokuz Khatun, Hulegu's main wife] and to get her advice in all circumstances."³⁶⁴ In addition to their private guidance, Mongol women were also appointed to official positions. The *Chronicle of Novgorod* mentions that during the second campaign on Russia, Batu sent a female diplomat to the Russian Prince of Ryazan to demand delivery of the ten percent tribute tax. Women in such prominent roles were ill-perceived outside the immediate Mongol society, however, as the chronicler's reaction suggests—he calls the envoy "a sorceress."³⁶⁵

Finally, some women in the Great Khans' immediate surrounding reached particularly prominent places in Mongol politics during the 1241-1251 decade. By the time of Ogodei's death in 1241, the widow of his brother Tolui, Sorkhokhtani Beki, essentially ruled the territories of Upper Mongolia and Northern China and the widow of his brother Jaghatai, Ebuskun, served as the regent of the Central Asian and Turkestan region. At the same time, Ogodei's favorite wife Toregene, who had already been largely in charge of the empire at the end of his reign because of his alcoholism, became the regent of the empire for five years because Ogodei's and Tolui's sons were unable to agree on a candidate to succeed Ogodei. Toregene assumed complete power in early 1242, becoming the first Great Khatun of the Mongol empire. She fired several of Ogodei's top administrators, including Chinkai and Yalavach, and appointed her own in their

³⁶⁴ *Secret History* 226-227, 229-230; Rashid al-Din 145.

³⁶⁵ *Novgorod* 81.

place, most importantly Fatima, a Persian Shiite who had been captured during the Khwarezmian campaign and who became one of the most powerful figures of the empire. When Guyuk was finally elected in 1246, he publicly dismissed and executed Fatima, but apparently more for strategic than misogynic reasons. After Guyuk's death two years later, his widow Oghul Qaimish took over as regent until Mongke's election in 1251. As Weatherford concludes, "neither gender nor religion hindered their rise to power."³⁶⁶ While women were mostly silent voices in the Middle Ages, enjoying neither power nor status, the Mongols' promotion of a tolerant society offered an alternative by allowing them access to larger roles, both in social and political life, to an extent unmatched elsewhere at the time.

Overall, the Mongols' unique non-military skills and innovations were key factors in their rise to hegemony. First, their novel, growth-oriented economic system supplemented their military might and enabled them to assert their superiority and influence over all competitors. Second, the Mongols' brilliant state-building efforts, resting on ten innovative political, administrative, and social reforms, enabled them to transform military victory into successful peacetime government by anchoring their territorial gains to a solid, lasting institutional structure. Both the economic and the political, administrative and social reforms helped the Mongols transcend the conflict-prone diversity of their empire and generate a cohesion that military conquest alone could not achieve. The reforms finally also allowed for the Mongols' redemption and generated a much-needed positive image for men that initially inspired fear and hatred because of their outrageous brutality.

³⁶⁶ Weatherford 160-161, 164, 166.

Conclusion

The failure of balance of power and the advent of a hegemonic power configuration under the Mongols was thus primarily due to a lack of trust-induced absence of cooperation (IV2.2) and a high vulnerability to deceptive tactics (IV1.3) on the part of the potential balancers, which made them unresponsive to the growing threat of the rising hegemon, and to unprecedented military adaptation (IV5) and non-military innovations (IV6) on the rising hegemon's part that propelled the Mongols out of the reach of all potential balancers. Somewhat surprisingly, physical (IV1.1) and perceptual (IV1.2) communication problems only played a minimal role. Similarly, although some cases of bandwagoning (IV4) occurred, their influence on the outcome seems negligible. Cooperation between the potential balancers was mostly absent altogether (IV2), and not laggard (IV3), and was the result of a lack of trust and a focus on immediate gains (IV2.2). There was little evidence of misperception of balancing interests (IV2.3) and no notable instance of buckpassing (IV2.3).

All four key causal mechanisms of Mongol hegemony that our evidence identified were decisive in triggering the outcome, as a quick counterfactual conjecture shows. First, had potential balancers not been so wary of their neighbors and intent on local gains and had they as a result achieved a vast coalition against the Mongols, it is more than likely that acting early, they would have been able to stop the rising hegemon, because their combined military skills—both quantitatively and qualitatively—far surpassed the Mongols' initial capabilities. Second, had potential balancers not been so receptive to the Mongols' adroit psychological campaign, they would have better been able to assess the real threat and would not have panicked and frozen at the prospect of fighting the Mongols. As a result, they would likely have fared much better both in internal and external balancing attempts, though it is unclear whether that would have sufficed

to stop the Mongols. Third, had the Mongols not exhibited above-average skills of adaptation and innovation in weaponry and military planning, they would not have been able to surpass powerful targets such as the Khwarezmian Empire or the Song, which took decades and tremendous efforts to conquer. The Mongols were only able to achieve victory against those strong competitors because they constantly matched all of their advances while the competitors failed to match the Mongols' own advances in return. Fourth, had the Mongols not complemented their military prowess with sweeping economic, political, administrative, and social reforms, they would most likely have been unable to retain their territorial additions and their growth would have been too ephemeral to engender hegemonic control.

By thus stressing the central role of the Mongols' own achievements (IV5 and IV6) in bringing about hegemony, this chapter confirms a major addition this thesis has sought to bring to balance of power theory. The Mongol example shows that balancers are not the only actors affecting the success or failure of hegemonic bids, as the theory postulates. The ultimate power configuration—hegemony or balance of power—depends partly on the success or failure of balancing movements undertaken to stop a rising hegemon, but it also largely depends on the skills of the rising hegemon itself. As the above counterfactuals demonstrate, the Mongols would most likely not have achieved hegemony without their considerable military and non-military innovations, regardless of the action or inaction of balancers. A potential hegemon is therefore unlikely to become a hegemon if it does not possess considerably superior military and non-military skills, whether other states counteract its rise or not. Hegemony is thus not solely about failed balancing, but also about empire-building, an aspect that Waltz's account of balance of power theory misses by focusing exclusively on system-level variables. Symmetrically, a lack of sufficiently superior skills may be a reason for many balance of power successes, which are often

credited solely to strong balancing movements. Whether balance of power or hegemony ultimately triumphs, then, depends on the balance of skills and capabilities between the rising hegemon and the balancers.

The few difficulties the Mongols experienced on their way to hegemony—against Egypt, Japan, and Southeast Asia—clearly demonstrate that the Mongols' hegemonic skills were crucial to their victory. Against Japan and some Southeast Asian states, the Mongols were defeated by forces beyond human control—they succumbed to uncontrollable elements of nature. Against Egypt and some other Southeast Asian states, however, the Mongols were defeated because their enemies either adopted and exploited their own innovations or shared their unusual adaptation skills and made some great innovations of their own. Thus, the only successful balancing was balancing aimed specifically at outdoing the Mongols' unique skills.

The first reason for some of the Mongols' rare losses had nothing to do with human intervention and amounts to sheer bad luck on the part of the Mongols—bad weather and climate. The Japanese were able to fend off the Mongols not thanks to their own power and capabilities, but because the Mongol expeditions encountered a series of massive storms. Although they were relative novices at naval warfare, sailing ships can do very little when happening upon a typhoon. The only time the Mongols managed to land on the Japanese islands in 1274, they easily won against the samurais. But storms and typhoons repeatedly destroyed their invading forces and compelled them to retreat at each attempt they made to invade the islands. Similarly, in some parts of Southeast Asia, resistance against the Mongols was partly successful, but mostly not because of its unusual strength but because the Mongols the tropical weather and conditions. In the early 1280s, for example, the Mongols were forced to withdraw

from Champa, despite winning initial battles against the Vietnamese, because hunting animals were too scarce in the jungles of Vietnam and the warriors were hit by epidemics of fever.

Notwithstanding those weather and climate-related failures, the only successful human opposition to the Mongols was led by opponents who took direct aim at the Mongols' special skills and tried to imitate or surpass them. One of the only enemies to inflict a battle defeat upon the Mongols were the Mamluks of Egypt, who sent an army up to Syria against the small contingent of roughly 10,000 warriors that Hulegu had left stationed there after heading back to Mongolia in 1260 to attend the kuriltai that followed the death of Mongke. The Mamluks overcame the Mongol army at Ain Jalut in Palestine on September 3, 1260, and then destroyed its few remnants at Hims in Syria on December 10, 1260. The Mamluks won because their army was specifically tailored to match the Mongols' innovative military tactics. First, the Mamluk army was composed solely of a cavalry of archers, just like the Mongols', equipped with similarly well-trained horses. Second, as Rashid al-Din points out, the Mamluk army was composed "in major part of Turcomans, and fugitives from Sultan Jelal ad-Din's army who ... had fled toward Syria." The Mamluk Turcomans were originally Kipchak and Slavic prisoners of war sold as slaves by Batu and eventually resold to the Sultan of Egypt. These former slaves thus shared the Mongols' agility with horses and arrows since they were steppes nomads themselves. Third, both the Turcomans and the former warriors of Jelal ad-Din, who also made up the Mamluk force, had first-hand experience fighting the Mongols in Eastern Europe and Khwarezm, respectively. They were thus fully aware of the Mongols' fighting techniques and did not hesitate to copy them to deny the Mongols any advantage. For example, Akanc' writes, at Ain Jalut a small contingent of Mamluks waited in ambush and "fell upon the Ta'tars [i.e., the Mongols]" from all sides, in an exact copycat of the Mongols' favorite tactics of speed and

surprise, and multiple front and encirclement. Rashid al-Din adds that the Mongols were seemingly able to ward off the initial attack, and went after the fleeing Mamluks, who were in reality carrying out the Mongols' third tactic of predilection, the fake withdrawal and luring tactic. The fleeing Mamluks led the Mongols straight to where the main Egyptian army was waiting, and they attacked the Mongols from three sides at once. The surviving Mongols fled and hid into the nearby woods, Mongol-style, according to Rashid al-Din, but again, the quick-thinking Egyptian commander did just what the Mongols would have done—he set the forest on fire.³⁶⁷

The Mamluks thus successfully stopped the Mongols at Ain Jalut because they learned and copied the Mongols' innovative military techniques. The Mamluks were the only balancers who emulated the Mongols' unique tactics, and we can speculate that they were able to do so because they shared the Mongols' culture and experience as self-reliant steppes nomads, which had inspired most of those tactics in the first place. Adopting the Mongols' military tactics was therefore easier for the Mamluks than it would have been for the other potential balancers. The Mongols' skills were in fact far beyond the grasp of their other targets who did not share like the Mamluks the Mongols' background and thus sensibility to those superior techniques.

The Mongols' more minor losses in parts of Southeast Asia only also confirm the centrality of the rising hegemon's skills in the equation for balance of power and hegemony. The Southeast Asian states that successfully booted the Mongols off their lands did so because they counter-innovated and surpassed the Mongols' skills. After taking Hanoi in their second invasion of Champa in 1287, for example, the Mongols were defeated in a decisive naval battle off the coast of Haiphong because the Vietnamese came up with a revolutionary tactic: they waited in high tide, then lured the Mongols into a shallow area where they had submerged iron-spiked

³⁶⁷ Akanc' 81; Rashid al-Din 343-353, 359; Phillips 115-116; Weatherford 159, 185; Turnbull 60.

poles, so that when the tide receded, the Mongol ships were damaged by the poles and forced to withdraw.³⁶⁸ In the end, the Mongols were only stopped by enemies that adopted their own innovations or came up with innovations that surpassed the Mongols' own. Because the Mongols' unique skills were key to their rise to hegemony, they also lay at the core of any successful attempt to balance against them.

In order for balancing to be successful, therefore, it may not be sufficient to simply increase the weaker side's capabilities by internal efforts or external alliances, as the structural approach to balance of power suggests. The weaker side's capabilities have to be built up with a purpose—to outwit the rising hegemon by matching and improving upon its unique skills. In the case of the Mongols, too few balancers were able to grasp this. The Mongols, however, did, since copying and outdoing their enemies' best weapons, techniques, and strategies constituted the core of their path to hegemony. As Turnbull concludes, wherever they went, "on the Russian steppes, outside China's walled towns ... the Mongol armies demonstrated that their reputation did not depend on mythical accounts of eternally galloping horsemen, but on something much more solid: the greatest example ever demonstrated of the ability to change."³⁶⁹ Unfortunately for the Mongols, their unique propensity to learn and improve also caused their downfall. By embracing new ways they eventually blended into the surrounding, preexisting civilizations of Persia, Russia, and China, and by doing so lost the very quality that gave them an edge over those civilizations in the first place—a simple, nomadic, steppes lifestyle.

³⁶⁸ Phillips 107-108; Turnbull 82-83.

³⁶⁹ Turnbull 93.

The Rise of Rome I Rome's Potential Balancers

The Roman empire constitutes the second largest hegemon the world has experienced. Unlike the Mongol empire, however, Rome did not reach the scale of historic global hegemon. Despite its overwhelming size, Rome remained a regional hegemon. While it ruled over the entire Mediterranean world, it did not control the majority of the world as it was then known, like the Mongols did. Roman influence did not extend to most of the Middle East or to East and South Asia, where other great powers, and even another regional hegemon, the Qin dynasty of China, simultaneously existed. Those lands were known to the Romans, and historical evidence shows that they recurrently interacted. The Chinese emperors sent embassies westward, for example, as early as the 2nd century B.C., and contacts multiplied between Rome and the Qin's successor, the Han, first through the lucrative silk trade and then through the exchange of diplomatic envoys.

Though not the sole hegemon of its time, Rome stands out by the vast extent of territory it controlled, in what was then the wealthiest, most densely populated, most powerful and richly civilized, and thus perhaps the hardest region to dominate, the Mediterranean perimeter. At the apogee of its power in the first century A.D., Rome spread over some 2.2 million square miles and ruled over about 120 million people, roughly two-fifth of the world's population, half of which were Roman citizens. Roman control stretched from Scotland and the Iberic peninsula in the West, to Iraq and Armenia and the Caspian Sea eastward, and from Germany and Romania in

the North to North Africa in the South, encompassing Morocco and Egypt and all lands in between. Rome not only aggregated the territories of previous empires, including the Hellenistic world with all of Alexander the Great's possessions, and the Near Eastern half of the former Persian Empire, but it expanded far beyond what they had achieved. And Rome's accomplishment is not limited to territorial expansion. The societal organization the Romans established, along with its administrative, legal, political, economic, social, and military structures, became the cornerstone of the modern European civilization.

The story of the rise of Rome is particularly remarkable because Rome started as an obscure city-state in a region dotted with a plethora of city-states and kingdoms, and while others, much better endowed entities like Athens and Sparta also attempted to become hegemon, they utterly failed, while Rome, the less likely one, succeeded. As Ancient historian G.H. Stevenson acknowledges, "no state in history may seem to have been so well qualified as fifth-century Athens to embark on a career of Imperialism." This naturally makes one wonder, "how was it then that Rome succeeded in performing a task which had proved too hard for Athens?"³⁷⁰

Even more puzzling is the fact that the Roman empire distinguishes itself by its longevity. Unlike the Mongols and most other regional empires like the Qin, Rome holds the record of the longest surviving hegemon as it ruled the Mediterranean world as a unitary hegemon for at least three full centuries until its permanent division into Eastern and Western empires at the death of Theodosius I in 395 A.D., testifying to a particularly successful recipe for hegemony. In its rise it planted the seeds of a long-lasting dominance, an inexplicable aberration for balance of power theory. While short-lived hegemon may be passed over by balance of

³⁷⁰ G.H. Stevenson, *Roman Provincial Administration Till the Age of the Antonines* (Oxford, UK: Basil Blackwell, 2nd Ed., 1949), 4-6.

power theory as accidental occurrences, a hegemon that thrived for a total of eight centuries from its inception to its downfall is impossible to ignore.

This chapter argues that the roots of Rome's successful and long-lasting rise to hegemony lay both in the potential balancers' difficulties and in Rome's unusual achievements. On the one hand, the potential balancers were unable to cooperate and invariably tended to bandwagon with Rome rather than oppose it. On the other, Rome displayed exceptional military and diplomatic skills, as well as a perceptive understanding of government. While the multitude of city-states and kingdoms in the region remained focused on short-term gains at the expense of their neighbors and became side-tracked by Rome's appealing offers of support, the Romans developed a highly effective military, political, social, and economic apparatus tailored to accommodate their growing possessions and generating the support rather than the resentment of their new subjects and allies.

Timeframe, Boundary of Study, and Chronological Considerations

Paradoxically, Rome's ascent to hegemony begun with the sack of the city by the Gauls around 387 B.C. Although Roman history was not systematically recorded in written form until the 3rd century B.C. when Roman rise was already underway, and histories of the earlier period rest on orally transmitted stories that may not be entirely reliable, all accounts suggest that the barely avoided annihilation of the city shocked the Romans into action. The Gallic tribe of the Senones made its way into Italy, defeated the Roman army some eleven miles north of the city, then moved into the city and fought the Romans in the streets of Rome. Polybius tells us that the Senones "occupied the whole of Rome with the exception of the Capitol," before eventually

withdrawing, distracted by an invasion of their own lands by a rival tribe.³⁷¹ Then a small, relatively isolated city-state, Rome realized that it needed to secure its backyard to ensure its survival and prevent a takeover by rival city-states or tribes in the future. The sack of Rome thus marked the start of a quest to control the periphery of Rome, envisioned in increasingly larger concentric circles as a "manifest destiny" required by survival.³⁷² Roman expansion took place in four phases: first, the immediate Italian peninsula (ca. 390 to 272 B.C.), then the whole Western Mediterranean with Spain and North Africa (264 to 201 B.C.), followed by the Hellenistic Mediterranean with the Greek isles, Macedonia and Asia Minor (214 to 146 B.C.), and finally Northern Europe and the Near East (133 B.C. to the first century A.D.).

At the time of the sack of Rome, there was no dominant power in Italy and the Mediterranean region, which constituted a highly fragmented multipolar system comprising a variety of political units—mainly kingdoms, city-states, and tribes. The Etruscans, whose kings had ruled Rome until Rome overthrew them around 509 B.C. and became an oligarchic republic, controlled central Italy North of Rome; Gallic tribes roamed around the Alpine regions; the Phoenicians from the Near East dominated Western Sicily, Southern Spain, and parts of North Africa through their colony at Carthage; and the Greeks had implanted their own colonies in the form of city-states all around the Western Mediterranean, from Southern Italy with Cumae and Tarentum, and Syracuse in Eastern Sicily, to Massilia (Marseilles) in Southern Gaul, with outposts in Spain.³⁷³

Rome's first move after the Gallic invasion was to establish its control over mainland Italy, starting with the Latium region to which Rome belonged. First, the Romans cooperated

³⁷¹ Polybius, *The Rise of the Roman Empire* (transl. by Ian Scott-Kilvert) (New York: Penguin Books, 1979), I.12, II.18, III.4.

³⁷² Chester G. Starr, *The Emergence of Rome as Ruler of the Western World* (Ithaca, NY: Cornell University Press, 1950), 23.

³⁷³ *Ibid.*, 14-16 and 17-20.

with the Latin League, a defensive alliance formed by fellow Latin city-states and villages, to repel the neighboring hills tribes of the Volsci and Aequi that continuously threatened the inhabitants of Latium. A subsequent war with members of the Latin League (340-338 B.C.), who were resisting Rome's increasing power in the League, enabled Rome to establish pre-eminence over the whole Latin region. After a series of wars (343-341, 326-304, and 298-290 B.C.) with the Samnite tribes that controlled central Italy Southeast of Rome and repeatedly attacked the Latins and Campanians South of Rome, Rome dominated the entire Italian peninsula through annexation, colonization, and alliance, except for the Northern Po valley still controlled by the Gauls and a few Greek city-states in the extreme South of the peninsula. Rome seized the opportunity to complete its takeover of Italy when the Southern Greek city of Tarentum appealed to King Pyrrhus of the Greek state of Epirus across the Ionian Sea ten years later, for help in a conflict against Rome over a naval treaty violation. Rome's defeat of Pyrrhus, its first opponent from outside of Italy, and takeover of Tarentum in 272 B.C., consolidated its supremacy over Italy but also "pointed the way to a new epoch ... in Roman expansion" directed overseas.³⁷⁴

In the second phase of its hegemonic rise, Rome focused its attention on its immediate overseas neighbors in the Western Mediterranean. By deliberately mingling in the affairs of Sicily just a few years after their victory over Pyrrhus, the Romans provoked a conflict with Carthage, Phoenice's powerful North African colony and one of the three great powers of the region besides Rome, that enjoyed naval supremacy in the Western Mediterranean and controlled the Western part of the island. The feud lasted over a century and resulted in three wars. Rome's victory in the first Punic War (264-241 B.C.) allowed it to dominate Sicily and eventually also the two neighboring islands of Sardinia and Corsica, which became the first Roman provinces outside of Italy. While Rome simultaneously secured victories against the Gauls (225 B.C.) and

³⁷⁴ Andrew Lintott, *Imperium Romanum: Politics and Administration* (New York: Routledge, 1993), 5-6.

was able to expand its territory to the Northern fringe of the Italian peninsula, it was unable to forestall Carthage's growth in Spain and its retaliatory invasion of Italy led by Hannibal in 218 B.C. Yet the remarkable comeback of Rome and its decisive defeat of Carthage during the second Punic War (218-201 B.C.), followed by a brief third Punic War (149-146 B.C.), spread Roman domination to Carthage's former settlements in Spain and North Africa, which also became Roman provinces. As Polybius points out, those exploits "encouraged [the Romans] to stretch out their hands for the first time to ... cross with an army into Greece and the lands of Asia" and pursue expansion beyond their close backyard.³⁷⁵ Indeed, at this point the region, though less fractured already due to the Roman consolidation in the West, still represented a multipolar system since there remained three great powers in the Hellenistic world: the Antigonid dynasty of Macedon; the Seleucid empire of Syria, the successor of Alexander the Great's empire; and the Ptolemaic dynasty of Egypt.

The Mediterranean configuration of power shifted dramatically in the third phase of Rome's rise, which witnessed the downfall of the three remaining Hellenistic great powers. In between the Punic Wars Rome had already begun turning its attention toward Greece by securing a few allies such as Sparta, Messene and the Aetolian League in the Peloponnese, and Rhodes and Pergamum in the eastern Mediterranean. Rome's growing relations with the Greeks soon led it to become entangled militarily across the Adriatic with Illyria, whose frequent raids along the coast and use of piracy caused grief among Rome's allies and disrupted Roman commerce with the Greeks. The Romans conducted and won three wars against successive Illyrian kings (229-228, 220-219, and 169-168 B.C.), after which they finally conquered the recalcitrant Illyria and made it a Roman province. Meanwhile, the regime of the Ptolemies in Egypt, which may have been considered the fourth and weakest of the region's great powers, suffered an internal collapse

³⁷⁵ Polybius I.3

in 207 B.C. as a result of a massive indigenous rebellion, creating a void that the remaining great powers, Rome, Macedon, and Syria, immediately sought to take advantage of.³⁷⁶ The powerful Philip V of Macedon had become increasingly menacing to Rome during the Second Punic War, having allied with Hannibal. Fearful that Philip might supply reinforcements for Hannibal's occupation of Italy, Rome preempted the threat and sent forces to Macedon (214-205 B.C.). The outcome of this First Macedonian War was indecisive but paved the way for further Roman military intervention in Greece, which occurred three years later when Rome got wind of alliance between Philip and Antiochus III of Syria, its two most powerful competitors.

After defeating Philip in 196 B.C. and forcing him to withdraw from all its possessions in Greece, Rome withdrew back to Italy, but Antiochus III, who had been raiding the weakened Egyptian kingdom for a few years, renewed the threat by deliberately leading his armies across the Hellenic peninsula into Greece in 192 B.C. Again Rome intervened on behalf of its Greek allies, severely defeating Antiochus in 188 B.C. and eradicating his influence from Europe and a large part of Asia Minor. When Macedon under Philip's successors Perseus and later Andriscus attempted to reassert its power, Rome defeated them in the Third and Fourth Macedonian Wars (172-168 and 150-148 B.C.) and finally annexed Macedonia as a province, pulling the entire Greek peninsula under its control.³⁷⁷ With the elimination of the Antigonids and the Seleucids, and the collapse of the Ptolemies, Rome remained the sole great power and firmly established its supremacy over the Mediterranean basin. With only a few weaker powers, mostly poorly

³⁷⁶ Theodor Mommsen, **Römische Geschichte** – *Erster Band: Bis zur Schlacht von Pydna, Drittes Buch* (Berlin: Weidmannsche Buchhandlung, 8th Ed., 1888), 694-704.

³⁷⁷ Lintott 7-9.

organized tribal entities and small city-states, remaining around Rome, the configuration of power in the region had become unipolar.³⁷⁸

In the final phase of Roman growth, Rome consolidated its rule over the Mediterranean perimeter by annexing the remaining weaker powers like the remnants of Ptolemaic Egypt and of the Seleucid empire in Asia Minor, and conquering more provinces in its tribal periphery: in the Northern frontier Rome acquired Gaul, Germany, and the plain of the Danube; in the South it added Portugal and Southern Spain and Northern Morocco; and in the East Rome eventually took over Thrace, Moesia, and Judea. By eliminating all other powers in the greater Mediterranean region, Rome attained hegemonic status. But this last phase was the slowest and longest, because it coincided with the internal transformation of Rome from an archaic city-state republic ill-equipped to handle such vast territory into a full-fledged imperial regime designed to administer the newly acquired lands. "The vast profits of foreign conquest placed a great strain on the Republican system of government," military historian Adrian Goldsworthy notes.³⁷⁹

Yet this mutation, which enabled hegemony to materialize by giving Roman rulers the capacity to effectively control the large Roman territory, did not occur smoothly. It was precipitated by a period of intense civil unrest beginning in 133 B.C. and streaked by civil wars, first between Rome and its Italian allies in the Social War (91-88 B.C.) that resulted in the formal incorporation of all Italians into Rome, and then between various Roman factions for reform and command of the regime, from 88 B.C. to the final struggle for power between Cleopatra and Mark Antony and Julius Caesar's adopted son Octavian, who prevailed in 31 B.C., unified the Roman provinces under the new imperial regime of the Principate, and became the first Roman emperor, Augustus.

³⁷⁸ Arthur Eckstein, *Mediterranean Anarchy, Interstate War, and the Rise of Rome* (Berkeley, CA: University of California Press, 2006), 1-3.

³⁷⁹ Adrian K. Goldsworthy, *The Complete Roman Army* (London: Thames & Hudson Ltd., 2003), 6-7.

By the end of Augustus's reign in 14 A.D., the great period of the Roman conquests was essentially over. Aside from a few exceptions—the conquest of Britain under Claudius (43 A.D.); the annexation of Dacia and Arabia and short takeover of Armenia, Assyria, and Mesopotamia under Trajan (d. 117 A.D.)—the boundaries of the empire were set and his successors focused primarily on consolidating them. The empire remained stable until the third century, and Rome presided over 250 years of hegemonic Pax Romana: in the South, the empire had reached the limit of the African desert; in the West, it stopped at the Atlantic Ocean; the Northern frontier was delimited by the Rhine and the Danube; and in the East, by the Iranian plateau.³⁸⁰ The first century A.D. thus provides the logical termination point for this inquiry. By the first century A.D., as military historian Arther Ferrill concludes,

except for barbarians, who could not compete militarily with a united and politically strong empire, Rome faced no military threat. Because of their primitive social, political, and economic organization, there was no arena of competition—diplomatic, economic or military—where the inhabitants of Free Germany [or Rome's remaining oriental neighbors] had any chance of overcoming the Romans.³⁸¹

The traditional explanation offered by a majority of Classical scholars in the literature for the formidable growth of Rome focuses on the character of the Romans. Emblematically, Greco-Roman historian William V. Harris argues that the remarkably violent, revisionist and revolutionary nature of Roman society constituted the driving force behind Rome's militaristic and expansionist foreign policy.³⁸² Yet characteristics highlighted by Harris remain insufficient to account for Rome's successful rise because they were hardly limited to Rome at the time, and

³⁸⁰ Moché Amit, "Les Moyens de Communication et la Défense de l'Empire Romain." *La Parola del Passato – Rivista di Studi Antichi* 20.102 (1965), 207-208.

³⁸¹ Arther Ferrill, *Roman Imperial Grand Strategy* (Lanham, MD: Publications of the Association of Ancient Historians 3—University of America Press, 1991), 19.

³⁸² William V. Harris, *War and Imperialism in Republican Rome: 327-70 B.C.* (Oxford, UK: Oxford University Press, 1979), 9-41, 51.

were instead largely embraced by all its neighbors in the Mediterranean, be they great-, second-rank-, or even minor powers. While much of the literature has emphasized the bellicose nature of Roman culture as a causal factor, "the conceptual framework and theoretical insights of Realism [and more particularly, of balance of power] have ... never been applied in detail to the study of Roman expansion," deplures historian Arthur Eckstein. This is "unfortunate, for Realist concepts of state interaction have much to contribute to our understanding of the emergence of Roman hegemony first in Italy and then in the Mediterranean," Eckstein concludes.³⁸³ Indeed, as the first section suggests, the fierce anarchical competition in the Mediterranean was largely responsible for the lack of successful balancing against Rome.

³⁸³ Arthur M. Eckstein, *Mediterranean Anarchy, Interstate War, and the Rise of Rome* (Berkeley, CA: University of California Press, 2006), 3 and 7.

Map 5.1: The Roman Empire at its Height, 125 A.D.



Source: Andrei Nacu, *Wikimedia*, released into the public domain.

The first root of Rome's successful rise lies in the potential balancers' actions. Strong balancing could have put serious hurdles in Rome's path and ultimately prevented its ascension to hegemonic status, but Rome's fellow powers in the Mediterranean failed to act. Among the great powers, Carthage was undoubtedly the most powerful besides Rome, followed by the roughly even Macedon and Syria, and Egypt at the bottom. The great powers were surrounded by a myriad smaller powers, both kingdoms, city-states, and tribal units. Competition among this unusually high number of political units intensified collective action problems and encouraged

potential balancers to focus on immediate gains and disputes with their own neighbors at the expense of the Roman threat. It also made Rome's neighbors more vulnerable to Rome's deceptive tactics and appeals to join, rather than oppose, the rising hegemon.

1. Communication Problems (IV1)

Just like the Mongols' potential balancers, the Romans' Mediterranean neighbors experienced few physical communication hurdles (IV1.1). They were mostly deprived of the correct information about Rome's hegemonic rise by their own misperceptions of Roman power and intentions (IV1.2), in many ways encouraged and reinforced by Rome's deceptive friendliness and cooperative posture (IV1.3).

Rome's early expansion throughout Italy, and then into Sicily, did not remain unnoticed (IV1.1). The movement of nomadic and semi-nomadic tribes constituted a first vehicle that spread the news of Rome's growth. The Alpine Gauls had been conducting raids into southern Italy periodically since decades before the sack of Rome and their brief alliance with the Samnites against Rome in 298 B.C. during the Third Samnite War confirms that they not only were aware of Rome's increasing power but also felt threatened by it. Since Gallic tribes migrated regularly around the Mediterranean in the 4th and 3rd century B.C., reaching as far as Thrace and Greece, information about Roman growth must have traveled with them from Northern Italy, and also back to their home territory in Gaul. Colonial relations constitute another way the information spread. The more than thirty Greek colonies in Southern Italy and Sicily were communicating with their mother-cities and -states in Greece, carrying reports of Italian developments with them. The Carthaginians were also informed through their Sicilian colonies. Trade relations, finally, were probably the furthest-reaching means of communication. The

Carthaginians, like the Greeks, were great merchantmen and their regular commercial sea journeys reached most lands between the Atlantic coast and the Black Sea and Orient, disseminating information about the growth of the Romans throughout the region and beyond.

Although the pace of news in the ancient Mediterranean remained unimpressive by modern standards, it was largely sufficient to warn Rome's close and distant neighbors in time to organize a response. Indeed, Roman armies were slow too during the Republic. As citizen-armies, they campaigned only in the hot summer months, and released their soldiers to their land duties for most of the year, which gave ample time for the news of their progress to travel, and for potential balancers to prepare. There is no doubt that everyone around the Mediterranean was aware of Rome's growing power, even though its expansionist predilections might not always have been obvious. Even before the end of the Second Punic War and Rome's expansion beyond Italy into North Africa and Spain, the Greeks knew that Rome had the means to threaten them, as an anecdote shared by Polybius demonstrates. In the summer of 217 B.C., the Greeks sought a settlement to the War of the Allies, a conflict that had been raging for three years between Macedon and the Achean League on one side, and Sparta, Elis, and the Aetolian League on the other, to which Rome did not participate. Polybius, an Achean himself, reports that during the peace conference that ensued, the Aetolian Agelaus of Naupactus warned the Greeks that the victor of the Second Punic War would seek hegemony in the region. Agelaus prophesized:

For it must be obvious to all those who pay even the slightest attention to affairs of state that whether the Carthaginians defeat the Romans or the Romans the Carthaginians, the victors will by no means be satisfied with the sovereignty of Italy and Sicily, but will come here [i.e., to Greece].³⁸⁴

³⁸⁴ Polybius, V.103-104.

Although no Roman ambassador had yet crossed the Aegean Sea and ancient modes of communication remained rudimentary, the Greeks clearly knew a hegemon was in the making. It is not physical communication barriers, thus, that precluded their balancing.

Could Rome's potential balancers, despite possessing the correct information, have misperceived Rome's intentions (IV1.2)? The question of Rome's hegemonic intentions is hotly debated in the literature—did Rome actively seek expansion or just happen to expand while pursuing a self-defense strategy?—but Polybius, perhaps the most reliable ancient source, certainly seemed to believe that regardless of Roman intentions, hegemony would result, as the above quote shows. Indeed, regardless of intentions, knowledge of Rome's sheer power should have indicated a threat and triggered decisive and all-around balancing. Instead, knowledge of Rome's growing power only triggered insufficient, episodic balancing because of Rome's clever deception, which led most potential balancers to dramatically underestimate the extent of Rome's rise and misperceive the danger it represented. Rome used deceptive techniques broadly, particularly in the first and second stages of its rise, effectively shielding itself from attention and retaliation until it had reached great power status and had already become difficult to balance.

Roman deception took on both strategic and tactical dimensions (IV1.3). At the strategic level, Rome showed considerable restraint, especially at the beginning of its rise, which resulted in cloaking the real extent of its power. Rome was careful not to engage in all-out annexations during the first few centuries of its rise. Until around 200 B.C., the Romans would intervene somewhere to weaken or defeat a competitor and then withdraw back to Italy to wait for new developments, even if it meant having to come back several times should the competitor try to regain power. This was the case in Illyria, for example, Rome's first intervention outside of Italy. In 229-229 B.C., the Roman army defeated Queen Teuta of Illyria, who had taken over Greek

settlements along the Adriatic coast and encouraged piracy detrimental to Roman commerce. After having expelled the Illyrians from the Greek cities, Rome installed an Illyrian leader, Demetrius of Pharos, to rule over Illyria, and returned to Italy. But Demetrius soon built a fleet of his own and renewed the expansionist enterprise of Queen Teuta, forcing a new Roman intervention in 220-219 B.C. while Rome was already occupied with the Alpine Gauls and with Carthage. But Rome had to send troops across the Adriatic a third time fifty years later in response to Illyrian provocation before it finally decided to formally conquer the area in 168 B.C. Illyria did not become a province for another forty years.³⁸⁵ The same occurred with Macedon, one of Rome's most powerful competitors. Rome generously withdrew back to Italy after defeating Philip V twice, in 205 and 196 B.C., instead of annexing his land. It was not until Philip's son Perseus renewed Macedon's aggressive behavior and Rome conducted two more wars against it that the Romans finally decided to establish a permanent presence in the Greek peninsula by making Macedon a province in 148 B.C.³⁸⁶

In addition to its moderation, Rome also acted as everyone's friend and liberator, which tricked potential balancers into viewing its power as benevolent and less expansionist than it really was. For example, upon withdrawing from the Balkans after the Second Macedonian War in 196 B.C., Rome claimed that it had freed the Greek peninsula and had no other interest in the region. According to Appian, who criticizes the conditions imposed by Rome on Philip as too mild and "inadequate," the Roman consul Flaminius reportedly declared: "The Roman people and Senate, and Flaminius, their general, having vanquished the Macedonians and Philip, their king, leave Greece free from foreign garrisons and not subject to tribute, to live under her own customs and laws." Unsurprisingly, many Greeks viewed the Romans favorably as a result, and

³⁸⁵ Appianus (Appian). *Roman History Vols. I-III* (transl. by Horace White) (Cambridge, MA: Harvard University Press, Repr. 1982), X.II.7-11.

³⁸⁶ Starr 43-44, 47-49.

allied with them instead of being wary of their growing power.³⁸⁷ Perhaps the best indicator of Rome's attempt to portray itself as a benevolent, helpful neighbor to conceal its real power and potential was that it always sought an invitation to intervene before engaging in a conflict. That way, Rome was never seen as an aggressor but rather as the protector against the aggression of others—a clever way of making military expansion appear benign and unthreatening, as Eckstein points out. Between the 4th century and 164 B.C. Rome mostly spread that way, by "answering pleas for protection" from unwitting smaller powers fooled by Rome's self-restrained behavior. The Greek city-states recurrently appealed to Rome, for example, like against Illyria, and again against Macedon and Syria, offering Rome a perfect excuse to intervene unnoticed instead of allying to stop its spread. Grasping the opportunity, the Roman Senate progressively redefined what constituted a security threat justifying intervention in an increasingly all-encompassing way.³⁸⁸

Rome employed a deceptive strategy from early on. While taking control of Italy, the Romans already displayed a certain reluctance in directly annexing land, preferring instead to lure their neighbors into friendship and collaboration so they would request Rome's help and leadership. For example, Rome cooperated with the Latin League to protect its members against common enemies like the Etruscans and the Volsci and Aequi tribes. But when some Latins requested that the League be transformed into a tighter-knit confederation under Roman leadership, Rome cautiously declined. Stevenson argues that Rome's refusal to the Latins was in fact "one of the turning points of the history of Italy and the world" because "had she [i.e., Rome] agreed to merge her institutions into those of the other members of the League the unity of Italy might have been attained and a powerful federation created," which would have differed

³⁸⁷ Appian, IX.IX.3-4.

³⁸⁸ Eckstein 179-180.

substantially from the slow, controlled expansion carefully pursued by the Romans.³⁸⁹ To be armed with a powerful federation from early on would have thrown greater suspicion and likely attracted much fiercer balancing than Rome faced by spreading gradually under cover of restraint and the benevolent rescuing of friends and allies. Rome's careful rise appeased and misled potential balancers. Instead of a federation, Rome relied on friendship, alliances, and colonies throughout Italy to establish its influence, occasionally granting Roman citizenship to some allies as a flattering gesture reinforcing their allegiance.

Rome's deceptive strategy proved very effective in concealing Rome's early rise and making its neighbors misperceive its intentions and perhaps even the full extent of its power. Although the Mediterranean powers were fully aware that Rome was powerful, the rising hegemon's efforts to appear friendly and restrained may have prevented them from realizing just how considerable its advance was. Most potential balancers, especially the second-ranked powers, were effectively tricked and did not take the indicators of Roman growth seriously. Once it became more powerful, especially once it was close to unipolarity, Rome did not hesitate to switch to a more heavy-handed strategy— direct annexation—abandoning its earlier subtlety toward the end of the Republic and under the Principate, when most competitors had been eliminated.

In addition to those strategic uses of deception, the Romans also excelled in tactical tricks. They frequently devised diversionary attacks to draw the enemy away from where they intended to attack, traps to divide enemy forces, and feigned withdrawals to surprise the enemy, not only in open battles but also in siege warfare. One great example mentioned by Polybius was the Roman conquest of the island of Pharos, which occurred in 219 B.C. during the second Illyrian War. Foreseeing a long siege because the main city had ample provisions and military

³⁸⁹ Stevenson 7-8.

supplies, Roman consul Aemilius Paulus decided to use deceptive tactics instead of a classic siege. He sailed to the island at night and ordered the majority of his troops to hide in wooded valleys close to the city. At sunrise he sailed with the rest of the men in twenty ships in plain sight into the main harbor. Demetrius fell for the trap. "When Demetrius saw this naval force, he was filled with contempt at its apparent weakness and led a sortie from the town to the harbor to prevent the enemy from landing," Polybius describes. But the Roman ships fought hard and forced Demetrius to draw more and more soldiers out of the town and onto the sea, until a majority of his men were involved in the naval clash. That is when the hidden Roman force arrived via a "concealed route" and cut off Demetrius's men from the town, which was soon taken over by the Romans.³⁹⁰

The Romans also did not hesitate to use psychology to erode its enemies' resolve in conflict. The best example is perhaps the 70-73 A.D. siege of Masada, an impenetrable mountain fortress in Judea, recounted by Jewish-Roman historian Josephus.³⁹¹ The stronghold contained only about a hundred enemies, and the fortress itself presented no strategic value to the Romans since the rest of the Jewish War was won. Masada was the only resistance left, and the Romans did not need the fortress since they had already amply demonstrated their superiority in the region. Yet governor of Judea Lucius Flavius Silva still meticulously focused his forces on the fortress, spending massive amounts of resources. The conquest of Masada required engineering wonders, including the construction of a gigantic ramp as high as the mountain, for Roman infantry and artillery to gain access. This unlikely commitment was made solely as a "calculated act of psychological warfare," military historian Edward Luttwak emphasizes, to deter other Jewish rebels that might have thought of revolting by demonstrating that the Romans would

³⁹⁰ Polybius III.18-19.

³⁹¹ Flavius Josephus, *The War of the Jews* (transl. by G.A. Williamson) (New York: Penguin Books, Repr. 1970), VII.VIII.1-7.

pursue them relentlessly to the last one, even to the most remote locations, regardless of expense.³⁹²

The Romans' tireless deceptive strategies and tactics required thorough intelligence. In fact, the Romans already had a very meticulous service of spies (*speculatores*) and scouts (*exploratores*) during the early Republic, though we must note one major intelligence failure in the invasion of Italy by Hannibal, which the Romans utterly failed to foresee, persuaded that Hannibal would be deterred by having to cross the Alps. During the Republic the spy services were attached to the different armies, and served a specific military intelligence purpose. Their role widened under Caesar and during the Principate, culminating in the development of a centralized secret service in the first century A.D.³⁹³

Roman spies and scouts were often disguised as army suppliers, for example as grain dealers, because of the wide access to information they could obtain that way. But they used all sorts of tricks to obtain information, as Polybius attests. While Scipio Africanus was in North Africa during the Second Punic War, for instance, he became friendly with the Numidian king Syphax but in order to communicate with him Scipio had to send envoys through Carthaginian territory. He took advantage of the opportunity to send spies as his envoys. His messengers

included ... men of tried experience and others of military capacity; they would be humbly and shabbily dressed, disguised in fact in the clothes of slaves, and their task was to examine and explore the approaches and entrances to both [the Carthaginian and the Numidian] camps without interference.³⁹⁴

Upon their return, Scipio had all the information he needed to succeed in the ensuing battle, known as the Battle of the Great Plains (203 B.C.). The battle itself was a diversionary tactic to

³⁹² Edward N. Luttwak, *The Grand Strategy of the Roman Empire From the First Century A.D. to the Third* (Baltimore, MD: Johns Hopkins University Press, 1976), 3-4.

³⁹³ Ferrill 25-27.

³⁹⁴ Polybius XIV.1.

trick Hannibal into leaving Italy where he was making much damage and coming back to Carthage by defeating the Carthaginians on their home turf in North Africa. The trick worked; after loosing the battle, the Carthaginians recalled Hannibal to Africa to continue the fight there, and Italian territory was freed of danger again.³⁹⁵ Once Rome had firmly established its hegemonic control in the first century A.D., however, the spies mostly reverted to internal, rather than external, surveillance, since for the hegemon's stability it was more important to keep track of detractors than gather intelligence about barbarian tribal movements across the border. As a result, the Roman spy services have been ascribed a nefarious reputation.

Thus, if communication problems prevented balancing at all, it was not because of physical hurdles associated with the lack of modern technology, since all of Rome's near and far neighbors seemed to be aware of its hegemonic rise. Rather, Rome's potential balancers frequently misperceived the danger posed by Rome and the extent of its power, most likely because of Rome's subtle deceptive strategies and tactics. But even if potential balancers had not been deceived by Rome, it is unlikely they would have balanced effectively, because their constant power struggles amongst each other prevented them from cooperating to stop the bigger threat, Rome. Most often this in-fighting precluded alliances altogether (IV2); but even in the few instances where the potential balancers managed to act in concert against Rome, the distrust resulting from their frequent conflicts mostly destroyed all common efforts before they could reach fruition (IV3).

³⁹⁵ Ibid., XIV.2 and 9, XV.9.

2. Collective Inaction (IV2)

One of the prime facilitators of Roman hegemony was its opponents' inability to cooperate to stop its rise. While some of Rome's strongest competitors—Carthage, Macedon, and Syria—attempted to balance on their own through internal means, very few tried to gather balancing coalitions and act together against Rome, even though combined efforts would have likely constituted the best chance of stopping Rome given the large number of small entities and thus the fragmentation of power in the region. Except for the three great powers mentioned above, and perhaps Ptolemaic Egypt, the Mediterranean consisted exclusively of a very large number of minor powers for whom external balancing through alliance constituted the only way to act against a rising hegemon. As Rome's power grew throughout the Republic, even the regional great powers would have needed to resort to collaboration to successfully stop Rome. Yet, though knowledge of Rome's growing power was widespread, there were hardly any alliances against Rome. This collective inaction was due partly to Rome's clever uses of communication, which indirectly hindered alliance efforts (IV2.1), but mostly due to the potential balancers' acute trust issues and preference for immediate gain (IV2.2). There is no evidence of insufficient interest (IV2.3) or buckpassing (IV2.4).

a. Communication Problems (IV2.1)

Communication problems may work directly by dampening *individual* incentives to balance, as argued in the previous section, but they can also help the rising hegemon indirectly by hurting the *collective* balancing efforts. Physical communication hurdles, as we have seen, were not pervasive enough to deprive Rome's neighbors of information about the rise of Rome, which circulated via a number of channels. But primitive diplomatic practices may have

prevented them from organizing a collective response, even though they might have wanted to. During Rome's early rise in the 3rd and 2nd century B.C., there was a clear barrier to collaboration because of the lack of official contact between polities, Eckstein argues, and the "prevailing primitiveness of diplomatic practices" in the Hellenistic state system. While envoys occasionally met for the mediation or arbitration of disputes, there were no regular diplomatic missions or permanent ambassadors. Those only became commonplace in the 1st century B.C., when Rome's power had already reached vast proportions. As a result, during Rome's early rise, contact between potential balancers remained intermittent. The absence of institutional channels of communication not only prevented the organization of a collaborative response to Rome, but it also facilitated the spread of conflict, misunderstanding, and mistrust between potential balancers, which itself rendered cooperation less likely. Without continuous diplomatic interaction and mechanisms of crisis diffusion and management, most contacts were only ad-hoc, with envoys sent once a crisis had started. While the potential balancers did thus not face physical hurdles gathering information about Rome's growing power, they experienced physical difficulties communicating with one another, which hampered their efforts to generate concerted balancing.³⁹⁶

In addition to those physical communication limits, Rome's use of deceptive communication also restricted potential balancers' incentives to collaborate. Roman restraint, for example, discouraged not only individual, but also collective balancing efforts by spreading a benign view of Rome. In fact, Rome devised deceptive tricks specifically with the purpose of preventing alliances between potential balancers. The Romans frequently resorted to divide-and-conquer tactics, in particular, to cause mistrust between likely allies and stop their cooperation. They did not hesitate to use such tactics even against their most powerful adversaries, Carthage

³⁹⁶ Eckstein 97.

and Syria, as Polybius attests. After his defeat against Rome in the Second Punic War, Hannibal sought refuge in Syria in 196 B.C., where he stayed with Antiochus III. Rome naturally needed to prevent any formal rapprochement between the two leaders. Since the Roman armies were already in the region battling Macedon, Rome seized the opportunity to wreak havoc in Syria, while plotting to throw the blame on Hannibal. As Polybius points out, the Roman envoys to Syria "made a point of paying attention to Hannibal in the hope of planting suspicion against him in the mind of the [Seleucid] king—which in fact was exactly what they achieved." Antiochus became so distrustful of Hannibal that Polybius even writes of an "estrangement." In this case the tactic did not prove as successful as the Romans would have liked, and Antiochus eventually hired Hannibal as a military advisor in his subsequent war against Rome (192-189 B.C.). However, the mistrust remained because there was never a formal alliance between Hannibal and Antiochus and the Seleucid king frequently failed to follow Hannibal's military advice during the war.³⁹⁷

b. Lack of Trust (IV2.2)

While communication hurdles, some physical and some deceptive, may thus have contributed to collective inaction, the main factor that prevented the cooperation of Rome's enemies was their exclusive attention to short term gain and the concurrent inability to trust each other (IV2.2). Conflict remained rampant among virtually all of Rome's potential balancers, who focused almost exclusively on extracting easily accessible advantages from their neighbors to the point of ignoring the looming long-term threat of the rising hegemon. The permanent belligerence in the Mediterranean rendered collaboration quasi-impossible, since it was expected

³⁹⁷ Polybius III.11-12.

that allies were not reliable and would turn hostile at the first opportunity. The large number of political entities just multiplied enmities. In addition, the third and second centuries B.C., key to Rome's early rise, were a period of offense dominance in strategy, tactics, and doctrine, as Eckstein points out, encouraging aggressive behavior and heightening security dilemmas that hampered cooperation even further. At the same time, the ancient Mediterranean did not have anything remotely resembling an international institution that could have monitored the enforcement of agreements and made states more comfortable with alliances. Instead, agreements were eagerly violated; there was no accountability for breaking them, and even concerns for reputation had very little relevance and did not prompt parties to observe agreements. As a result no state could trust others' compliance. The only way to achieve compliance remained force.³⁹⁸ The difficulty to cooperate was also exacerbated by the diversity in political entities and allegiances proper to antiquity. States around the Mediterranean were not rigid national units but rather composed of an array of sub-groups and tribes with their own organization and military capabilities often loosely articulated around a king. Thus, it was frequently impossible to gather unity even within a certain state, and accordingly, it was a herculean task to unite different states. Rome, of course, was only too happy to exploit such manifold and incessant hostility to ensure collective inaction.

The examples of potential balancers fighting each other rather than allying against Rome are simply unending, facilitating Rome's rise from the early days of the Republic all the way to the Principate and Rome's accession to hegemony. Already in the beginning stages of Roman consolidation within Italy, Rome's spread was much helped by the Italians' constant quarrels. The Etruscan cities frequently attacked each other, competing for territory since the 4th century. There existed a league of twelve Etruscan cities but it only had a religious role, and no political or

³⁹⁸ Eckstein 96, 100-102.

military prerogatives. Some Etruscan cities also regularly attempted to conquer the Greek city-states of Southern Italy during that period. Meanwhile, the Latin League was also tainted by member disputes and was therefore an alliance only hypothetically. Italy, both North and South of Rome, was thus no more than accumulation of war-prone city-states. They occasionally fought against Rome, but since they quarreled at least as much against each other those few attempts against Rome can hardly be considered collective balancing efforts. Moreover, there is no evidence that the various Etruscan attacks against Rome in the 4th century B.C.—by Veii, Tarquinii, Volsinii, Caere, and Felerii, for example—, nor the Samnite attacks around the same time, were coordinated efforts. Similarly, the Gauls, though terribly feared by the Romans following the 390 B.C. sack of Rome, kept primarily attacking each other and taking on Rome on their own rather than joining forces with other Roman enemies, who were often afraid of the Gauls themselves. In fact, the Gauls carried such a reputation for random violence and unpredictability that no other Roman adversary except for the Carthaginians dared to ally with them against Rome between the 4th and the 2nd centuries.³⁹⁹

The Greek world was also torn by constant warfare since the death of Alexander the Great in 322 B.C. The Greek city-states were unable to agree on anything, not even on protecting their common holy sites. Thus, temples were recurrently destroyed by another Greek army, as happened to Dodona in Epirus, Thermos in Aetolia, the Nicephorium at Pergamum, and even those sites supposedly off-limits according to official decrees of the time. The Hellenistic armies grew exponentially during that period, and the descendants of Alexander the Great worshipped victory to such an extent that war simply became endemic as the natural tool of diplomacy for all Greek states. Many reasons are advanced by historians for the decline of Greece, which became increasingly apparent in a variety of fields after 200 B.C., but the primary cause of the Greeks'

³⁹⁹ Ibid., 123-9, 136-7, 80-86.

downfall, as Starr points out, was their constant betrayal of each other. "If one studies the record of war among the Hellenistic states after 250 B.C., one cannot but feel that these states first tore themselves to pieces by their own rivalries," Starr writes. In other words, their continuous in-fighting and absence of trust allowed the Romans to take them over. "The Hellenistic world invited—almost required—conquest," Starr concludes.⁴⁰⁰

When the Romans first turned to Illyria and Western Greece after the First Punic War, the area was already a chaotic web of constantly shifting allegiances and League memberships, driven by deep mistrust. At the time the Aetolians had besieged the city of Medion, which they sought to draw into their League. As a result, the local great power Demetrius of Macedon hired Illyrian king Agron to restore order, and Agron freed the city in 231 B.C. Emboldened, Agron's wife and successor Queen Teuta sent raids onto the Greek cities of Elis and Messenia in Achea, and on the way took over the Epiran city of Phoenice. But she had to call her fleet back home because part of her Illyrian territory had seceded to the Dardanians in the meantime. The Epirots, who had at first asked for the aid of the Achean and Aetolian Leagues against the Illyrians, then turn against the Leagues and signed a pact with their Illyrian aggressor, a perfect example of the unreliability and unpredictability of commitments. The Romans eventually intervened in the midst of this in 229 B.C. in retaliation against Illyrian piracy, but the area was so rife with internal conflict and animosity that it would have been impossible for the Greeks to build a solid coalition against the Romans even if they had tried.⁴⁰¹ Indeed, the Greeks again engaged in massive in-fighting a decade later during a conflict paradoxically called War of the Allies, in which basically all the second-rank powers in Greece undermined each other while Rome watched on the sidelines. This time the conflict began with the city of Cynaetha in Arcadia,

⁴⁰⁰ Starr 41-42.

⁴⁰¹ Polybius II.2-7

governed by pro-Acheans but with a strong Aetolian minority which helped stage an Aetolian invasion of the city and surrounding towns in 220 B.C., initiating a three-year long conflagration of Greece.⁴⁰²

Repeatedly, balancers against Rome took on the rising hegemon on their own, without any coordination with other Roman enemies engaged in simultaneous conflicts with Rome. As a result, Rome could simply deal with each individually, one after the other, instead of all at once, which would have been much more challenging since until the Marian reforms of 107 B.C., the Roman army only held a maximum of two legions on the ground at any one point. For example, when a few years after the defeat of Queen Teuta the Romans found themselves provoked by the Gauls, the Carthaginians, and Teuta's successor Demetrius of Pharos at the same time, they encountered no difficulty in stalling one conflict while taking care of the others. They first defeated the Gauls (224-223 B.C.), the threat closest to their homeland, then tackled Demetrius (220-219 B.C.), and only after that took on the most powerful adversary, Carthage, in Spain (Second Punic War, beg. 218 B.C.).⁴⁰³ Had the Gauls, the Illyrians, and the Carthaginians put concerted pressure on Rome, the Romans would have risked attack from three geographically opposite theatres, which their two-legion and two-consul army would not have been able to face.

In addition to balancing without coordination with other balancers, Rome's competitors often acted erratically, reinforcing the mistrust of potential allies instead of rallying them. Philip V of Macedon offers a prime example of such behavior. As one of the strongest contenders against Rome and the heir to Alexander the Great's Hellenistic empire, Philip was the most likely candidate to unite the Greek world in a coalition against Rome. Yet during the successive wars between Macedon and Rome, Macedon was largely on its own, with most of the Greek states

⁴⁰² Ibid., IV.17-21.

⁴⁰³ Ibid., III.16

opposing rather than supporting it. Philip failed to secure Greek allies against Rome because both before and during the wars, he aggravated most Greek states by seizing territory here and there at their expense. When Philip called up a conference in Demetrias in November 198, in the midst of the Second Macedonian War, to negotiate as he faced setbacks against Rome, the resentment of the Greeks toward him was very apparent. Delegates from every corner of Greece voiced vindication against him: Pergamum and Rhodes for his incursions; the Acheans for his conquest of Corinth and Argos; the Aetolians for his annexation of some of their cities. By his erratic conduct throughout Greece, Philip aroused the mistrust of the Greeks and undermined his ability to garner their collaboration in his efforts against Rome, as Roman consul Flaminius himself pointed out to him at the Demetrias conference. "Of course you are alone by this time, Philip: you have killed off all the friends who could give you the best advice," Flaminius reportedly mocked. Naturally the Romans volunteered to become the spokesmen and defenders of the Greek cities and states violated by Philip, and encouraged them to send delegates to the Roman Senate to appeal against Philip who "cheat[ed] the Greeks out of their hopes of liberty."⁴⁰⁴

On top of acting erratically within their spheres of influence, the great powers of the Mediterranean also jumped on every opportunity to take advantage of each other, even though this only weakened their position and reinforced Rome's relative power. When the Ptolemaic Dynasty collapsed in Egypt in 207 B.C. due to domestic turbulences, the two other Hellenistic great powers, Macedon and Syria, immediately turned against their weakened neighbor and seized some of its border territory. In an act of "unbounded rapacity," as Polybius depicts it, Philip and Antiochus entered a pact to partition and seize Ptolemaic possessions. As a result the leading Greek states, Rhodes, Pergamum, and Athens, turned against the two great powers and

⁴⁰⁴ Ibid., XVIII.1-7 and 11.

even appealed to Rome to join in against them. Macedon and Syria's predatory actions, focused solely on their immediate gains, were thus doubly counterproductive: they detracted potential allies toward their most powerful enemy, Rome, and drove that enemy into the conflict. Indeed, the great powers failed to see that they should have buttressed rather than dispossessed Egypt and collaborated to protect themselves against Rome, who would soon insert itself in the power vacuum created by the decline of the Ptolemies and conquer them one by one. As Eckstein concludes, "a key to what occurred ... was that the geopolitical horizons of Philip and Antiochus turned out to be too narrow."⁴⁰⁵

In the far Eastern Mediterranean, finally, lack of trust and focus on short-term gains similarly forestalled collective balancing and facilitated Rome's hegemonic rise. The Sea of Marmara and the Bosphorus Strait, which connect the Aegean Sea and the Black Sea (Pontus to the Romans), constituted the sole maritime trade route between Greece and the vast Black Sea region and was thus a pillar of economic power in the Eastern Mediterranean. However, the Bosphorus Strait itself belonged to the Greek city of Byzantium (later Constantinople and now Istanbul), while the very fertile land above it and along the Sea of Marmara belonged to Thrace, itself controlled by a number of chieftains constantly at odds with each other and unable to agree on any common venture. Byzantium and the Thracian chieftains remained locked in a "perpetual and insoluble state of war" while Rome was growing throughout the Eastern Mediterranean.⁴⁰⁶ Had they teamed up instead, they would have combined control of the Greek trade routes with a high agricultural output, providing a very powerful counterweight to Rome in its attempt to control Greece and Asia Minor. Instead, because of their short-sighted rivalry, Byzantium and Thrace remained weak and the Romans easily incorporated them into the empire.

⁴⁰⁵ Eckstein 106-109, 111, 115; Polybius XV.1-26.

⁴⁰⁶ Ibid., IV.45.

Further East beyond the Black Sea the collective inaction problem was similar. Mesopotamia was home to the Parthian Empire, which remained, along with the Germanic tribes, Rome's biggest external challenge under the Principate. The Romans never fully managed to pacify the Parthians, but they never posed a severe threat because they were mostly divided and fighting amongst themselves, just like the Thracians. The Parthians were nomadic tribes living in the desert and arid mountainous region of Mesopotamia, led by warring feudal warlords who were technically ruled by an Arsacid king, though in reality the king proved unable to unite them. For example, British Roman historian Graham Webster emphasizes, "it was never possible to maintain a permanent army ... under the Arsacids." Each warlord maintained his own army, over which the king a virtually no authority. The Parthians had tremendous potential as balancers of Rome, in particular because they were highly skillful horsemen and archers. Yet because of their constant in-fighting they were only able to occasionally harm Rome with isolated squirmishes and raids.⁴⁰⁷ Had they overcome their own mistrust and belligerence and formed a common front against Rome, their insurgency-style desert warfare may have done considerable damage to Rome's regular army of heavy-armored legionaries.

Thus, trust issues and attention to short-term gains were the primary cause of the widespread collective inaction that paved the way for Rome's smooth hegemonic rise, helped by some communication hurdles. But even in the instances where Rome's balancers managed to cooperate, as the next section demonstrates, the same issues soon doomed their collective efforts.

⁴⁰⁷ Graham Webster, *The Roman Imperial Army of the First and Second Centuries A.D.* (University of Oklahoma Norman Press, 1998), 29.

3. Laggard Balancing (IV3)

Though collective inaction was extensive during the rise of Rome, some attempts at collaborating against Rome occurred. Far from being successful, those efforts constituted laggard balancing, being too late or too weak to make any difference, mainly because they were plagued by the same trust issues and attraction to short-term gains endemic to the region that prevented most alliances in the first place. Attempts at external balancing thus failed in each phase of Roman rise: efforts by the Italians, the Celts, then the Carthaginians, and later the Greeks, the Macedonians and the Syrians, and eventually the Parthians and the Germans to unite against Rome all collapsed as individual states and tribes sought to take advantage of each other despite the growing threat of Rome. While the communication shortfalls and deceptive divisive tactics of Rome mentioned above naturally enhanced the balancers' innate suspicion of each other in their collaborative efforts (IV3.1), the laggard nature of their balancing was ultimately a result of their own refusal to transcend their quest for immediate gain (IV3.2). Again, I found no lack of sufficient interest (IV3.3) and no evidence of buckpassing (IV3.4).

In the early period of Rome's rise already, the few cooperative efforts organized by various Italian tribes remained weak because they were riddled with internal competition. Tribes in several Italian regions formed Leagues in the 4th century, like the Etruscans and the Latins, but most remained a collection of individual entities and leaders that were too distrustful of each other to put their resources in common and intervene as a collective when facing a crisis. When confronted with Rome, thus, these Leagues were soon overcome. The Samnite League, an alliance of the tribes in the Southern Apennines, was relatively more cemented than other Italian Leagues. Cooperation was restricted to war issues, but in those cases the Samnite League was led by one supreme general. Unity of command rendered the Samnite League militarily much more

effective than the other Leagues. And indeed, Rome clashed with the Samnites twice (340s and 327-304 B.C.) over the control of Campania before the Romans were able to make incursions into Samnite territory. The two wars were close victories for Rome, demonstrating that collaboration pays off for balancers.

The Samnites must have realized that because they subsequently built a coalition with some Etruscans, Umbrians, and Senones Gauls in 302 B.C. But the alliance was not united enough on the field; at the ensuing Battle of Sentinum (295 B.C.), each ally attacked the Romans on its own, and the coalition was crushed despite its larger numbers. As Eckstein points out, the allies were "individually ... formidable" military opponents for Rome, but by fighting on their own they doomed their chances. In fact, the coalition crumbled after that defeat, and Rome was then able to overcome them not as a collective, but one by one. The allies had kept their autonomy in the battlefield because each sought to watch their backs against any treachery from the others. The Etruscans, in particular, were wary of the Celts, who had expanded their own territory at the expense of some Etruscan city-states in the Po Valley in 350 B.C. and had invaded other parts of Etruria as recently as 299 B.C., despite the existence of the pact against Rome.⁴⁰⁸

Further South in Italy, the various Greek city-states between Naples and Sicily also formed a coalition of their own, the Italiote League, as early as 420 B.C., to shield themselves against the constant attack by indigenous tribes and neighbors, and later against Rome. As with other Leagues of Italy, the Italiote League lacked common purpose, this time not because of the lack of participation of its members but because of the dominance of some members. Because it was forcefully led by one member, first Croton and then Tarentum, and headquartered in that member's territory, the League was viewed suspiciously by other members as an instrument of

⁴⁰⁸ Eckstein 129-131, 140-2, 144-7.

the leader to curb fellow members' power rather than to deal with external threats. As a result, the League was not a reliable common instrument and membership suffered from a high turnover rate. In addition, local Italians resented the Greek settlers and recurrently attacked the Italiote League in the late 4th century rather than recruit it to help them fend off Rome. Isolated, Tarentum resorted to call on a new ally, the city-state of Epirus in Greece, to battle Rome. That alliance raised the hurdles for Rome, and Epirus and Tarentum, also aided by other Southern Italian Greek cities, first defeated the Romans at Heraclia in 280 B.C. before Rome was able to inflict considerable losses on the Greek forces and take control of the colonies eight years later. Rome won in the end because Pyrrhus, the king of Epirus, became distracted by his own ambitions for power in Southern Italy, which led him to overreach by trying to simultaneously dislodge the Carthaginians from Sicily. Once again, a collective effort was disrupted by an obsession with immediate gain.⁴⁰⁹

In Northern Italy, common efforts by the various Gallic tribes similarly failed because of their own local belligerence and resulting mistrust, which was recurrent throughout centuries of Gallic history. Gallic alliances were ephemeral and constantly shifting. The sack of Rome itself could have permanently incapacitated the Romans, had the Senones tribes received support from other Celtic tribes and formally occupied the region around the city. Instead, the Senones were forced to leave Rome almost immediately because their own territory in Gaul, around the present-day Eastern Parisian region, was threatened by a Gallic tribe from Armorica (Brittany), the Veneti. The Senones were further distracted by local conflicts as "they became involved in domestic wars" with several Alpine tribes, and subsequent attempts to stop the progress of the Romans (356 and 344 B.C.) failed because "dissensions broke out within their own rank," leading the Senones to withdraw again, Polybius tells us. In 299 B.C., the Senones made a

⁴⁰⁹ Ibid., 148-50, 152-8.

successful incursion into Etruria, but "no sooner had they arrived home that they began to quarrel about obtaining a larger share of the spoils, and in the end destroyed the greater part of their own army and even the plunder itself." Such a self-destructive focus on immediate gains was not the sole characteristic of the Senones, Polybius emphasizes; instead, in-fighting was a "common occurrence among the Gauls," who were unable to mount any successful common enterprise against the Romans between 295 and 241 B.C.⁴¹⁰

In the early 240s B.C. the Senones sought to organize a coalition of Celtic tribes, but other tribes like the Transalpine Boii, for example, hesitated to trust them. The Boii sided repeatedly with the Etruscans, to the delight of the Romans, who had gotten ready for a Celtic invasion by then and had started moving troops North to meet the Gauls. "But when they [the Romans] learned of this act of self-destruction by the enemy they returned home." It is not until the Romans began colonizing the land taken from the tribes in Northern Italy that Gallic collaborative efforts bore fruit. The Senones, Boii, and Insubres allied and even obtained support from other tribes from the Alps and Rhône regions. However, the lure of short-term gains soon returned as the Romans were easily able to convince some tribes—the Veneti and the Cenomani from Cisalpine Gaul—to defect and attack other Celts, and the allies became so distrustful that they dispatched troops to guard each other. They nevertheless managed to conquer Etruria with little resistance, and went as far as three days away from Rome. But their lack of unity and fear of each other made them no match for the Romans, whose reinforcements decisively crushed the Gallic allied armies at Telamon in Tuscany in 225 B.C., enabling Rome to then take control of all of Northern Italy and the fertile Po Valley between 220 and 180 B.C.⁴¹¹

⁴¹⁰ Polybius II.18-25.

⁴¹¹ Ibid., II.25-33.

When Rome eventually sought to invade the Gallic homeland over a century later, it was again considerably aided by the mistrust and in-fighting of the Celtic tribes. While Vercingetorix, the leader of the Averni, famously united many tribes against Julius Cesar in the 50s B.C. and led a common army with clever tactics, he was unable to transcend Gallic factionalism and hostility entirely as some tribes declined to join him and sided with the Romans instead. Vercingetorix and the allied Gallic armies were eventually destroyed at Alesia in 52 B.C., sealing the conquest of Gaul and ending Gallic balancing efforts against Rome.⁴¹²

In the second phase of Rome's rise, Carthage offered the most substantial opposition to the aspiring hegemon, yet its collaborative efforts failed as well. During the First Punic War, Carthage allied with Hiero II of Syracuse against Rome, but that alliance was wrought with difficulties. First, it was only an ad-hoc alliance, the result of circumstances preceding the war, and not planned or organized in any way. Second, the Carthaginians, who controlled Western Sicily, and the Syracusans, from Eastern Sicily, fought battles against Roman forces separately and were thus only political, not military, allies. Their reluctance to collaborate was most likely due to the fact that Carthage and Syracuse had been bitter enemies for nearly two centuries, battling for the control of Sicily. One of Hiero's ancestors even attempted to invade Punic North Africa in 310-308 B.C. Thus, during the First Punic War, the "two allies ... greatly distrusted each other," as Eckstein underlines it. Indeed, after Rome pushed back the Syracusan forces early in the war and Syracuse faced the risk of defeat, Hiero disengaged, preferring to cut his immediate losses rather than help Carthage continue the fight. The coolness between the allies did not escape the Romans. Consul Claudius Appius, seeing his opponents divided and not supporting each other, immediately attacked the Carthaginian forces and scored a quick victory,

⁴¹² Julius Cesar, *Commentariorum De Bello Gallico Liber Primus (Gallic War)*, transl. by W. A. McDevitte and W. S. Bohn (New York: Harper & Brothers, 1st Ed., 1869), Books I-IV.

then turned to Syracuse again and besieged the city. With no Carthaginian help in sight, Hiero capitulated to the Romans in 263 B.C. and was forced to support them against Carthage for the rest of the war.⁴¹³

Syracuse's fickleness and predisposition for short-term gains resurfaces fifty years later during the Second Punic War, when Hiero's successor Hieronymus decided to sign another alliance treaty with Carthage, in ca. 215 B.C., violating the alliance treaty with Rome Hiero had signed only one year earlier. Hieronymus's change of heart was not motivated by the necessity to balance the growing threat of Rome but by his desire for short-term gain. Having heard of Rome's massive defeat at Cannae in 216 B.C., he anticipated a Carthaginian victory in the war and signed the treaty with Hannibal solely to extract territorial spoils from the pre-supposed victor. The terms of the treaty provide clear evidence of his opportunistic reasoning: after Carthage expelled the Romans from the island, it was to help Syracuse take control of all of Sicily, parts of which were still owned by Greek colonies, in exchange for Hieronymus's military support against Rome. Thus, Syracuse was not a reliable ally but could switch allegiances again as events progressed on the mainland—though one could argue that this was a calculated risk for Hannibal since Sicily was only a minor theatre of events during the Second Punic War. In the end, Hieronymus was assassinated before he could provide the help promised in the treaty, so the alliance never materialized.⁴¹⁴

Carthage was not much more successful with its other allies in the Second Punic War. Hannibal used two types of allies: whole political entities like states, city-states and tribes, such as Syracuse; and individual mercenaries from various areas, some allied and others not. On the battlefield, first, Hannibal managed to draw numerous mercenaries from a wide variety of areas,

⁴¹³ Polybius I.11-12 and 16; Eckstein 160-170.

⁴¹⁴ Polybius VII.3-5.

and Polybius praises him for having been able to maintain his army together despite its disparate elements, principally Africans, Spaniards, Ligurians, Celts, Phoenicians, Italians, and Greeks, "men who had nothing naturally in common, neither in their laws, their customs, their language, nor in any other respect." Yet while Hannibal may have had some authority over his contingents, their diversity often hindered effective, united military action, which soon became problematic because as the war drew longer and the Carthaginian ranks grew thinner, Hannibal relied increasingly on foreign mercenaries, up to almost 1/3 of his totally army at the last battle of the war at Zama in North Africa in 201 B.C. The mercenaries, unlike the Roman legionaries, were often not reliable because of their different military training, the language barrier impeding the giving out of orders, and their own customs. The Gauls reportedly became drunk at the eve of Hasdrubal's last battle in Italy, for example, and missed the battle because they were asleep.⁴¹⁵

But in addition to the difficulties resulting from the diverse background of the mercenaries, Hannibal was also unsuccessful in ensuring the loyalty of his mercenaries, particularly the Spaniards and the Celts. Although the mercenaries' loyalty was harder to earn than the Carthaginians' because their stake in the conflict remained lower, Hannibal made no special attempt to obtain their trust. Quite the contrary, he recurrently treated mercenaries as expendable, placing them in the riskiest positions in the front and flanks of an attack and selecting them for quasi-suicide missions and raids to spare his own African troops, and as a result mercenaries always suffered the greatest losses. Hannibal's actions, largely motivated by calculation and distrust for the mercenaries, triggered bitterness and resentment and exacerbated the mercenaries' isolation within the army instead of fostering their loyalty and solidarity.⁴¹⁶

⁴¹⁵ Ibid., XI.19.

⁴¹⁶ Ibid., XV.11.

At the battle of Cannae, for example, Hannibal placed his troops in a reverse crescent-shaped formation with the Spaniards and Celts in the front center, which would be hit first. As Polybius points out, "the object of this arrangement was to begin the battle with the Spaniards and Celts and use the African troops as a reserve to support them." At Cannae over four times more Celts died than Africans. Similarly at Zama, Hannibal placed his mercenaries, this time mostly Ligurians, Balearians, and Celts, in front to take the hardest blow, while the native (non-Carthaginian) Libyans and Mauritians were behind them, and Hannibal's own Carthaginian troops brought back from Italy remained further behind, as a second reserve. Hannibal effectively created a social and perhaps even racial ranking system, which as the outcome demonstrates failed to enlist the trust and reliability of the mercenaries. Indeed, the mercenaries displayed great fighting skills, but when they saw that the Carthaginians were letting them take the hit by themselves while safely watching them being slaughtered from behind the battle line, "it seemed to them that they had been abandoned by their own side, and so as they retreated they turned upon the soldiers in their rear and began to cut them down." Carthage subsequently faced both the Romans and their own mercenaries, and suffered a fatal defeat. Thus, Hannibal's actions, by generating the mistrust of his allies, battered their trust and doomed their support.⁴¹⁷

Hannibal behaved similarly with the political entities Carthage managed to enlist to help it balance Rome—some Spanish tribes, Italian cities, and Celtic tribes—and was equally unable to keep these allies' support. The Spanish tribes South of the Ebro River were the first to be recruited when Carthage began expanding into the South of Spain after the First Punic War. But Hannibal and his lieutenants treated their Spanish allies with such contempt, Polybius notes, that they abandoned Carthage at the first opportunity instead of reinforcing Carthage's balancing power. Thus, by the time the Romans landed in Spain at the onset of the Second Punic War, even

⁴¹⁷ Ibid., III.113 and 117.

the "most loyal supporters of the Carthaginians" in Spain, the powerful chieftains Andobales and Mandonius, "had for a long time been disaffected and were only awaiting the right moment to revolt." The alliances were in fact hardly voluntary, and more often than not coerced by the Carthaginians through terror tactics to extract benefits from the locals. Hannibal's brother Hasdrubal, one of the Carthaginian leaders in charge of the Spanish theatre during Hannibal's invasion of Italy, demanded large payments from his Spanish allies and frequently took the chieftains' wives and daughters hostage. Roman proconsul Scipio Africanus, in charge of the war in Spain, did not fail to notice the opportunity such poor alliance management offered for the Romans. He noted to his men in 209 B.C.:

Because of their [the Carthaginians'] overbearing treatment of their allies, they have alienated them and turned them into enemies. The result is that some of them are already negotiating with us, while the rest, as soon as they can pluck up the courage and when they see that we have crossed the river [Ebro], will gladly take our side ... because ... [of] the outrages they have suffered. But most important of all is the fact that the enemy's commanders are quarrelling among themselves and are not willing to fight as a single army, while if they attack us separately it will be all the easier to defeat them.

Carthage and its Spanish allies, riddled by severe trust issues resulting from the Carthaginians' mistreatment, were thus unable to cooperate against Rome. Because of the danger of being double-crossed by their local allies, the Carthaginians divided their troops into three armies spread throughout Southern Spain, away from their home base and supplies depot at New Carthage, greatly facilitating Scipio's takeover of Spain.⁴¹⁸

The Romans quickly understood the failures of Hasdrubal. Andobales reportedly told Scipio of all the "injustices and insults" he suffered from the Carthaginians. While Scipio's men

⁴¹⁸ Ibid., X.6.

did not hesitate to use terror tactics against those openly opposing the Romans, they were particularly attentive to the appeal of the Spanish locals willing to collaborate with them, in stark contrast to the Carthaginians. Scipio liberated the Spaniards taken into slavery by Carthage and returned them to their families. He freed the wife and children of Edeto, chief of the Edetani tribe, who then befriended other local chieftains and convinced them to switch allegiance too. Scipio also took care to reward the local tribes for their military support in defeating Hasdrubal. Carthage's mistake had been to believe its position secure and exploit its allies rather than develop a partnership based on mutual trust.⁴¹⁹

Unlike the Romans, though, the Carthaginians failed to learn from that mistake and repeated it with other allies. On his way to Italy and through the Alps in 218 B.C., Hannibal successfully recruited a number of Alpine Gallic tribes—chief amongst them the Boii and the Insubres—in addition to mercenaries, to support Carthage in the invasion of Italy. Those alliances remained somewhat shaky because they were mostly based on Hannibal's promise to reward them with luxurious gifts. Even though many of the tribes were at odds with Rome because of Rome's establishment of colonies on their territory, their collaboration with Carthage was more the result of a desire for immediate gain than a quest to balance Rome, and thus, they did not constitute reliable allies for Carthage. In fact, while some tribes helped guide the Carthaginians through the mountains, provided them with supplies, and even participated to fighting the Romans, not all Celtic tribes were equally cooperative. Hannibal's men frequently had to pay bribes to pass through Gallic territory, or buy goods and supplies from the tribes, and were attacked by the Rhône River and repeatedly ambushed while crossing the Alps, costing them much time and numerous lives. After passing the Alps the Carthaginians repeated the

⁴¹⁹ Ibid., X.15 and 34-36.

brutality that had distinguished them in Spain, this time against the Taurini tribe who refused their offer of alliance.⁴²⁰

"The slaughter of the Taurini had an immediate effect on the neighboring tribes," ancient historian John Prevas explains. "Many of the Gauls were intimidated by how quickly and brutally Hannibal had taken the city," he continues, and while some joined Hannibal out of fear, as the Carthaginian leader had planned, many more were alarmed for their own safety and became swayed by Rome's efforts to draw them away from Carthage. "He [Hannibal] discovered that some of the Celtic tribes who lived between the valleys of the Trebbia and the Po had made a treaty of friendship with him, but were at the same time negotiating with the Romans," Polybius mentions. In fact, during the winter 217-216 B.C., Hannibal was reportedly so worried about the unreliability of the Celts that he feared for his own life and had several wigs made, which he alternatively wore with matching outfits to look like different men of different ages so the Celts would not recognize him.⁴²¹

Even though he had lost many of his Gallic allies by pursuing terror tactics, Hannibal still had not learned from his mistake by the time he reached the Italian mainland. During his campaign in Italy, Hannibal came very close to annihilating the rising hegemon, winning massive victories at the Trebia River, Lake Trasimene, and Cannae. Yet in the end he was unable to fully defeat Rome for one reason, Starr argues: "he was born and bred in an empire which rested more on force than on consent" and he thought he could coerce Rome's Italian friends and allies into supporting him by terror tactics and displays of force. Hannibal lost the Second Punic War for the same reason Hasdrubal lost Spain: because he was unable to sway populations and generate allies, and instead inspired only fear and mistrust. Even after Cannae, with Rome's best

⁴²⁰ Eckstein 264; Polybius III.34 and 40-44.

⁴²¹ John Prevas, *Hannibal Crosses the Alps: the Invasion of Italy and the Punic Wars* (Cambridge: Perseus/Da Capo Press, 1998), 178; Polybius III.50-56, 60, 66-71, 78.

armies annihilated and its leadership in complete disarray, only a few city-states like Capua joined Carthage against Rome. While advancing through Italy, Hannibal collected a massive amount of booty, numerous slaves, and repeatedly seized property, making himself very unpopular. He mistakenly believed that displaying ruthlessness would lead the Italian tribes and city-states to join him out of fear. Such tactics worked when employed by the Mongols because they only sought surrender, not voluntary alliance, and their targets had no viable alternative. But Hannibal failed to convince the locals in Italy because his erratic behavior caused him to be ultimately regarded as the largest, most immediate threat, and the enemy to beat instead of Rome, which had by contrast granted them freedom and autonomy.⁴²²

Even Hannibal's mild treatment of his non-Roman Italian prisoners, whom he generally sent back to their home without the usual ransom demand in the hopes that they would convince their towns to support him, failed to detract the Italians' attention away from his brutality and treachery. Polybius recounts that as soon as they had passed the Alps and reached Etruria, "the Carthaginians began to ravage the countryside and columns of smoke rising on all sides bore witness to the devastation." As they continued South toward Rome and along the Adriatic, Hannibal "amassed so much plunder that his army could neither drive it nor carry it with them, and he also killed a number of the inhabitants on his route." That is an understatement; in reality, the entire adult population of some cities was killed—hardly the most efficient way to gain the trust of the locals and coax them into joining forces.⁴²³

The cooperation attempts among the Greeks during the third phase of Rome's rise tells again the same familiar story of mistrust and focus on short-term gain at the expense of other potential balancers. Greece, like Italy, comprised a multitude of city-states, and as a result, also

⁴²² Starr 44; Polybius III.17 and 77.

⁴²³ Ibid., III.82, 86, 88, 90, 100, 107.

organized in Leagues from early on. The Peloponnesian Leagues (Aetolian League, Achean League, and Leagues around Sparta, Elis, Corinth, and Thessaly) served as collective security organizations to help defend their small members in case of aggression, intervene abroad jointly for added power, and provide mediation and arbitration in international disputes. Collaboration was also fostered during the 3rd century B.C. by a bolstering of the Pan-Hellenic identity, as evidenced by the organization of a number of joint religious festivals like the Artemis festival of 208 B.C., and collective rebuilding efforts after wars and natural disasters, like the rebuilding of Cytinium in 206 B.C. after an earthquake. But just like in Italy, the Leagues in Greece were far from close-knit communities that transcended their differences in the face of danger. They only engaged in loose cooperation, with members frequently quarrelling and attacking each other, so that members always remained suspicious of each other and membership in the Leagues was constantly changeable. In fact, once Rome started being active in the Greek Mediterranean, it soon took over the mediation and arbitration function of the Leagues, with much greater success.⁴²⁴

In-fighting was so prevalent the Leagues were virtually paralyzed in the face of Rome's spread through the region. Polybius describes, for example, that after the Achean League was established, its twelve member-cities "became so ill-disposed and even hostile to one-another that they all broke away from the League and began openly to act against one another's interest." Eventually by the 280s B.C. "the cities underwent a change of heart," he conceded, and reformed the Achean League with more solid institutional roots, with representatives from each city leading it on a rotational basis, but the damage was done and the suspicions remained.⁴²⁵

⁴²⁴ Eckstein 80.

⁴²⁵ Polybius II.41.

Not only were the member of the various Leagues quarrelling, but the Leagues themselves remained on hostile terms and repeatedly sought to take advantage of each other, rendering a united Greek front against Rome or any other intruder highly unlikely. The Cleomenian War (227-222 B.C.) provides a perfect example of their shortcoming. In 239 B.C., before Rome's arrival in Greece, the Aetolian and Achean Leagues managed to seal an alliance with the goal of expelling Macedon, then the most powerful contender for the control of Greece, from Peloponnesian lands. But as the Achean League grew with new members, the Aetolian League became wary of its competition. As a result, instead of balancing the larger threat of Macedon, the Aetolians bandwagoned with Macedon and with Cleomenes of Sparta, who had also been seeking to seize Achean territory. With Cleomenes marching through Achea, the Achean League also called to their enemy Macedon for help against the Aetolians. Itself worried of Sparta's sudden growth, Macedon eventually entered the war and reclaimed all the Peloponnesian territory taken by Cleomenes. After defeating Sparta, however, Macedon had to quickly call its army back home because its Western neighbors, the Illyrians, had simultaneously begun invading its territory. Yet the defeat of Cleomenes failed to put an end to the Greek's internal strife. Any opportunity to take advantage of one's neighbors and score short-term gains at their expense was seized. Immediately following Macedon's retreat, the Greeks went to battle again in the War of the Allies (220-217 B.C.), this time pitting the Aetolians, Sparta and Elis against the Achaeans and Thessalians. This was the dire state of affairs just as Rome was about to enter the Greek theatre. Given the constantly shifting allegiances of the Greek Leagues and their continuous in-fighting, it is hardly remarkable that Rome was able to pitch itself as the savior and liberator of the Greeks, aptly disguising its own growth.⁴²⁶

⁴²⁶ Ibid., II.44-70.

Prompting Rome's involvement in Greece was Macedon's attempt to gather alliances. Following the War of the Allies, the Aetolians, Achaeans, and Macedon signed the Peace of Naupactus (217 B.C.), which stipulated their common response in case of attack by Rome or Carthage. It looked as if Philip V of Macedon was becoming the new Alexander, ready to unite the Greek world. But just like Hannibal, Philip V of Macedon destroyed his chances to lead an alliance of the Greeks by engaging in erratic behavior and taking advantage of his own allies, as Polybius points out. Philip surrounded himself with the abrasive Demetrius of Pharos, who was only intent on securing Philip's support to gain territory for himself at the expense of other Greeks, instead of following the advice of other, more prudent allies like Achaean leader Aratus. Philip sought to expand his own possessions and became aggressive toward his own allies, even "inflicting [grave] suffering upon some," like Messene, which Macedon destroyed in ca. 215 B.C. Thus, because of the lure of immediate gain, "he lost both the goodwill of his allies and the confidence of the rest of the Greeks," making it very easy for Rome to find a sympathetic audience in Greece.⁴²⁷

Besides the failed alliance with the Greeks, Macedon simultaneously sought to join forces with Carthage, after hearing of Hannibal's victorious advance through Italy in 218-216 B.C. Philip and Hannibal concluded a formal alliance treaty in 215 B.C. The treaty specifically mentioned the common cause of fighting Rome, and it resulted in opening a third front for Rome, who was drawn across the Adriatic to fight the First Macedonian War (214-205 B.C.) while also battling Carthage in Italy and Spain. Moreover, Philip nurtured the idea of eventually joining Hannibal on the Italy mainland and invade Italy too. "This is the moment to strike a blow, when the Romans gave suffered a disastrous defeat," Philip reportedly told Demetrius of Pharos upon hearing of Hannibal's victory at Lake Trasimene in 217 B.C. In the end, however,

⁴²⁷ Ibid., V.105, VII.11-14.

the two allies remained distrustful of each other and kept their efforts separate and in their respective theatres. Philip never crossed into Italy to support Hannibal, and Hannibal did not provide any aid to Philip's efforts in Greece. Thus, the impact of the alliance and of the diversion it created remained dubious at best.⁴²⁸

Just like Hannibal and Philip, Antiochus IV, the last remaining great power in the region in the 2nd century B.C., also acted erratically and managed to alienate most of the Greeks. After the fall of Philip of Macedon in 196 B.C. that in effect turned the Mediterranean into a bipolar world, he raided Macedon's former possessions, crossed the Hellenpont into Greece, moved into Thrace, and finally invaded central Greece in 192 B.C. Antiochus not only lost the support of the Greeks, but just like Philip and his aggressive behavior before, precipitated them toward Rome, the larger though more restrained power. The Aetolians were the only ones who initially sided with Antiochus, but they were exclusively motivated by profit and short-term gain at the expense of the other Greeks. Feeling cheated by the limited territorial expansion granted to them by Rome in 196 B.C. in exchange for their help against Macedon, the Aetolians virtually invited the Seleucids into Greece, hoping they would be more generous than the Romans. As Antiochus progressed through Thessaly and reached as far as the Adriatic, he was able to convince more Greeks dissatisfied with the Roman settlement to become his allies, but he was only putting together a coalition of the greedy. When the Romans declared war on Antiochus in 192 B.C., the Seleucid king's allies only marginally contributed to the fight. When the Romans and Syrians finally clashed at Thermopylae in Thessaly (191 B.C.), a battle lost by Antiochus, Appian even reports that the Aetolians "fled ... in disorder" from the battlefield. Here again, balancing was

⁴²⁸ Eckstein 266-7, 271; Polybius V.101, VII.9.

laggard because the balancers were more intent on obtaining spoils for themselves than on stopping the potential hegemon.⁴²⁹

Finally, in the last phase of Roman rise, Rome's few remaining foreign enemies—the Germans to the North and Parthians (Persians) to the East—were similarly hindered by their infighting and the primacy of immediate gains in their calculations. As Goldsworthy points out, "the effectiveness of the Parthian army has been greatly overrated," mostly because it inflicted a resounding defeat upon Roman general Crassus at Carrhae in 53 B.C. during the Triumvirate. But Goldsworthy contends that Crassus lost the battle because he was overconfident and lacked military skills, not because Parthia was particularly brilliant. In reality, Parthia suffered from severe internal divisions and was plagued by recurrent civil wars that hampered its unity. Though the Arsacid king technically acted as the Parthian army's commander-in-chief, in reality each local family controlled sections of the army, and with no loyalty to each other or to the king, military leadership was rarely coordinated between sections.⁴³⁰

Germany beyond the Rhine and the Danube resembled Parthia in many ways. When a few decades later, Emperor Augustus attempted to conquer the Germanic tribes' territory, he concluded that their land could not be incorporated into the empire because it comprised a collection of nomadic tribes who never permanently settled but mostly raided their neighbors, so they presented too much disunity to organize them as a province with a local administration, taxation system and laws. The Germanic tribes occasionally collaborated and in a few instances even met considerable success. In 9 A.D., for example, Roman-educated Germanic leader Arminius of the Cherusci tribe sealed a secret alliance with other tribes that had traditionally been hostile, the Marsi, Chatti, Bructeri, Chauci, and Sicambri, and led them to ambush and

⁴²⁹ Appian, Book XI.IV.17-20; Eckstein 295-306.

⁴³⁰ Adrian K. Goldsworthy, *The Roman Army at War 100 B.C.-A.D. 200* (Oxford: Clarendon Press, 1996), 61-63-67.

decimate three Roman legions in the Teutoburg Forest in Lower Saxony. Yet while this battle marked the halting point of Rome's northward expansion, such cooperation remained rare, and most often the Germanic tribes focused on their own local squabbles and rivalries and would have likely proved unable to halt a massive Roman annexation campaign.⁴³¹

Thus, from Rome's early days to its accession to hegemony, collaborative attempts to stop Rome's growth were largely laggard because they were plagued by mistrust and hampered by their members' prioritization of immediate gain, and they ultimately failed to cripple Rome. Yet collective inaction and laggard balancing were not the only missed opportunities for potential balancers to stop the growth of Rome. In fact, a significant number of potential balancers even directly contributed to Rome's rise. As the previous section already hinted at, bandwagoning was rampant in the ancient Mediterranean. Many entities that could have participated in balancing the aspiring hegemon instead joined its ranks, further dooming the few balancing efforts undertaken.

4. Bandwagoning [IV4]

An unusually large number of potential balancers against Rome systematically bandwagoned with Rome instead of balancing. One must wonder why bandwagoning, though a counterproductive strategy to stop a hegemon, was so particularly prevalent in the ancient Mediterranean, while it was much rarer during the Mongol rise, for example. A first explanation is that many smaller states simply bandwagoned with Rome out of fear, not only of Rome itself but also of the other larger powers of the region [IV4.1]. Eckstein argues that the Mediterranean theatre was especially prone to fear-based bandwagoning because it comprised an unusually

⁴³¹ Publius Tacitus, *Anales (Annals, transl. by A.J. Woodman)* (Indianapolis: Hackett Publishing Company, Inc., 2004), I.3, 10, 43, 55-71, II.7, 41, 45; Webster 33.

large concentration of small states and city-states, which naturally sought the protection of larger entities to shield themselves from the fierce competition stemming from their large number and the constant risk of being attacked by larger neighbors.⁴³² In other words, the smaller states did what they must, as Waltz puts it, to ensure their survival in the short term, including allying with Rome when another great power was threatening to conquer them.⁴³³ While great powers in the region—Carthage, Macedon, Syria—may have been able to balance, the smaller states could not afford that luxury if they wanted to survive. This natural inclination was reinforced both by Rome's perfidiously restrained behavior, which, as we have seen, portrayed the rising hegemon as more benevolent than it really was, and by the other great powers' inherent tendency to take advantage of their superior power to mistreat their neighbors and allies.

As a result, smaller Mediterranean powers recurrently called to Rome for help against their neighbors, virtually inviting Rome's spread to their area and in effect enabling Rome's "invisible" rise. The examples are too numerous to list. From early on, Rome encouraged the Italians to bandwagon against the Celts, for example, by a clever propaganda stressing that they represented civilization against the threat of the barbarians. As Eckstein reminds us, "... the Etruscans and other Italian polities acquiesced in Roman hegemony not only because they feared Rome (they did) but also because they feared Celtic invasions, which they themselves were unable to prevent ... ; Rome provided what protection was available." Rome repeated the same scheme with the Aegean Greeks, and later again against the Germanic tribes and Parthia. In the Greek world, both Rhodes and Pergamum allied with Rome in 200 B.C., and Athens joined them soon thereafter, because of the predatory actions of Philip V of Macedon and Antiochus III of Syria that threatened their lands following the downfall of the Ptolemaic regime in Egypt in 207

⁴³² Eckstein 119.

⁴³³ Waltz, *Theory of International Politics*, 127.

B.C. The Ptolemies also flocked to Rome when they realized Philip and Antiochus planned to take advantage of Egypt's weakness to help themselves to Egyptian territory, and Rome succeeded in deterring an attack on Egypt, at least for a while. The Aetolians, Achaeans, and Illyrians also time and again appealed to Rome for protection against Macedon. But, as Eckstein concludes, such calls for Rome's help were based on a fundamental misunderstanding of the immediacy of local threats v. the Roman threat and in the end carried the "grave consequence" of enabling Rome's accession to hegemony.⁴³⁴

A second explanation for the prevalence of bandwagoning despite the risks it entailed is that many Mediterranean states believed they could extract profits from the rising hegemon in exchange for their collaboration [IV4.2]. Again, there are too many examples to list all. In one instance, Numidian king Syphax secretly became friendly with the Romans and negotiated with Scipio around 204 B.C. with the hopes that the Romans would rid North Africa of Carthage and open the way for his own supremacy there. In the past Syphax had supported Carthage and had even married Hasdrubal's daughter to seal their alliance, but he then realized he could obtain more leverage in North Africa if Rome defeated Carthage and thus approached Scipio to switch allegiances. As always, bandwagoning, particularly for profit, is dangerous, and Syphax paid a hefty price. When Scipio launched his attack on Carthage a few months later, he simultaneously attacked Syphax and defeated him too as the Romans did not need the Numidians. Another example of profit-driven bandwagoning is that of the Mamertines, a Sicilian colony of former Italian mercenaries responsible for sparking the First Punic War. After being defeated by Syracuse in ca. 274 B.C. and seeing themselves considerably weakened, the Mamertines appealed for help from both Rome and Carthage to maximize their chances. The Romans responded first, and the Mamertines anticipated that they would be able to accrue great gains

⁴³⁴ Eckstein 108-9, 137-8, 141-3.

from the war spoils. Again, bandwagoning did not bring them their expected benefits. Ironically, the ensuing war took on such great proportions that Mamertines became lost to history and there remains no historic trace of them.⁴³⁵

It is still puzzling, though, that so many states in the Mediterranean became tempted with both fear- and profit-driven bandwagoning even though it was mostly not paying off. The unusual prevalence of bandwagoning behavior in this case compared to the Mongol case leads us to suspect that something else was at play besides the individual states' fear and profit motivation. In fact, when looking more closely, we can see that Rome played a particular role in encouraging bandwagoners. The Romans came up with an innovative way to entice possible rivals into bandwagoning with them instead of opposing them by reinforcing and combining the fear and profit incentives, the *clientelae* system [IV4.3]. Exploiting the paradox between the potential balancers' quest for immediate gain amidst the widespread local divisions on the one hand, and the states' weakness and lack of resources to acquire those immediate gains on the other, the Romans offered the ideal solution to many Mediterranean states: become our friend and support us and for a small price, we will help you achieve those immediate gains and guarantee your own territorial integrity and sovereignty at the same time. This was the essence of the friendship or client system that Rome established throughout the region, and it worked because it resolved the Mediterranean states' dilemma by assuaging their fears for their own survival while satisfying their desire for profit, a double goal most of them would not have been able to meet with their resources alone.

In practice, Rome gradually became 'friends' with a multitude of political entities around the region, starting in the first phase of its rise with its Italian neighbors and progressively expanding to an increasing ring of friends all around the Mediterranean. This was often

⁴³⁵ Polybius I.10-11, XIV.1-2.

accomplished informally, and frequently without clear process, with an alliance treaty (foedus/foedera) or simply an oral agreement. Rome established client relationships not only with states and city-states but also with tribes and non-state entities like the Germanic tribes, Luttwak explains, thus cleverly constructing a wide network of support while still relatively weak, and obtaining justification for intervening just about everywhere in the region in defense of those clients. The concept, which to us is closest to protectorate, was based on the patron-client relationship of Roman municipal life, where rewards (beneficia) were granted by patrons in exchanged for services (official) rendered by the client. The very first clients of Rome were the cities of Latium, with whom Rome sign the Foedus Cassianum (Treaty of Cassius) in 493 B.C., pledging to protect the cities against nearby hills tribes like the Volsci and Aequi who constantly raided the coast. Multiplying such agreements throughout Italy and then beyond enabled Rome to inflate its power and rise toward hegemony. As Eckstein points out, the Romans had "found a new way, via diplomacy, to mobilize the resources of Latium as a whole" in their favor. The Foedus Cassianum was an "early exercise of Roman skill in alliance management," which was then repeated throughout Italy, Greece, North Africa, and beyond. When some of the clients eventually became incorporated into Rome as provinces, the Romans simply developed new client relationships beyond the provinces. As a result, Webster concludes, both under the Republic and the Empire, Rome developed a wide web of friendly client-states that acted as buffers between the Roman provinces and its enemies, a very cost-efficient way of securing its periphery.⁴³⁶

It is obvious, then, that being Rome's client friend came at a price. Clients were promised Rome's help and in return, were expected to volunteer their support to Rome. But in reality, while Rome was never bound to assist a client, the client kings and leaders could not refuse to

⁴³⁶ Luttwak 20-23; Eckstein 247-9; Webster 29-30.

help Rome. As historian P.C. Sands explains, "a king who consented to be called the friend and ally of a vastly superior power found it difficult to refuse assistance when it was requested, whereas the superior power by reason of its strength did not lie under the same necessity of lending its assistance." As always, due to the uneven nature of the relationship, bandwagoning remains a dangerous game. "When a king, then, accepted from Rome the title of 'friend and ally', he accepted therewith a position of inferiority," Sands concludes. In the end, virtually all of the friends and clients of Rome became Roman provinces or parts thereof, ending up under direct Roman administration and loosing the independence they sought to preserve by bandwagoning in the first place.⁴³⁷

Even during their client relationship with Rome, Roman friends were never entirely independent. Becoming a client already meant forfeiting a large part of their sovereignty to Rome. The terms of the friendship treaties and agreements were generally fundamentally unequal, and very restrictive for the client. The first condition, that the authority of the client king or leader be approved and legitimated by the Roman Senate, limited the client's autonomy in domestic political matters. In practice the rule meant that Rome controlled all successions and ascensions to the throne in their client states. Every change of leadership had to be ratified by Rome, otherwise the friendship relation was not renewed. Sometimes Rome simply acknowledged the successor, but in the more strategically located states Rome actually named successors, and during the Principate the Roman emperors even developed the practice of crowning new client kings themselves. Thus, Rome largely controlled the domestic politics of its clients and could ensure that they were always ruled by pro-Roman leaders and remained loyal.

⁴³⁷ Percy Cooper Sands, *The Client Princes of the Roman Empire under the Republic* (New York: Arno Press, 1975), 46-7.

In addition, clients were required to send hostages to Rome, often the sons of the rulers, which enabled Rome to Romanize future rulers.

A second highly restrictive condition imposed upon clients was their subordination in most foreign policy matters. Clients not only had to refrain from any international behavior hostile to Rome, but Rome's consent was also required for the client to engage in any war or conclude any treaty and clients were required to accept Rome's arbitration in cases of conflict with other clients. Client leaders were allowed to take down internal revolts on their own, but if attacked by a third party could only defend themselves until the Roman Senate had given its position. In addition, Rome curtailed the power of its clients by limiting the size of their armies and navies and delimitating the sphere of influence beyond which they could not reach. Clients could only expand in certain directions in which Rome had no interest. To monitor the enforcement of those measures Rome could send commissioners, either officially or undercover, to inspect the clients' relationship with their neighbors and their respect of the army and naval restrictions. When Antiochus IV, for example, raised a new fleet in violation of his agreement with Rome in the 160s B.C., Roman commissioner Gnaeus Octavius was dispatched to investigate and was then instructed by the Senate to burn the fleet, which he did. Moreover, starting in the last years of the Republic, Rome made it mandatory for client leaders to accept summons (*evocatio*) of Roman generals demanding their presence at their quarters. Finally, while tribute and taxes were generally only collected in conquered lands and provinces as taxation was based on Roman landownership, clients were liable for irregular exactions and aid requested by Rome, whether in military supplies or manpower. Naturally, once clients, including

tribal clients, became annexed as provinces, they were required to pay the full imperial taxation, which consisted of a monetary and troop contribution to the Roman army.⁴³⁸

The clients' support was sustained via a simple system of carrots and sticks, combining the profit- and fear-based incentives of bandwagoning. Loyal client rulers were rewarded with personal honors such as Roman citizenship or special titles, territorial rewards to satiate their thirst for immediate gain, and protection from external aggression. Non-subservient clients were punished by the replacement of disloyal leaders or, when necessary, reprisals and annexation, highlighting once again that bandwagoning constituted a dangerous gamble. For most clients the threat of being invaded by Roman legions was sufficient to deter disloyal behavior, though.⁴³⁹

The extent of Roman control over its clients and the fundamental asymmetry of their relationship led anthropologist Jean Baradez to describe the client-states as "political fictions." Studying the North African kingdom of Mauritania, which became a Roman client under Augustus, Baradez notes that the kingdom was only "theoretically independent." In reality, it was administered by Roman prefects, even though it was at that point not yet a province, and was officially run by Juba II, the son of a former Numidian king defeated by Julius Cesar, who had been educated in Rome, granted Roman citizenship, and was entirely dedicated to Augustus. So in the surface client-states were led by locals—though whenever possible thoroughly Romanized locals—, likely as a move to appease their populations, but frequently the Romans took effective control of the local administrations. The "fiction" did not last long, as Juba's son and heir Ptolemy was assassinated by order of Emperor Caligula and after 42 A.D. Mauritania became

⁴³⁸ Luttwak 27; Sands 49-57, 89-98, 100-106.

⁴³⁹ Luttwak 31-7.

formally annexed as a Roman province, once again proving that bandwagoning does not pay off in the long term.⁴⁴⁰

The first clients to voluntarily join Rome outside of the Italian peninsula were Hiero of Syracuse (263 B.C.), Massinissa of Numidia (206 B.C.), Attalus I of Pergamum, Ariarathes of Cappadocia and Prusias of Bithynia (190-188 B.C.), out of fear of being overtaken by Carthage, Macedon, and Syria, respectively, and out of profit. "All these smaller monarchs had much to gain from Rome's support," Sands writes, which legitimized their rule (Hiero) and helped them gain territorial expansion (Massinissa) and favorable arbitration (Pergamum) at the expense of their neighbors. Antiochus IV, the ruler of a weakened Syria, also became a client of Rome in 175 B.C., clearly voluntarily and not as a result of the Syrian defeat against Rome, which occurred some fourteen years earlier. Rome concluded multiple other client agreements: a few examples are Ptolemy VI of Egypt in 168 B.C.; the Asiatic kings of Anatolia in 133 B.C.; Antiochus I of Commagene in 68 B.C.; Tigranes of Armenia in 66 B.C.; and the Jewish high priests in Judea in 63 B.C. All these client-states, and many more, eventually became full-fledged Roman provinces.⁴⁴¹

In the end, then, Luttwak concludes, Rome's client network was just a "vehicle of Roman control." Rome was able to rise by benefiting from its clients' military means, while simultaneously keeping them strictly subservient and non-threatening by regulating their domestic and foreign policies. By cleverly disguising its expansionist policy behind the cloak of "armed suasion," emphasizing political and diplomatic rather than purely military means, the rising hegemon retained an appearance of restraint which, through its stark contrast with the

⁴⁴⁰ Jean L. Baradez, "Organisation Militaire Romaine de l'Algérie Antique et l'Evolution du Concept Défensif de ses Frontières." *Revue Internationale d'Histoire Militaire* No. 13 (1953), 26, own translation.

⁴⁴¹ Sands 1-7; 21.

erratic behavior of the other regional great powers, attracted even further bandwagoners.⁴⁴² The terminology used by the Romans was in fact hardly accidental, as Sands contends. To avoid "offend[ing] her dependents by laying stress upon their subordination," Rome did not use the term 'subject' until the late imperial period, but instead called its collaborators 'friends' and 'clients', masquerading the relationship as an alliance on equal footing. For example, the Romans used flattering, honorific ceremonies like the recognition of a king's royal title or the signature of a 'Treaty of Friendship' as a seal of the relationship. Yet that terminology also allowed for Rome's commitment to remain loose: an assurance of friendship, with no mention of a defensive alliance, carried no obligation for active help and left Rome free to pursue its own interest at its will. While most clients were simply granted the title of 'friend', even in the cases where they were officially labeled 'ally', the term still retained a loose meaning and was given as a compliment or reward for loyal service, always at the discretion of Rome, without entailing any guarantee of assistance. Rome was only bound to a friendly intent toward its clients, while the clients carried the bulk of the burden.⁴⁴³

While the clients paid with the loss of their autonomy, Rome reaped large benefits from the deal. The client-states were not simply neutral buffer zones around the empire, but increasingly "absorbed" risk and cost by taking on the task of providing external security to the periphery of the empire. The clients became a convenient, cheap "complement" to Roman military power, Luttwak highlights. As most client-states were expected to provide their own internal security and deal with the low-intensity raids and incursions from the periphery, they essentially worked for the growing Roman empire and freed the Roman army to focus on other areas. For Luttwak, this delegation explains in large part why Rome's rise was so successful

⁴⁴² Luttwak, 38-9.

⁴⁴³ Sands 7-11, 44-5.

compared to other hegemonic bids. Delegating parts of the border security to client-states made Rome a hegemonic empire, he argues, an empire that does not need direct control over all its territory to survive, while many rising powers attempt to become territorial hegemons, exercising direct supervision over every inch of their territory. Once that territory grows, though, it is virtually impossible for an aspiring territorial hegemon to maintain control and be everywhere at the same time to prevent uprisings and secure borders. By delegating, like Rome, the hegemonic empire uses its forces much more efficiently and is able to maintain oversight of a larger territory over the long term. Indeed, in the instances where a client-state was unable to handle all internal and external security threats alone Roman forces intervened, but even then Rome did not need to establish a permanent presence in the client-state to resolve the issue. Rome pursued a limited supervisory role even in such emergencies, leaving most of the work to the client-state. While Rome was rising, this reliance on bandwagons constituted the very economical way in which the aspiring hegemon was able to maintain the security of its growing borders. It enabled Rome to rise while maintaining a relatively small army and no permanent garrisons along the borders at all until the late Republic. In case of problem at the periphery, clients would provide the forces and "geographic depth," while the rising hegemon suffered no loss of prestige in case of client defeat.⁴⁴⁴

In conclusion, potential balancers bore a large part of the responsibility for enabling Rome to fulfill its hegemonic aspirations. While the few great powers of the region that were powerful enough to counter Rome—Carthage, Macedon, and Syria—all attempted to balance, they remained mostly on their own, diminishing their chances of success. Those great powers, as well as the numerous minor powers around them were plagued by mistrust and greed for

⁴⁴⁴ This changed later during the imperial period, where Rome switched to a more active and permanent presence along the borders. Luttwak and others identify that shift toward territorial hegemony as a central cause of Rome's eventual downfall. See Luttwak 20-26.

immediate gain, which led to massive collective inaction, doomed virtually all efforts at external balancing they undertook, and generated massive bandwagoning. The rising hegemon encouraged such counter-productive behavior by its deceitful benevolence and enticing offers of support. But the potential balancers' shortfalls only provided part of the explanation for Rome's success.

When Hegemons Arise: Explaining Balance of Power Failure in World History

VOLUME II of II

A Dissertation

Presented to

The Faculty of the Graduate School of Arts and Sciences
Brandeis University

Department of Politics
Robert J. Art, Advisor

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By

Eliane Tschaen Barbieri

February 2011

The Rise of Rome II The Rising Hegemon

While together, the powers of the Mediterranean could have stopped the rise of Rome, it is not their lack of balancing alone that enabled Rome to become a hegemon. Rome also succeeded because of its own unique skills: it built a formidable military apparatus capable of overtaking any adversary and at the same time transcended military conquest by developing innovative state-building tools to consolidate its growth and foster a unified empire. By pioneering new political institutions, social relations, and economic growth, coupled with a powerful and efficient military, Rome succeeded in building the bases of a sustainable, long-lived, respected empire. As Greek orator Aelius Aristides proclaimed at the height of Rome's supremacy in 144 A.D.: "You [Romans] have made the name of Rome no longer that of a city but of an entire people."⁴⁴⁵

The second root of Rome's successful rise thus lay in its unique mastery of modern military and institutional tools and organization. Rome would not have reached hegemony without being proactive itself. Though the potential balancers' lack of strong opposition certainly facilitated Rome's task, it would not have been able to overtake its competitors and build such a vast empire had it not designed a unique, new military and state system.

⁴⁴⁵ Starr 133.

1. Military Skills and Innovation (IV5)

The superiority of the Roman military apparatus came largely from its professional, highly effective organization and strict discipline (IV5.3), reinforced by unique strategies and tactics (IV5.1) always infused by state-of-the-art techniques and weapons borrowed from competitors and improved upon to enhance them further (IV5.1). Most importantly, the Roman army did not emerge as a success overnight; rather, it improved gradually through a series of reforms and adjustments over the centuries as new challenges and new enemies arose, showing the Romans' incremental, adaptable approach to hegemonic rise.

a. Army Organization (IV5.3)

The novel, advanced organization of the Roman military, which became increasingly refined and perfected through the centuries of Roman rise, played a key role in making Roman soldiers the implement of Rome's expansion and control of the Mediterranean region.

At the dawn of Rome's rise, the Roman military was similar to the military of most other Mediterranean states, copied on the then-successful Greek model, but soon began to surpass its competitors by evolving into a highly skilled and disciplined professional army. The early Republican armies were ad-hoc Greek-style phalanxes of conscripts led by two Roman consuls, magistrates who shared the political and military leadership of Rome and were elected yearly to that position by the Senate. Each consul's army was called together for a specific campaign that usually did not last more than one season, as fighting generally occurred solely in the warmer summer months. Once the campaign was over, the army was always disbanded. Service time was not set in advance and depended solely on the duration of the military enterprise at hand. After

demobilization, the soldiers mostly went back to their previous occupations.⁴⁴⁶ But soon Roman leaders realized the shortcomings posed by their army: the phalanx formation was limiting, and the reliance on a civilian militia hampered the size of the army and the quality of its military skills.

The gradual transformation in both the tactical and structural organization of the Roman army, coupled with the army's unique role as a cement of Roman identity and its strict system of army discipline that complemented and reinforced its innovative organization, resulted in creating one of the most effective instruments of hegemonic rise in history, used as an model ever since. However, because Roman improvements in army organization were far-reaching and systematic, involving virtually every aspect of the military structure at once, from battle formation to recruitment, hierarchy, logistics, training, camps and much more, it took considerable time for others to copy it in a way that could have rivaled Rome. While Rome's competitors might have emulated a few aspects here and there, they lacked the comprehensive overhaul that gave the Roman military the ultimate edge.

1. Tactical Organization: From the Phalanx to the Maniples

The sack of Rome in the 3rd century B.C. demonstrated the inadequacy of the Greek phalanx against Rome's new barbarian enemies and initiated Rome's organizational innovation. The phalanx was likely introduced to the Romans by the Greek colonies on the Italian coast. It was a traditional formation of heavy-armored spearmen carrying large circular hoplon shields, tightly huddled together in one large rectangle 6 to 8 ranks deep and hurling onto enemy as a collective force. Most of the fighting—and losses—came from the front ranks, which would stab the opponent on contact, while the back ranks replaced them as casualties rose and provided

⁴⁴⁶ Amit 208.

moral support. Against another similar formation, the phalanx was powerful and victory largely depended on having a greater number of spearmen. But the phalanx was a heavy, awkward formation suited mostly for large battles in open country. The phalanx thus proved disastrous against the mobile, small units of Celts who took advantage of dead ground and terrain variation.⁴⁴⁷ As Webster explains, "this solid mass of heavily armored hoplites with their large spears had in its day been superb and invincible ... but against lightly armed, fast moving barbarians it was like a tortoise in a rat race."⁴⁴⁸

In 390 B.C., therefore, Rome engaged in its first military reform under Camillus, as French archeologist Paul Couissin points out, and introduced the legion of maniples, replacing the slow and inflexible phalanx system with smaller, more flexible formations designed to fight hills people and barbarian tribes on various terrain.⁴⁴⁹ The maniples system, which was gradually refined and improved, consisted of several layers of subdivisions decoupling the massive phalanx into multiple smaller, more flexible units that could operate in concert or independently and were set in a checkerboard pattern rather than one solid bloc to maximize coverage and impact-surface. Though the numbers included in each unit varied somewhat at times, Rome retained this exact system of subdivision throughout the Republic and Principate. In the main force, the infantry, the smallest unit was the contubernium, generally composed of eight soldiers who would always live together in the same tent, share meals, and could be rewarded or punished collectively, contributing to strong comradeship and attachment between soldiers reinforced their dedication in battle. As Goldsworthy notes, the new tactical organization encouraged "a very close bond between its [i.e., the army's] members, of the type observable in

⁴⁴⁷ Goldsworthy, *The Complete Roman Army*, 21-24.

⁴⁴⁸ Webster 5

⁴⁴⁹ Paul Couissin, *Les Armes Romaines: Essai sur les Origines et l'Evolution des Armes Individuelles du Légionnaire Romain* (Paris: Librairie Ancienne Honoré Champion, Ed., 1926), 177.

the small units of modern armies." In fact, "contubernalis" became "a word for close comrade and was used by officers and men alike," creating a sense of belonging and identity unique to the Roman army and unknown to its opponents' much larger units.⁴⁵⁰

Ten contubernia were grouped into a century, which was led by a centurion. The centuries originally contained roughly 100 men but by the time of Caesar were reduced to 60 to 80 men. The century became the main tactical unit after the Marian reforms of 107 B.C., emphasizing the Roman commanders' shift in focus toward smaller units and maneuverability. Then, two centuries formed one maniple and three maniples made one cohort. Caesar used the cohort as the main tactical unit, because the centuries and even the maniples had become too small to stand alone as the scope of the Roman conquests grew, and their large number made them harder to coordinate. Ten cohorts, or 30 maniples, made up one legion. Typically, a legion would contain 4,000 to 6,000 men, mostly infantrymen but also roughly 300 cavalrymen, who were divided into ten *turmae* (companies) themselves subdivided into three *decuries* of ten horsemen each under the command of a *decurion*. Under the Republic, each consul was generally handed two legions to command in campaign, though in times of extreme danger like during the invasion of Italy by Hannibal the consuls were each given four legions. After the Marian reforms the legions became permanent units and developed their own names and identities. For example, Caesar used between six and eleven legions in the conquest of Gaul. During the Principate, the total number of legions varied from 25 to 35.⁴⁵¹

The introduction and mastery of the maniples enabled the Roman army to become the tactically most proficient military in the Mediterranean because unlike the phalanx, which was used by most of Rome's competitors including the Carthaginians and the Greeks, the manipular

⁴⁵⁰ Goldsworthy, *The Complete Roman Army*, 51-54; Ferrill 8.

⁴⁵¹ Albert Harkness, *The Military System of the Romans* (New York: D. Appleton and Company, 1887), xxv-xxix, xxxiii; Polybius III.107.

formation was adaptable to any circumstance and any enemy. As Polybius explains, the maniples could not only engage in a collective frontal charge like the phalanx, when facing another phalanx, for example, but it also allowed for smaller, streamlined formations to be deployed in steep terrain and even against insurgencies, and most importantly, it allowed for individual combat within the units, which the phalanx, with its huddled mass, ruled out. In fact, Polybius concludes, the flexibility of the Roman manipular system effectively rendered the phalanx obsolete. "If it is quite possible, even easy, to evade its irresistible charge, how can the phalanx any longer be considered formidable?" he observes. While the phalanx was paralyzed as soon as the enemy refused to come out and fight in an open, level area, the maniples were adaptable to ground with obstacles, holes, ridges, rivers, hills, trees, etc. In fact, the Roman army underwent tremendous tactical progress thanks to the maniples, thereafter using terrain and hiding places to its advantage to surprise and deceive the enemy, and always favoring the better, higher ground as a rule rather than the plain. The manipular system enabled it to conceal its forces and keep reserves, which were particularly effective when used to attack an enemy phalanx from the rear, where it was vulnerable.⁴⁵²

In fact, after being defeated by Hannibal's phalanx at Cannae, the Romans soon realized that the maniples offered many more options than simply taking on a massive phalanx in the open and from the front, as was traditionally done. At the Battle of Zama in North Africa, which concluded the 2nd Punic War, for example, Scipio refused to be drawn into a head-on clash against Hannibal's phalanx again and cleverly focused the initial assault on a cavalry duel on the edges of the phalanx, in which the Romans were superior. Once the sides of the phalanx were broken down and Scipio was thus assured they would not execute a pincer movement around his forces, he lined up his infantry in different maniples, carefully leaving holes between the

⁴⁵² Polybius XVIII.29-32.

maniples through which they let Hannibal's elephants run, thus rendering them harmless.⁴⁵³ The Roman soldiers were trained to deploy in a variety of battle formations designed along the unit subdivisions introduced by the manipular system, rendering the force versatile and usable both in deep formation in thin lines depending on need. Every soldier was trained to serve a variety of roles and positions, unlike in a phalanx system where positions were rigid. In a phalanx, for example, the weaker troops were invariably placed in the middle, while in the Roman manipular army, Ferrill points out, there was "no place for weak troops" and all performed a vital role in the formation, further reinforcing each soldier's sense of worth and belonging.⁴⁵⁴

Thus the manipular organization, by increasing the range and flexibility of the Roman army, enabled Rome to score big against all of its enemies despite their military diversity. "In changing the mode of fighting so drastically the Romans showed their genius for adaptability," Webster concludes about the end of the phalanx. By creating a new tactical organization that combined strictly defined units and roles and flexibility to terrain and environment, the Romans essentially created an invincible army with the capacity to win against any enemy and in any circumstance. This novel organization eventually became the standard for all modern militaries. And because the innovation remained unequaled by Rome's competitors, it played a key role in the success of Rome's hegemonic expansion.⁴⁵⁵

2. Structural Organization: From Militia to Professional Military

The evolution from a civilian militia to a professional military took a little longer to materialize, but it gave Rome an even clearer edge over its competitors. The transformation effectively took place with the Marian reforms of the 1st century B.C., but the seeds of the

⁴⁵³ Starr 45-46.

⁴⁵⁴ Ferrill 7-8.

⁴⁵⁵ Webster 5.

reforms undertaken by Marius were sown much earlier in the Republic already. The temporary civilian militias raised by the consuls necessarily remained limited in size, since drafting too many working men in the summer months would have disrupted the economy and social organization of the state. Thus, only propertied men were eligible to serve in the military. The annual census was used to identify those citizens whose property made them eligible for service, and the Senate would for each campaign decide how many troops needed to be raised. The soldiers were thus generally farmer landlords who had enough wealth to afford weapons, which the recruits had to procure themselves—the wealthiest men served in the cavalry, the equites, since they could provide horses—and to delegate their civilian duties and who had a vested interest in the integrity of the state and its territory to preserve their land possessions. Under the militia system, serving in the army was therefore considered patriotic, a "duty owed to the state," according to Goldsworthy. The tactical organization of the militia similarly reflected the socio-economic rank of the soldiers, as the main infantry line of the phalanx was arranged by wealth and age. For example, the hastati, the poorest citizens among those eligible to serve, who could provide the most modest weapons and material, formed the light infantry and generally composed the first, most vulnerable battle line, while the triarii, the oldest, wealthiest, and most heavily armed soldiers, were set in the rear and often were spared the battle altogether as the previous lines had won before it was their turn to intervene.⁴⁵⁶

This relatively small, temporary militia remained sufficient throughout the early to mid-Republic, as long as Rome's conquests—Italy, Carthage, and the Western Mediterranean—were located in its immediate vicinity. But as Roman expansion spread further beyond Italy, campaigns started stretching over not just one season but over several years, requiring garrisoning abroad and creating impossible dilemmas for civilians with full-time occupations. In

⁴⁵⁶ Goldsworthy, *The Complete Roman Army*, 7-8, 26-28; Starr 51.

addition, the militia system posed growing problems of quality and consistency of military skills. The men were discharged after every campaign and returned to non-military activities and thus had to be retrained each time a new army was raised. In addition, they could not retain the experience they accumulated at war because they were assigned to different units under different officers each time a new consul was elected and a new army raised, in general every year. Moreover, because of the lack of a standing army the Romans' use of their military was always reactive, and never preventive. This seriously weakened the credibility of their deterrent and coercive positions. Marius's 107 B.C. reforms set out to fix those problems, replacing the temporary civilian militias with professional legionaries, permanent soldiers for whom the military was a career and who signed up for decade-long service.⁴⁵⁷

The professionalization of the military was initiated long before the Marian reforms, though, after the battles of Lake Trasimene and Cannae in 217 and 216 B.C., where Rome was severely defeated on its own soil by Hannibal. Just like the sack of Rome prompted military innovation in the abandonment of the burdensome phalanx, the disasters at Lake Trasimene and Cannae triggered the transformation from militia to professional army that would eventually be ratified by the Marian reforms, by demonstrating the qualitative superiority of Hannibal's army even after its long, decimating journey through the Alps and despite its hodgepodge ethnic composition, and the need to improve the proficiency of the Roman army. The Romans first realized they needed a more able and unified leadership and appointed a number of dictators, or temporary extraordinary magistrates combining the ultimate civilian and military authorities, in the years following their defeats by Hannibal to oversee the dual, rotating consuls. In addition,

⁴⁵⁷ Goldsworthy, *The Complete Roman Army*, 43-46.

they propelled Scipio Africanus, a commander with genial tactical skills, to lead the army and defeat Carthage.⁴⁵⁸

Scipio began the process of professionalizing the army itself by two steps: first, he developed a more demanding training program for the Roman soldiers, designed to help them master both high speed tactics and surprise attacks but also methodical planning; second, Scipio began relying increasingly on Roman volunteers at the expense of conscripts, a trend that would continue after him until Marius sealed the fate of the militia by officially making the Roman military a volunteer-based career army. When in 205 B.C. Scipio, elected consul, decided to pursue the Carthaginians on African soil to end the 2nd Punic War, the Senate was wary of the additional hardship onto the Roman populace. Over a decade of war in Spain and Italy had rendered military service highly unpopular and expanding commercial prospects with the East had multiplied the civilian ventures citizens were involved in. The Senate thus forbid Scipio to use conscripts. As a result, Scipio raised and trained an army entirely made-up of volunteers, who displayed tremendous military skill during the campaign in Africa. Future consuls remembered the qualitative superiority of Scipio's volunteers and continued relying heavily on volunteers. Consul P. Licinius Crassus, for example, led an almost-exclusively volunteer-based army to defeat Perseus in the 3rd Macedonian War.

The first legal evolution toward the Marian reforms occurred in 152 B.C., when automatic conscription was revoked and replaced by a lottery system limiting service abroad to six years. The shift was prompted not only by the need for better trained and better skilled soldiers, but also by a growing social crisis resulting from the disruption conscription had caused among the mostly agricultural basis of the Roman economy. The lengthening campaigns had rendered small landowners bankrupt and led them to flock to the capital as larger, better

⁴⁵⁸ Webster 15-17.

endowed landowners took over their farms, creating a growing class of nouveaux-poor and a battery of social unrest that eventually fostered the civil war that took down the Republic in the 1st century B.C. The newly emerging volunteer army provided a response, a new career opportunity for the landless class, and Marius, seizing upon the recruitment potential that class represented for an army that had traditionally suffered from a limited size, abrogated the property requirement to join the army and essentially opened up service to Rome's new proletariat. Just like Scipio before him, Marius, upon being elected consul and assigned to replace the Roman commander in Numidia in 107 B.C., was forbidden by the Senate to raise legions with conscripts, as many senators opposed the war. He had to rely on volunteers, thus, but for the first time, Marius called on the poorest citizens, those without enough property to qualify for military service. Many eagerly joined, glad to be granted an escape from their dire everyday life, and proved just as skillful soldiers as the propertied men. The gradual professionalization thus not only improved upon the Roman army's quality, but also upon its numerical capital.⁴⁵⁹

With the Marian reforms, the army officially became a volunteer force made up of career-soldiers who signed up for 16 to 20 years. They remained assigned to the same unit, under the same commander, fostering a new identity shifting away from the Roman state and associating strongly with the unit and military commander instead. The class distinctions and privileges traditionally upheld in the Roman military, including the hastati and triarii, were similarly erased. Experience and technical knowledge was preserved overtime and specialization was encouraged, with special units of expert craftsmen and engineers formed. The Roman army became better trained and disciplined also because they now benefited from more consistent,

⁴⁵⁹ Webster 17-19, 30.

capable officers. Soldiers were from poorer, less socially prominent backgrounds with virtually no political influence and were thus less bribable.⁴⁶⁰

With the professional army also came the standardization of pay. While in the early Republic the citizens in the army armed and supported themselves, this was no longer possible once the non-propertied class became eligible to serve. The new post-Marian professional soldiers were thus given all equipment and armament by the state as most of them could not afford it. And while some compensation had been given to soldiers since 396 B.C. to make up for the lost time and income at their regular jobs during service, this was hardly a source of income, so after the Marian reforms the new career soldiers received full pay equivalent to that of the laborers in Rome, as well as substantial bonuses, mostly in the form of booty confiscated from the enemy, and presents. A small amount of their pay was subtracted for supplies furnished by the state.⁴⁶¹ The pay system was perfected overtime. It is difficult to calculate the legionary's exact pay because parts were received as equipment, ration, rewards, and booty, parts were deducted and the pay changed overtime and gradually increased with inflation, but it was set and clearly regimented by rank. Caesar reportedly doubled the pay of legionaries to 225 denarii a day. It was still 225 under Augustus, and remained that way until Domitian increased it to 300 denarii. Officers were usually paid 1.5 to 2 times as much as legionaries. Centurions were likely paid at least 5 times as much as ordinary legionaries, and senior officers twice as much as ordinary centurions. Bounties, donatives, and discharge grants were all in proportion to rank also. There are virtually no records of the pay of auxiliary troops, but experts assert that they were paid considerable less, perhaps as little as half the salary of citizen legionaries, though they

⁴⁶⁰ Starr 66-67; Goldsworthy, *The Complete Roman Army*, 46-48.

⁴⁶¹ Harkness xxxii.

shared in the spoils of war and other advantages.⁴⁶² By the early Principate, Rome had a number of regimental savings banks where soldiers were required to deposit half of their pay, as Vegetius describes, "to prevent it from being wasted by them through extravagant and the purchase of useless articles." A few decades later in 89 A.D. under Domitian, however, the deposit requirement per soldier was capped at a certain maximum amount after the governor of Germania, Lucius Antonius Saturninus, who held a grudge against the Emperor, managed to mount a revolt using the combined savings of two legions.⁴⁶³

The next set of reforms came with the Principate. The professionalization of the military had the perverse effect of switching the loyalty of soldiers from the state to their commanders—consul or legate, legion commander—who were responsible for providing them support, generally a land grant from recent acquisitions, for their life after the army. Those pension lands had played a crucial role in the spread of the empire because they were an instrument of administrative, political, and military control of the newly acquired, frontier-areas of the empire. The veterans discharged from the legions were sent there by their commanders to essentially Romanize those areas by establishing colonies, and occasionally act as a "ready-made militia" to "defend their [new] hometowns in the event of an attack and hold out until [Roman regular] forces could arrive on the scene," Luttwak explains. The fact that commanders were responsible for their soldiers' future after service naturally created a very strong bond between the troops and their commanders, and eventually enabled the civil wars that tore apart the Republic during the 1st century B.C. When some commanders eventually turned against each other, their subordinates followed them.⁴⁶⁴

⁴⁶² Webster 264-8.

⁴⁶³ G.R. Watson, *The Roman Soldier* (Ithaca, NY: Cornell University Press, 1969), 104-5.

⁴⁶⁴ Luttwak 19; Eckstein 135-6.

Augustus, after winning the civil wars, solved the problem by providing a unitary leadership for the military in the Principate. He placed the legions and auxiliaries under the sole command of the Emperor, who was himself also in direct charge of the more unruly provinces that could not be trusted to a governor and could take over war command whenever necessary. This new system of direct, permanent, and centralized military command provided much clearer, focused authority than the rotating multiple-hierarchy leadership of the consuls of the Republic. All military personnel owed allegiance to the Emperor. In addition, the Emperor took the responsibility of compensation for the veterans away from unit commanders and set a fixed compensation, the *aerarium militare*, to remedy the problem of soldiers' exclusive loyalty to their commanders. Retiring soldiers and officers were given a set cash amount or equivalent land grant upon completing their service. Augustus created a special fund subsidized by his own money and two special taxes on the Roman citizens to that effect. Thanks to this shift to a unitary leadership structure, Goldsworthy concludes, "under the Principate the Roman army reached the peak of its efficiency, completing the conquest of the Empire and then preserving Roman rule."⁴⁶⁵

Augustus also finalized the professionalization of the military by banning non-soldiers from bearing arms. This was another very novel concept unique to Rome at the time, that served to separate the military and civilian spheres and removed the use of force from everyday Roman life, making the empire an attractive place to live. With a population of about 50 million and an army of roughly 300,000 by the early Principate, Rome had a very low percentage of men under arms (0.6%), sign of a cost-efficient military that could be financed without much strain on the population or economy. Because the vast majority of the population participated to economic activity and not to the army, the Roman empire was thus able to preserve a high growth rate

⁴⁶⁵ Amit 208; Couissin 177; Goldsworthy, *The Complete Roman Army*, 8, 49; Ferrill 3-5.

while at the same time maintaining a sufficiently large army.⁴⁶⁶ By creating a new type of army, entirely composed of specialized, career soldiers, Rome not only maximized the skill level of its military, it also maintained military expansion as a separate activity and thus kept its population content and focused on other areas of growth while its opponents concentrated all their effort on their military.

3. Ethnic Organization: Legionaries and Auxiliaries

One of the other unique traits—and strengths—of Rome's army was its exclusive reliance on citizens to fill its ranks. Whether they served temporarily and then went back to their regular occupation, like in the militia of the early Republic, or whether they were career soldiers like after the professionalization of the army, Rome's legionaries were always citizens, not hired mercenaries. In contrast, most of Rome's opponents, notably Carthage, used mercenaries, professional foreign fighters contracted by the government to conduct wars. While mercenaries might have accumulated better military skills and experience, their major drawback was that they were motivated to fight mainly by the monetary compensation promised to them and did not display any particular loyalty toward the government that hired them, quite the opposite. For example, Polybius reports that before the battle of Lilybaeum in Sicily during the First Punic War, the Carthaginian commander had to promise his mercenaries all sorts of rewards to prevent them from handing the city over to the Romans. Mercenaries in the Mediterranean tended to lend their services to the side that paid more, and did not hesitate to switch sides in the middle of a war if the opponent offered more. "The Carthaginians came very near to losing everything they had in Lilybaeum through the treachery of their mercenaries," Polybius deplores. This was not an isolated incident. In addition, the mercenaries' more extensive fighting experience did not

⁴⁶⁶ Ferrill 34.

necessarily result in a better quality military. "The Carthaginians [despite being excellent seamen] largely neglect their infantry," Polybius continues. "The reason for this is that they employ foreign mercenary troops, whereas those of the Romans are citizens and natives of their own country," who were more dependable and dedicated to the war effort because they had a vital, not merely monetary, stake in its outcome and were willing to sacrifice their lives and take larger risks to prevail. The Roman citizen army was thus superior, Polybius concludes.⁴⁶⁷

Perhaps the biggest problem professional citizen armies encounter is to generate enough recruits, particularly when the needs are very high, like in long wars. On a few occasions during the Principate the Roman army had to resort to conscription to fill its ranks despite having a professional military, during some revolts, notably, but it kept the instances rare because it angered the population. Fortunately, the Romans had another, much larger source of soldiers. While the main legions were all composed of Roman citizens, each legion was attached to an auxiliary unit composed of soldiers of foreign nationalities. Unlike the citizen-legions, the auxiliary units were generally drafted, with the exception of some volunteers, from among Rome's client-states and later provinces but commanded by a Roman leader, first a prefect and later a cavalry commander. During the Republic Rome acquired a large number of auxiliary levies from states or tribes as a counterpart for accepting them as a client or friend. Once those clients and friends became provinces during the late Republic and Empire they continued providing troops for the auxiliary units. Some auxiliary units were armed and trained by Rome, others by their home country or province, depending on their agreement with Rome.⁴⁶⁸

Relying on a foreign force for half of its army was a brilliant strategic move by Rome. Auxiliary soldiers provided the perfect way to swell the ranks of the Roman army beyond any

⁴⁶⁷ F.E. Adcock, *The Roman Art of War Under the Republic* (Cambridge: W. Heffer & Sons Ltd., 5th Ed. 1981, or. Harvard University Press, 1940), 5-6; Polybius I.43, VI.52.

⁴⁶⁸ Ferrill 24-25.

level reachable by its competitors and also "relieved the pressure on the available pool of citizen manpower ... and [simultaneously] reduced the financial burden on the Roman empire," as Luttwak points out, since auxiliaries were supported by Rome but did not earn the large pensions of retired legionaries. In addition, the auxiliary soldiers brought to the Roman army military specialties it did not have before or was not proficient in, for example mounted archery. As Goldsworthy explains, the auxiliary units were "important in providing soldiers whose fighting styles were particularly suited to the conditions of the region" where the fighting occurred. Thanks to the auxiliaries, the Romans became even more adaptable tactically.⁴⁶⁹

The auxiliary units were generally in equal numbers to the citizen-legions, according to Tacitus, but their composition was different. The auxiliary infantry was organized in cohorts and sub-units similar to the legion, but it was light, not heavy, infantry. In addition, it had a contingent of cavalry about three times larger than that of the legion. The auxiliary in fact gradually became the main cavalry force, while the legion's main focus and skill was infantry. In fact, by the early Principate, the legions' cavalry was cut down to a few hundred riders maximum whose missions were mainly to be "messengers, escorts, and scouts," Ferrill notes. The auxiliary cavalry was the only one to participate on the battlefield.⁴⁷⁰ Auxiliary units therefore provided fundamental support to the regular legions, not only because of the large number of additional soldiers they entailed, but also because of their distinct role as light-infantry and cavalry units. This explains why in the early Republic the auxiliary force was called "alae" or wings because its cavalry and light infantry often operated at the sides of the main phalanx. Its role only became more important as the number of Roman allies grew. In his conquest of Gaul, for example, Caesar used on average 20,000 auxiliary troops each year, and at a maximum 70,000. His

⁴⁶⁹ Luttwak 16, 27; Goldsworthy, *The Complete Roman Army*, 55.

⁴⁷⁰ Tacitus IV.5; Ferrill 6-8.

auxiliary cavalry was composed mainly of fellow Gauls, Germans and Spaniards. The auxiliary infantry had Balearian archers and Cretan slingers, who had a reputation for remarkable military skill. By the time of Augustus and the Principate, the auxiliary became a regular and professional force just like the legions.⁴⁷¹

The Roman army thus relied very heavily on foreigners, and one may think it to have raised a problem of reliability of the troops, similar to that of the mercenaries. It seems like a dangerous gamble to operate with half of one's army having different origins and thus potentially contradictory allegiances. The Romans, however, managed to construct auxiliary service into an inclusionary system that fostered support and loyalty for Rome from its allies and provinces. Service in the auxiliary force soon became a vehicle for entry and integration into Roman society. While the army remained tolerant of ethnic and religious practices, it promoted the Roman way of life and Roman cultural traditions. One tool used for that purpose was language, by requiring the use of Latin for all official military interaction, such as command and army administration. Auxiliary recruits thus had to promptly learn the Roman language, though we have no estimation of auxiliary or even legionary literacy. This requirement increased both the sense of unity and the ease of command in the Roman army. Ferrill contrasts the Roman army's surprising cohesion in spite of its diversity to the pre-World War I Ottoman army, which presented a similar ethnic diversity. He notes that while the Romans always issued orders in Latin, the Ottoman hierarchy sent out the mobilization order for World War I in fifteen different languages, generating much confusion.⁴⁷²

Besides subjecting them to Rome's cultural influence, being part of the army also awarded the auxiliary soldiers a unique opportunity to become Roman citizens themselves. At

⁴⁷¹ Harkness xxv-xxix; Goldsworthy, *The Complete Roman Army*, 57.

⁴⁷² Ferrill 19.

first citizenship was granted exceptionally to auxiliary units who displayed extraordinary bravery, as happened to a Spanish contingent after the siege of Asculum in 89 B.C. Then, to reward its Italian allies' participation in the expansion of the empire, Rome granted citizenship to all Italians who had not rebelled against Rome in 88 B.C. As a result of becoming Roman citizens, those Italians serving in the auxiliary force became absorbed into the regular legion. Emperor Claudius eventually extended the privilege to all auxiliary units as a reward for service, granting them citizenship upon discharge. Claudius fixed auxiliary service to a maximum of thirty years, and the auxiliary soldier was eligible for citizenship, along with his wife and children, after 25 years of service. The citizenship reward became a massive incentive for provincials to join the Roman army and serve loyally. In addition, auxiliary troops benefited from bonuses and pensions if they made it to the end of their service, though less advantageous than those of regular legionaries, and they also had access to the same quality medical care and hospitals than the legionaries. The Roman army, by playing an inclusionary social role, thus contributed to nurturing the idea of Rome as a nation transcending the multiple ethnic origins represented in the empire. In reality, there are very few instances in which the auxiliary units proved unreliable. Goldsworthy can find only one example, a 212 B.C. battle against Carthage lost by Rome after an auxiliary contingent of Celtiberians abandoned the Roman army.⁴⁷³

The combination of citizen legions and non-citizen auxiliaries not only supplied Rome with a superior military force in both quantity and quality to complete its hegemonic expansion, but it also built the social foundations of its empire by attracting foreigners to the Roman culture and national ideal and granting them an individual opportunity to integrate and benefit from its success. The army, by its inclusiveness, served as a cement for the Rome's new hegemonic identity. In addition to the legions and auxiliaries, Rome kept a navy, urban cohorts and

⁴⁷³ Goldsworthy, *The Complete Roman Army*, 47, 56, 80; Webster 142-3.

vigilantes—the equivalent of our police and fire departments—in charge of maintaining order in Rome itself, and under the Principate also praetorian guards who served as the Emperor's personal guard, a militia to keep order on the Italian peninsula (since the legions were always outside Italy) and a training camp for future high officers of the regular army. Yet the legions and auxiliaries provided the military cornerstone of Rome's force rise to hegemony.⁴⁷⁴

4. Sustaining the Organization: Rome's Unparalleled Military Discipline

Structural and tactical organization alone did not make the strength of Rome's military apparatus. Rome was only able to sustain its military preeminence thanks to its unparalleled and unprecedented discipline. The literature overwhelmingly credits the Roman army's unique, rigid disciplinary system for Rome's achievements. Discipline, Harris contends, was the Roman army's greatest asset because it enabled the Romans to conduct ordered and controlled campaigns and show exceptional resilience under duress. Similarly, Harkness writes, "the success of the Roman army was due largely to discipline and military drill."⁴⁷⁵ Ferrill further argues that ...

... because the Roman army was so well trained and professional, it always had operational and tactical advantages over its opponents, usually even in the wars against the major states of the Eastern Mediterranean. Against barbarian armies Roman forces were invincible, barring stupid mistakes of generalship.⁴⁷⁶

The Roman army's disciplinary excellence took on many aspects: training and obedience, hierarchy, marching and battle orders, camp order and etiquette, field communication, and logistics.

⁴⁷⁴ Amit 208-9.

⁴⁷⁵ Harris 103; Harkness xxxii.

⁴⁷⁶ Ferrill 8-9.

The Roman military, because it used a civilian force, took training particularly seriously. Vegetius, in his late-Empire military treatise, details the content of that training, which was centered on physical fitness and repetition. When a soldier joined—or was drafted, in the militia years—he would first undergo basic training for several months before becoming a full member of a unit. Then, according to Vegetius, the Roman soldier's routine included running and jumping exercises, short-order drills, long marches in full gear at both normal and fast paces, often with extra load, weapons training with blunt-tipped or wooden swords and spears, and mock battles. Training remained continuous throughout the service, during both peace and war times. Some emperors ordered large-scale exercises involving entire legions, and even mock campaigns. Hadrian, for example, was well-known for touring all his provinces and ensuring that the troops were properly trained and up to par. Josephus further contends that Rome's non-stop training routine was very different from that of its opponents that mostly lacked the serious preparation Roman soldiers underwent, though he admits that the effort was somewhat followed less dutifully in the provinces furthest away from Rome.⁴⁷⁷ Ferrill adds that Roman soldiers developed unequalled military skills because their training was particularly intense. "Modern states do not permit their troops to be treated as harshly as Roman soldiers were," he explains. Ferrill estimates that thanks to their training, Roman armies benefited from a 4-to-1 superiority, at least against barbarian enemies. When factoring in logistics and equipment in addition to the training, the advantage was "immeasurable," he writes.⁴⁷⁸

The thorough training was accompanied by a strict code of obedience, including rewards and particularly harsh punishments, to incite good behavior. In fact, while serving Roman

⁴⁷⁷ Publius Flavius Vegetius Renatus, *De Re Militari (The Military Institutions of the Romans)*, transl. by Lieut. John Clark, ed. By Brig. Gen. Thomas R. Phillips, U.S.A. (Westport, CT: Greenwood Press Publishers, Repr. 1985); Josephus; Goldsworthy, *The Complete Roman Army*, 80-81, 92.

⁴⁷⁸ Ferrill 8-10.

soldiers essentially lost many of the progressive legal rights Roman civilians enjoyed: they could be beaten or even executed at the mere order of an officer. Some crimes were punished much more severely than in civilian life. "Both legally and ideologically, a marked distinction was maintained between the status and appropriate behavior of Romans at home (in domi) and at war (in militiae)," Goldsworthy emphasizes, and the resulting docility and resilience is largely responsible for making the Roman army superior to all other contemporary armies.⁴⁷⁹ Polybius informs us that the Roman army had a pre-established system of rewards and punishment. Minor offenses were answerable by soldiers to their officers and by officers to the consuls. For severe offenses, a court martial was to determine guilt. Some offenses carried a death sentence, generally administered by beating: failing to submit guard duty report tablets or falling asleep while on guard duty, stealing from the military camp, giving false evidence, engaging in homosexual acts, fleeing the battlefield or displaying cowardice, and having been punished three times for the same offense. Sometimes punishment was applied to a whole unit. In case of severe misbehavior by group of soldiers—if a whole unit deserted a battle, for example—the rule was to randomly select one in ten men from the offending group and apply the death sentence to those selected, while humiliating the rest of the group by feeding them barley instead of wheat and forcing them to sleep outside the camp without defenses. The brutal randomness was designed to work preventively. Polybius insists that just as punishments were strict and systematic, rewards were also always distributed for courage and bravery on the battlefield, for example when men deliberately exposed themselves to danger or saved fellow citizens or allies. Rewards included public praise by a consul in front of all the troops and various gifts from the consul such as specially decorated cups, spears, shields, or horse-trappings. Regardless of its nature, a reward

⁴⁷⁹ Goldsworthy, *The Complete Roman Army*, 33.

implied great prestige for the recipient, not only in the army but also upon returning home. For example, reward holders could claim special honorific places in religious processions.⁴⁸⁰

In spite of its harshness the Roman army's system of rewards and punishments worked because it was applied equally to all, generating a sense of justice that reinforced the loyalty of the men. Because it was applied "scrupulously," Webster argued, with gradual punishment depending on the severity of the offense, the disciplinary system was perceived to be fair and was thus respected.⁴⁸¹ The Roman military displayed the same level of fairness in determining the career advancement of officers. While it is undeniable that the aristocratic class of senators provided most of the provincial governors and senior army officers, and every military position was clearly not open to the average Roman, the Roman system rewarded experience and merit at least as much as wealth and personal background. While letters of recommendation were always needed to move to a higher position, achievement and proven leadership remained key to advancement. This was particularly true under the Principate when the emperor's approval was required for the appointment of any legionary commander (legate) or provincial governor, and emperors naturally preferred capable and loyal men to well-established socialites who were more likely to challenge their rule. The ladder to climb to pursue a public career necessarily involved military service in Rome, so the military often served as stepping stone for men without extensive personal relationships and privileged family backgrounds to distinguish themselves and build their public career.⁴⁸²

Besides career advancement, individual rank promotion within the army followed an even more meritocratic process. While the highest positions in a legion required equestrian or senatorial—i.e., aristocratic—rank, these were not necessarily the officers with the most say in

⁴⁸⁰ Polybius I.17; VI.37-39.

⁴⁸¹ Webster 13-4.

⁴⁸² Goldsworthy, *The Complete Roman Army*, 62-3.

military decisions. In fact, those senior officers appointed because of their wealth, social position, or personal and political friendships, like the six tribunes of the legion, mostly served as administrators for the legion or staff officers for higher officers. They were only occasionally given command of actual detachments, and when they were, they generally commanded small units for special missions. The real military power lay with the subaltern officers, the centurions, whose position was accessible to any deserving soldier by promotion. As Adcock explains, "when the legion fought as a whole, the leadership of its several tactical units was vested, not in aristocratic officers, but in experienced soldiers of the social class of the private." He adds, "thus what mattered most in the Roman conduct of a battle, its tactical performance, was left in the hands of men of the social class of the ordinary legionaries, who lived with their men, who had to maintain themselves ... by their professional competence and bravery and endurance in the field."⁴⁸³

Those officers not only determined most tactical moves, as Adcock points out, but they also often took part in strategic decisions as the legion commanders generally opened up their war council meetings to the centurions, especially the senior centurions, and welcomed their comments. It was standard practice, already in the Republic, for the commander to hold such a "consilium" to formulate war plans for a projected campaign or operation. The officers would regularly give their recommendations, even offer a certain course of action. And although the commander was ultimately free to decide and override them, the process was one of debate where a plurality of opinions and analyses were presented, contributing to a higher-quality decision. The centurions technically commanded only a century, a maniple, or a cohort, Harkness underlines, but in reality they ran the whole legion. There was a total of sixty centurions in a legion, all generally with substantial military experience, who reached the

⁴⁸³ Adcock 17-18.

position through promotion "as a reward for good service." Thus, Harkness concludes, the Roman system "aimed to stimulate ambition and reward fidelity" and kept all sides happy by cleverly combining hereditary and social appointments with merit-based appointments.⁴⁸⁴

In addition to keeping a strict hierarchy, the Roman army also kept a strict order in the ranks at all times, whether for simple marching, for battle, or even for plunder. The Roman army had a codified, rigidly enforced order of march for all movement, with exact distances to be maintained between cohorts, units, and individuals following a number of pre-defined marching formations. A given day's marching order was "determined by the potential hazards of the terrain and danger for the army," with different possible formations depending on the imminence of threat, Webster describes. While marching the troops were always preceded by a vanguard. Only then came the main body, and then a rear guard. The baggage train always came last, with animals carrying the tents, tools, and supplies, though each soldier carried his own basic bundle of supplies in case they became separated from the baggage train. The rigid organization was designed to minimize surprises, but because commanders had a variety of marching formations available and the transitions from one formation to another were routinized and internalized for increased efficiency, they also had the necessary flexibility to adapt to unforeseen circumstances. At any time positions and locations in the legion were highly regimented, and each individual involved always knew what his role was and where he was expected to be, when, and for what purpose. Yet troops also knew to shift smoothly and fast to different roles and positions with just a command, following a "textbook pattern" that every man was familiar with. For example, when a legion left its camp, it was in the same sequence each time. At the first bugle call, soldiers would take down and pack the tents and baggage. At the second bugle call, they would load the baggage on the pack animals. At the third bugle call, the column would start marching. The strict

⁴⁸⁴ Harkness xl-xliii; Polybius VI.19-21, 24, 26; Goldsworthy, *The Roman Army at War*, 131-133.

regimentation of roles and positions in the marching ranks was continued throughout the imperial period, a testament to its success.⁴⁸⁵

An even greater diligence was applied to the battle arrangements, and every man similarly knew his roles and positions and was trained to switch rapidly if told, providing for both orderly and flexible maneuvers in combat. Just like with the marching order, the army leadership had at its disposition an array of preset combat formations internalized by the troops to respond to different situations, depending on where the enemy was expected and what type and number of enemy units were anticipated. As Harkness emphasizes, in many of the combat formations the cohorts and even smaller units were disposed in three lines for attack, in order to maximize their strengths: the first and second lines would alternatively relieve each other so as to sustain combat as long as possible, while the third line was kept in reserve and sent to the front only if necessary. The constant provision of reserves highlights that the Romans were far ahead of their time in military thinking. It represents an innovation in doctrine that foreshadowed the key role reserves would play in modern force employment.⁴⁸⁶ Most of Rome's competitors, in contrast, focused all troops on the onset of battle and moved at once in one mass to surpass the enemy in troop concentration and momentum, but their lack of depth was recurrently a fatal flaw against the Romans. If the first onslaught was not victorious, they had no plan B as the Romans launched their reserves from behind the battle line. Thus, Polybius contends, the novel combination of rigid, pre-arranged formations and capacity for rotation and flexibility upon demand made the Roman army quasi-invincible:

The order of battle used by the Roman army is very difficult to break through, since it allows every man to fight both individually and collectively; the effect is to offer a

⁴⁸⁵ Webster 14-15; Harkness xliii-xlix.

⁴⁸⁶ See Stephen Biddle, *Military Power: Explaining Victory and Defeat in Modern Battle* (Princeton, NJ: Princeton University Press, 2004), Ch. 3, particularly 46-8.

formation which can present a front in every direction, since the maniples which are nearest to the point where danger threatens wheel in order to meet it.⁴⁸⁷

Strict order went even beyond marching and battle formations. The intolerance for disorder in the Roman army led commanders to regulate everything down to plunder, which was central to the army's war activities. All armies at the time engaged in plunder after taking over a hostile town or area and it contributed a vital share of their financing. Plunder was expected upon defeat, but some armies displayed particular ruthlessness while others plundered in more civilized ways, respecting the lives and integrity of those defeated and concentrating on extracting their material riches. Roman commanders, unless the defeated enemy had shown particular brutality itself, generally insisted upon following certain rules to make plunder fairer, not only for the defeated, but also for Roman soldiers. Polybius describes the plunder following Scipio's defeat of the city of New Carthage, Spain, in 209 B.C. The booty was collected then brought to the central marketplace, where the senior officers divided it among their respective units. It was then divided equally within each unit, down to the individual soldier. In general, Polybius explains,

after a city is captured, the Romans adopt the following procedure with the spoils. Sometimes a proportion of the soldiers from each maniple and sometimes the whole maniple itself—... depending on the size of the city—is detailed to collect the plunder. They never use more than half the army on this task ... All those who have been detailed to collect the plunder then bring it back, each man to his own legion, and ... the tribunes distribute the proceeds equally among all, including not only those who have been left behind in the protecting force, but also those who are guarding the tents or tending the sick, or who are absent on any special duties.

⁴⁸⁷ Polybius XV.15; Harkness lix.

The organization of the plunder was thus just, because Roman leaders understood that granting rewards for a victory, but also preventing jealousy and grievances stemming from discrimination and preferential treatment, were key mediums for a content, obedient force.⁴⁸⁸

Once marching, battle and plunder were over, the next target of Roman discipline was the camp set up for overnight respite. The rigorous discipline extended to the camp because it was vital not only to the legion's defense and battle preparation, but also to the battle itself. If a battle did not succeed as planned, a solid camp was necessary as a safe escape ground. Thus, whenever they needed to retreat, the Romans rarely displayed chaos or faced the risk of a general massacre because as a rule they always kept a solid camp in rear to get to safety. In addition, since the Romans were constantly traveling through enemy country, they expected their camps to come under attack frequently. As Vegetius emphasizes, the Roman camps were thus always fortified, even for a brief, one-night stay, and always guarded, even during a battle. Even when there was no immediate threat, a Roman army on the move or in maneuver always built a temporary camp to dwell overnight, and the camp was by rule surrounded by a wooden palisade and a 9-foot-wide and 7-foot-deep trench. In case of elevated danger, the trench was increased to 12 feet in width and 9 feet in depth. Camps for longer term stays were of course built with even larger and deeper entrenchments.⁴⁸⁹ Archeological finds in France show that Caesar's camp in Gergovia was built with a 15-foot-deep moat, while his camp at Alesia featured a 20-foot-wide trench.⁴⁹⁰

Guard duties naturally followed rigid rules also. In general, camps were always set up in advancing positions, for example in slight altitude rather than on an open plain, so that guards could more easily spot enemies and hurl their javelins onto oncoming attackers. Camps were guarded at all time, and the responsibility for guard duty rotated between the various cohorts.

⁴⁸⁸ Polybius X.16.

⁴⁸⁹ Vegetius I.24, III.8.

⁴⁹⁰ Ferrill 10-11.

The guard's task was prescribed with minutia, and each guard was required to submit a report at the end of his shift. Any digression immediately led to a court-martial, and inattention on the job was punished by death, as mentioned above.⁴⁹¹

The strict discipline extended to the inside organization of the camp. Each legion, cohort, maniple, contubernium, and individual was assigned to a specific location within a pre-defined amount of space. Every camp was set up the same way, so that each soldier and officer always knew where everyone and everything was located. The similar nightly arrangement also brought the troops comfort and reassurance. The camps housed not only the legionaries but also the auxiliaries and was thus another way to make them feel like they belonged with Rome and the Romans. As Polybius describes it, "one simple formula for a camp is employed, which is adopted at all times and in all places." The praetorium, or headquarter of the legion with the consul's (later the legate's) and senior officers' tents, was located in the center of the camp, in a 10,000 sq. feet square easily accessible from all sides of the camp. The subaltern officers' tents were then set up in parallel to the sides of the praetorium, and the legionaries' tents in parallel to those subaltern officers' tents. When the camp housed a consul's two legions, which happened frequently until Emperor Domitian banned multiple legion camps after the Saturninus mutiny of 89 A.D., each legion would occupy one side of the praetorium, and the auxiliary and allied armies would occupy the other two sides. The troops' quarters were arranged along a system of streets and crossroads for easy access and communication. In the end, the camp resembled a series of squares representing blocks of tents, leading concentrically all the way up to the praetorium. This simple organization was easy to expand or retract and was adaptable to the size

⁴⁹¹ Harkness xlix-lvi.

of the army, which constantly varied, and was thus another example of the Romans' inventive adaptability.⁴⁹²

Besides accommodation, field communication was another key aspect of the Roman military that was closely regulated with the same strict discipline. The Romans used both visual and sound signals to communicate on the field. Those signals were crucial during battle, but also before and after battles, to dispatch information from the front to rear-area defenses. Maintaining effective lines of communication was critical to enable Rome's use of reserves and in-depth positioning. Unfortunately there are few descriptions of field communication in the primary literature. One common tool was standards and banners, which carried a different meaning depending on their position and location. For example, if the standard was pitched in the praetorium in front of the commander's tent it instructed the troops to get ready for battle. If it was waved in front of the legions ready to engage in combat, it was the signal to charge. Each legion, cohort, and cavalry unit possessed its own distinctive standard and several banners. The trumpet was also used to signal different actions and movements. A special tune indicated attack, retreat, and given formations. In the cavalry the trumpet was replaced by a horn.⁴⁹³

Polybius reports that he personally contributed to improving the Roman army's torch system, a primitive ancestor of the Morse code. The Latin alphabet was divided into five groups of five letters. A first set of torches would indicate what group of letters a letter belonged to—1 through 5—and a second set of torches, situated a short distance away from the first, would signal the position of the letter within the group—again, 1 through 5. That way, a message could be sent letter by letter, with torches at night and smoke columns during the day.⁴⁹⁴ Vegetius describes a semaphore line that was used to relay more complex messages over larger distances,

⁴⁹² Polybius VI.26-31.

⁴⁹³ Harkness lvi-lviii; Webster 254.

⁴⁹⁴ Polybius X.43-47.

using wood poles that would be raised and lowered from towers.⁴⁹⁵ One can spot depictions of such signal stations on Emperor Trajan's column in Rome, showing three watch towers along the Danube with signal torches flashing from the windows. There was also generally a pile of wood ready at the foot of each tower that was likely lit as an emergency beacon or distress signal if needed and was designed to make a high fire surpassing the vegetation. Trajan and other emperors used the system of Roman forts built along the frontiers to send messages, and the torch signs presumably made up an alphabet or semaphore signals. There are archeological remains of signal towers and platforms inside the Antonine Wall in Scotland, for example, and behind Hadrian's Wall in England, as well as in Syria along Roman roads, and along the Rhine and Danube forts. The signal stations were generally small and easily concealable in the surrounding landscape.⁴⁹⁶

The roads that crisscrossed the empire were similarly built as part of the same tight organizational discipline and quest for effectiveness that pushed Rome's innovative military agenda. As Raymond Chevallier, expert on Roman roads, explains, the Romans built roads wherever they expanded, starting with Italy around Rome. Those building them were most often consuls or provincial governors in charge of conquering armies, so the roads bear their name, like Via Appia, for example, the first road South of Rome to Capua, built in 312 B.C. by consul Appius Claudius Caecus. The Romans then built roads to the North of Rome, with the Via Flaminia (220 B.C.) along the Tiber River and the Via Aemilia (187 B.C.) along the Adriatic, all the way into Northern Italy, Gaul and the Alps, Spain and the farther provinces as the Roman conquests progressed. That was a revolutionary transformation, as the region was soon crisscrossed with a network of roads. In fact, Chevallier contends, "before Romanization, the

⁴⁹⁵ Vegetius III.5.

⁴⁹⁶ Webster 255-6.

road system [in those places] ... remained in a fragmentary state. The Romans were ... responsible for straightening roads and linking them with each other in addition to carrying out more permanent engineering work." The Romans built roads originally with a clear military purpose in mind, which was not only to ensure the chain of command with their always farther front lines of conquest, but also to be able to move troops rapidly toward the faraway province to quell rebellions or barbarian assaults on the periphery of the growing empire.⁴⁹⁷

The Romans' engineering genius led them to spread their roadwork even to the most inhabitable places to connect all of the empire. The African provinces were particularly challenging because the land combined high mountains, arid steppes and desert. Still, the Romans were able to build "incredible ramifications of the network of imperial roads" there too, North Africa specialist Jean Baradez argues, and they formed a "gigantic cobweb of roads encircling the mountain ranges and cutting through the torrid and desolated steppes zones." When necessary they even built canals, like in Egypt, to ease troop movements. Most importantly, they also maintained the roads, even the most difficult roads to reach such as the desert highways of African and the Middle East, providing water tanks and overnight halts and posts. The Romans even marked each course with milestone markers, generally rocks but when necessary, reeds stuck in the ground visible in sandstorms, like in the Sinai, for example. Caravan trails snaking through the desert in Egypt were marked with stone cairns strategically leading to oases and had special posts set up for guarding key areas like pass entrances or narrow valleys. All Roman roads were designed with great care and were strictly organized to accelerate

⁴⁹⁷ Raymond Chevallier, *Roman Roads* (Berkeley, CA: University of California Press, 1976), 132-5, 145-8; Amit 210.

Roman communication and troop movement to and from the frontier areas for information, resupplying, and extracting resources.⁴⁹⁸

Along with the Roman roads, the *Cursus Publicus*, or Roman Imperial Post, was another innovation that showcased the Romans' watchful discipline and organizational skills. The Imperial Post was a state-run courier service set up by Augustus, possibly inspired by a similar system that existed in Persia earlier, to transport imperial messages, officials, and official documents, as well as tax money across the empire using the Roman roads. Its purpose was not only to dispatch orders and information, but also to serve as an instrument for the central supervision of governors and other provincial officials, British anthropologist A.M. Ramsay explains. A number of major Post stations were arranged along the road system. In addition, thousands of relay stations, not unlike the ones later reinvented by the Mongols, were set up 18 to 22 miles apart along the Roman roads, providing horses, carriages, fodder, food and lodging for messengers and officials, who needed a special permit signed by the Emperor, the diplomatic passport tables, to use the free transportation service. The stations were supplied by the local town where they were located. The Imperial Post operated two parallel messenger services: a regular service with oxen and carts, presumably for officials, tax revenue transports, and low-priority messages, and a fast service with horses. The Imperial Post was operational in Italy and most of the developed provinces by the early Principate. There was only one Roman road equipped with stations in Egypt and Asia Minor, however.⁴⁹⁹

The Post fulfilled a crucial military purpose: it enabled the legions to ask for reinforcement fast and to send field reports to Rome in one way, and it enabled Rome to send instructions and updates to the legions. The mounted messengers could travel at a pace of about

⁴⁹⁸ Baradez 30.

⁴⁹⁹ Chevallier 181-4, 186, 189; Amit 210-211.

40 to 65 miles a day, depending on the nature of the message, the number of hours of daylight, the weather, terrain, and how detailed the area's Post system was, with the possibility of an even faster delivery in case of emergency. In that case the messenger would attach a feather to his spear to indicate urgency and bad news, and a laurel to indicate good news. The average speed covered was about 4.6 miles an hour, which seems slow to a modern person but was as fast as possible in antiquity and was definitely a major improvement over the earlier Republic, when communications depended mostly on private entrepreneurs and ceased entirely in the winter.⁵⁰⁰

One last area where strict discipline and tight organization clearly led the Romans to surpass their competitors was with the army's support functions, and principally logistics, which Roman roads considerably improved. The effective defense of a growing empire like Rome required that troops be always well supplied and reinforced all through the empire and necessitated an easy and fast flow of troops and supplies wherever it was needed. The Romans developed a "highly sophisticated," flawlessly organized logistics system, Ferrill contends:

The Roman system of supplying food, arms, and armor was as advanced as any system would be down to the 19th century of our era, and it served, alongside the monuments of Roman military engineering, as an instrument of psychological warfare, intimidating Rome's ... enemies by its obvious superiority over anything they could ever hope to achieve.⁵⁰¹

Though developments were gradual, by the early Principate, arms were manufactured partly by private contractors but mostly by the legions themselves, which had armories in their camps where most weapons were crafted and repaired. The legions were in effect almost self-sufficient, Vegetius writes. They produced not only their own weapons, but also their armor, transports,

⁵⁰⁰ A.M. Ramsay, "The Speed of the Roman Imperial Post." *The Journal of Roman Studies* Vol. 15 (1925), 60-61, 66, 68-69, 71, 73-74.

⁵⁰¹ Ferrill 27-29.

carpentry, etc., as they employed a number of artisans and craftsmen and other experts in their ranks. The Roman army also directed most of its own food supply. It employed centrally controlled grain suppliers—who often simultaneously worked as spies, as mentioned above—with regional depots in the various provinces and could easily furnish food on short order. No Roman enemy had such a highly institutionalized and comprehensively organized logistical support system, particularly not Rome's barbarian opponents. As a result, Roman soldiers, even when fighting at the edges of the growing empire, were well-fed, -housed, and -supplied with arms, and could focus on training and improving their military skills. That was all a result of prior learning. In the early Republic there are accounts of Roman soldiers forced to forage for their own foods outside their camps, sometimes with disastrous military consequences.⁵⁰²

On a day-to-day basis, the Roman army's needs varied considerably depending on its location. If it had to make deep incursions into enemy territory, it had to carry most of what the soldiers needed, including equipment, food, and animals, in their baggage train, in order to be self-sufficient. If it operated on the periphery of the empire, the army could afford to carry much less and thus be faster, and rely instead on convoys from the nearest supply bases to resupply them. If it established a long-term garrison in a newly acquired but faraway territory, like in England for example, it made more sense to contract locals and even cultivate some food items themselves. Armies often brought herds of animals with them on campaign for meat, and at times also confiscated animals from enemy or local populations or were supplied by allies. Besides meat, rations generally also included bread or biscuits, wine, and oil. Records only exist for the later Roman empire, once Rome was already a hegemon, but they indicate daily rations of 3 lbs. of bread, 1 lb. of meat, 2 pints of wine, and 0.1 pint of oil per soldier. A legion of 5,000 men would accordingly consume 100 bushels of grain per day and slaughter 12.5 oxen, 120

⁵⁰² Vegetius II.11.

sheep, or 38 pigs (or a combination of those) every day—the need for disciplined, reliable logistical support is thus apparent. Without a state-of-the-art logistical organization, the Roman army would have been paralyzed.⁵⁰³

In addition to logistics, a few other support functions were vital to the Roman army and therefore also benefited from the same attention and detailed organization. First, the Roman army was doted with surprisingly modern medical services. Although the Romans were traditionally keen on superstitious rites and incantations, they also imported the medical practices of the Greeks, which were much more advanced. The Romans adopted and expanded the Greek focus on healthy diet, exercise, hygiene, and natural drugs. They also mastered simple but vital operations like wound stitching and amputation. In fact, Webster point out, the Romans "recognized by the early empire that a well-organized and competent medical service was essential not only for the general well-being of the troops but for the prompt and efficient treatment of battle casualties," and as a result made it a priority to offer the best possible medical care to the troops and attract the best doctors into the army. The most qualified doctors were given the rank of centurions, and the attractive status, pay, and security as well as the opportunity for hands-on experience, particularly as military surgeons, led many of the best doctors of Rome to join the army, guaranteeing an excellent medical care in the military hospitals. Each legionary fortress and some of the larger auxiliary forts had a military hospital, built with double walls designed to shield the wounded and sick against noise and extreme temperatures. In addition, to support the medical philosophy highlighted above, the Roman army "ranked highly" in personal hygiene and effective sanitation. Every permanent camp had latrines with complete drain systems and bath houses with rooms at different temperatures and humidity levels. Before constructing a camp the Romans would ensure that large supplies of fresh water were available

⁵⁰³ Goldsworthy, *The Roman Army at War*, 287-295.

nearby. Camps entailed elaborate water transport systems to supply the latrines and bath houses, using gravity to flow water via aqueducts into tanks and reservoirs and then through pipes to the different parts of the camp. The Roman engineers ensured that drains would discharge lower than watering points for humans and animals, and if that was not possible, the Romans also developed innovative arrangements akin to modern septic tanks. In the end the Roman soldier was in much better shape and much better cared for than any of its rivals.⁵⁰⁴

Second, the Roman army also developed a specialized and highly disciplined intelligence and reconnaissance function. Special units of "exploratores" were responsible for collecting intelligence. They alternatively used local populations in friendly or neutral territory and enemy captives and deserters to extract information, but their most reliable tools were scouts and patrols, which kept an eye on the enemy day and night and were trained to become inconspicuous and get very close to or even mingle with their targets. Espionage was facilitated by the fact that many Roman enemies, particularly barbarian tribes like the Gauls or Germans, frequently moved in disorderly ways, unlike the Romans, so that their movements were particularly easy to pick up. The Romans were also expert at observing their natural surrounding and interpreting natural signs. For example, they could easily determine the size, distance, and nature (cavalry or infantry) of an enemy contingent by observing dust clouds. Even though they did not have modern topographical maps, the Romans understood the importance of terrain so part of reconnaissance consisted in assessing the terrain and determining where the most tactically advantageous ground to fight was. High commanders sometimes participated in reconnaissance trips themselves to better assess the weaknesses of enemy defenses, especially

⁵⁰⁴ Webster 257-262.

when a siege assault was planned. Caesar personally made two reconnaissance observations of Gergovia before launching the attack, for instance.⁵⁰⁵

The Romans considered reconnaissance and intelligence crucial for victory, a concept that went hand in hand with their forward-thinking focus on planning. A trademark of the Roman army was to never hastily engage in a conflict, but calculate all the details, including the best terrain, alternative approaches, and supply and escape routes. In fact, the few near-disasters against Carthage in the Second Punic War can in large part be blamed on the Romans' unusual failure to follow those tenets and be tempted by hasty heroism. Their traditional emphasis on cautious, level-headed, disciplined planning, reinforced by careful gathering of intelligence stands in stark contrast to many of their opponents' often counterproductive military impulsiveness, which the Romans frequently sought to exploit with deceptive tactics. As Frontinus explains, Roman intelligence units were often sent on patrol to exploit the enemy's spontaneity and disseminate wrong information to prompt its action. Those units were particularly good at using "stratagems ... [to] convinc[e] the enemy that you plan to do one thing, before doing another," Frontinus writes.⁵⁰⁶

Thus, the Romans created an innovative military apparatus, which propelled them to surpass their competitors. They progressively improved both the tactical and structural organization of their army and reinforced it with a particularly extensive system of army discipline. But in addition to its unique organization, the Roman army also became unbeatable because of its superior strategic and tactical skills.

⁵⁰⁵ Goldsworthy, *The Roman Army at War*, 125-128.

⁵⁰⁶ Sextus Julius Frontinus, *Stratagemata (The Stratagems)*, transl. by R.I. Ireland (Leipzig: Teubner, 1990), II.8; Webster 23; Goldsworthy, *The Roman Army at War*, 128-131.

b. Roman Strategies and Tactics (IV5.2)

While a highly efficient organization certainly contributes to the superiority of an army, it is not sufficient to guarantee success over its adversaries. But the Romans were able to couple their innovative, disciplined organization with similarly inventive strategies and tactics, enhanced with the best techniques and weapons they were able to copy from their competitors and further improve themselves.

1. Superior Strategy and Grand-Strategy

The Romans succeeded in their hegemonic bid because they understood the strategic and versatile nature of military power. They used their military power for political purposes, instead of fighting their neighbors for the sole sake of military conquest and without larger plan as most of the Mediterranean states that surrounded them did. Roman expansion was always part of a larger strategic ensemble. The Romans were thus an early example of Clausewitzian use of force, where the military was channeled as a tool of political warfare. Because the political purpose always came first, the Romans did not shy away from using the military power sparingly and indirectly, as long as it enabled them to reach their political goals, which earned them more popularity and support than their opponents' frequent arbitrary brutality. As Luttwak points out, "the Romans ... learned that the most desirable use of military power was not military at all, but political; and indeed they conquered the entire Hellenistic world [i.e., the Eastern Mediterranean] with few battles and much coercive diplomacy." This is evident by the Romans' constant use of deterrence, compellence, and alliances. Luttwak further argues that this understanding of the primacy of the political purpose led the Romans to seek to "conserve force," avoid reckless advance, and always carefully plan military campaigns to limit risks and surprises. The Romans

favored coherent campaigns integrating troop deployments, defenses, road networks, signaling, etc. Many of Rome's competitors, particularly in the Greek world, were much more reckless. The "sophistication of Roman strategy" resulted in the superior quality and resilience of their army.⁵⁰⁷

Thus, from the mid-Republic into the first century A.D., Roman strategy was designed to maximize hegemonic reach, with every aspect meticulously focused on expansion. This was the clear priority of the Julio-Claudian system of client-states and mobile legions, which involved a "combination of diplomacy, direct force, and fixed infrastructure" coupled with specific "operational methods" reflecting this goal. There remains some debate in the literature as to whether Rome's strategy was primarily offensive or defensive. Adcock, for example, argues that at first the Roman Republic was clearly defensive, as it rarely made the first move in conquering Italian and Greek neighbors and instead expanded by responding to security threats and neighbors' calls for help, but then became caught up in an expansionist cycle as those early operations had "widened the horizon of Roman strategy."⁵⁰⁸ Goldsworthy retorts that such analyses are misleading because they assume that Rome sought geographic expansion. Rather, he argues, Rome fought wars not to control territory but to control political entities, such as tribes or states, and thus, the Roman army favored offensive postures that could break the enemy's political will and bring it quickly under Roman influence. "The Roman emphasis on the offensive in all forms of warfare [including economic, siegecraft, high intensity, and low intensity warfare] was [an] attempt to dominate the enemy's collective willpower and suggested the inevitability of Roman victory," Goldsworthy adds.⁵⁰⁹ His analysis is much more in par with Rome's focus on political goals, and also explains Rome's novel development of the concept of total war. Starting with the Second Punic War, instead of focusing on one main theatre of

⁵⁰⁷ Luttwak 1-4.

⁵⁰⁸ Adcock 51; Luttwak 4-5.

⁵⁰⁹ Goldsworthy, *The Roman Army at War*, 76-79, 114-115.

operations as had previously been the case, the Romans began simultaneously keeping troops and supplies to operate in several theatres at a time. Again, this required considerable planning and an increased reliance on coercion in addition to brute force. "It was the success of this strategy and the brilliant victories in Spain which prevented Hannibal from completing his work in Italy and eventually sealed his fate," Webster points out.⁵¹⁰

Once Rome began expanding, one of the most crucial aspects of its grand-strategy became how to protect the empire's ever-increasing borders. Here again, Rome showed innovative skills. Previous empires and contemporary great powers had faced tremendous difficulties keeping their possessions after conquest, and the vaster the territory the more arduous the task of preventing border flare-ups. By the late Republic the square footage of the empire had reach such extents that it became evident the Roman army could not guard every inch of border—a grand strategy known as perimeter defense—as most past empires had done. Thus, Rome came up with a novel grand strategy. The solution resided with a flexible, peripheral defense strategy: Rome relied primarily on client-states and outside provinces to safeguard their own borders and act as buffers, but a mobile Roman army could supplant the locals and intervene both to secure borders and ensure domestic security when necessary thanks to Rome's state-of-the-art infrastructure. During this period the Roman army had no permanent headquarters but was sent on mission wherever it was needed. "By virtually eliminating the burden of maintaining continuous frontier defenses, the net, 'disposable' military power generated by the [Roman] forces was maximized," Luttwak concludes. This was particularly beneficial to Roman conquest since "the total military power that others could perceive as being

⁵¹⁰ Webster 10.

available to Rome for offensive use—and that could therefore be put to political advantage by diplomatic means—was also maximized."⁵¹¹

But by the early Principate, as the empire had grown even further, it became clear that the peripheral defense based on clients and provinces and occasional Roman intervention alone was insufficient in the most unstable border regions, primarily those adjacent to barbarian lands like Germany, Parthia, Scotland, and North Africa, which suffered constant raids by tribes across the border. Again, Rome showed great adaptability to the circumstances and created a hybrid defense system. Starting with the end of Augustus' rule and continuing under his successor, Rome fundamentally kept its client-based peripheral defense system for most of the empire but gradually introduced a dose of perimeter defense in those volatile regions. Then, Hadrian essentially stopped the expansion of the empire, conscious that further expansion was not sustainable defensively, and focused on consolidating what was already acquired and reinforcing the limited perimeter defense against the barbarians. Essentially, along the Rhine and Danube, as well as around Parthia and in England, the Romans began building permanent garrisons and placed their legions on the borders in those crucial areas. Tacitus mentions, for example, that in 23 A.D. under Emperor Tiberius, there were 8 legions on the Rhine, 4 on the Danube, 2 in Dalmatia, 4 in Syria, 3 in Spain, which experienced frequent revolts, 2 in Egypt, and 2 in the province of Africa, where there were also occasional tribal uprisings.⁵¹²

Those legions, though residing in permanent garrisons, kept a dual purpose of internal and external security and a mobile intent and could easily be moved to support other areas whose defense was left to a province or client-state if necessary. When on the move, they could use the Roman roads to travel fast and the *Cursus Publicus* stations to get supplies. The establishment of

⁵¹¹ Luttwak 19.

⁵¹² Ferrill 11-13, 15.

more permanent bases, which began under Augustus, was also advantageous in that it enabled the Romans to standardize and perfect their supply routines and also to rely less on confiscating foods from the local population, a practice that had created resentment in the past. This new strategy was only possible because the empire possessed the best road network that enabled information to flow between the legions and Rome and because Rome was about to reach hegemony and major expansion was therefore not needed anymore.⁵¹³

There is clear archeological evidence of the switch to a hybrid grand strategy, as Emperors Augustus, Trajan, and Hadrian built fortifications on the borders, or limes, to serve as the perimeter defense infrastructure for the legions assigned to the volatile regions. "By the ... 2nd century the Roman world was encased in an armor of stone or wooden walls or frontier roads which were garrisoned at intervals and guarded," Starr explains.⁵¹⁴ The best know is probably Hadrian's fortification system across Northern England from Carlisle to Newcastle, which was complete with a roughly 73 mile-long, 7 feet-thick and 15 feet-high stone wall from coast to coast, sixty fortresses set a regular intervals and an earthen rampart and road inside the wall and running parallel to it, the Vallum, to guard the new province of Britain against the Picts and their ancestors. Second-century Emperor Antoninus Pius later constructed a second, 39-mile long wall dotted with 26 forts about 100 miles north of Hadrian's wall, again from coast to coast between the Firth of Forth and the Firth of Clyde, as Rome conquered more English territory. Under the Judeo-Claudian and Flavian emperors the Romans also constructed advanced fortifications to keep out the Germanic and Eastern European tribes, using the natural barrier of the rivers as a border to keep out the barbarians. Along the Rhine, from Rotterdam to Basel, as well as between the Rhine and Danube, from Koblenz to Regensburg, the Romans built a first fortified line

⁵¹³ Amit 212-215.

⁵¹⁴ Starr 105-107.

composed of 176 forts. Then, along the Danube and all the way to the Black Sea, from Regensburg to Budapest to Belgrade, the Romans set up a second fortified line with another 220 forts. Webster thus identifies a total of 396 defensive positions, mainly forts, along the Northeastern limes of the empire, based on archeological remains. In North Africa, and to a certain extent in Mesopotamia, where the Romans battled the neighboring Parthians, there is similar evidence of fortifications, but with much fewer forts.⁵¹⁵

The fortifications along the empire's most volatile border, like Hadrian's wall, were not meant solely as a defensive wall to station soldiers on, British historian R.G. Collingwood argues. They also played the vital modern role of frontier marker as an "obstacle to smugglers, or robbers, or other undesirables." The function of controlling who got in and out of Roman territory is clear when considering that in those regions one could only cross into the empire at certain checkpoints, presumably after showing credentials.⁵¹⁶ In addition, French archeologist André Piganiol explains, a Vallum like the one running along the inside of Hadrian's Wall in England also served as a military road to convey army personnel and supplies, and was for that purpose protected on both sides by mounds of earth and a ditch to shield the army traveling on the road.⁵¹⁷ The Romans were thus particularly resourceful and inventive with their hybrid defense strategy, and, as Baradez emphasizes, revolutionized the traditional perimeter defense in the troubled borders where they implemented it. While perimeter defense was generally applied horizontally, following the thin line of the border, under the Empire the Romans gradually reformed it to embrace a more vertical concept of border, focusing on depth with a network of roads and forts that ran along the border rather than simply on a line.⁵¹⁸

⁵¹⁵ Webster 28, 48-49, 52-53, 60-63; Ferrill 20-23.

⁵¹⁶ R.G. Collingwood, "The Purpose of the Roman Wall." *The Vasculum* Vol. 8 (1920), 4-9.

⁵¹⁷ André Piganiol, "La Notion de Limes." *International Congress of Roman Frontier Studies* 5th (1961), 119-122.

⁵¹⁸ Baradez 28-30.

Overall, despite minor domestic revolts in some provinces, the early empire had very little need to maintain internal security by brute force, eliminating the need for a central reserve of troops, which could thus be assigned to safeguard the empire's periphery. The hybrid defense strategy enabled the empire to ensure its security while keeping its troop numbers under control—roughly around 28 legions (about 140,000 men) and 2 naval bases (about 20,000 sailors) from the early Principate on—resulting in a very cost-efficient force. The Romans' incremental, inventive grand-strategy paid off, as the end of Augustus' reign inaugurated over two centuries of Pax Romana during which the empire thrived in peace.⁵¹⁹

2. Superior Tactics

In addition to their forward-thinking strategies, the Romans also displayed tactical inventiveness. Again, their superiority was not immediate but a result of experience and constant attention to improving and learning from their mistakes. As Polybius points out, the Romans made costly tactical mistakes early on against Carthage because it was the first time they battled such a powerful opponent, but they learned quickly. For example, while stationed in Western Sicily during the First Punic War, Roman commanders sent their troops to harvest grain in the countryside outside their camp. The Carthaginians took notice and attacked the Roman camp while it was half empty. Similarly, after defeating Carthage in the siege of Heraclea in Southern Sicily, the Roman troops were too excited and "failed to keep watch as strictly as usual," thinking they were safe since they had just routed the Carthaginians. Hannibal seized the opportunity, broke out of the besieged city with a few men and was able to fill the Roman

⁵¹⁹ Amit 219; Ferrill 1-3, 9-10.

entrenchment with straw-filled baskets unobserved. Fortunately for the Romans the damage was not serious but they learned from both mistakes and "took stricter precautions" in the future.⁵²⁰

Roman tactical prowess transpired through a variety of techniques specific to the Roman army. Roman commanders as a rule thrived to be on the offensive on the field, thought not recklessly, from very early on, to take advantage of movement and surprise and set their own pace to the battle. Already in the War of Sentinum (302-293 B.C.), the Romans are described as always on the offensive, while their opponents the Etruscans were defensive, even immobile, in siege behind their fortified cities and towns, abandoning the whole countryside to the Romans. Polybius argues that domestic politics, at least during the Republic, had much to do with that constantly offensive posture. The two Roman consuls who headed the army were only elected for a one year term each, so they were naturally pushed to excel to secure their own reputation, obtain sought-after honors from the Senate, and perhaps be reelected. Because of this political system, consuls were encouraged to take the offensive in campaigns, and push their armies forward constantly. On a few occasions it drew the consuls too far, for example at the Battle of Trebbia in 218 B.C., and they would rush into a battle for fame before being ready, but over the course of Rome's rise, it is safe to say that the offensive incentive served as a driving force for victory and a factor of heightened tactical efficiency in the Roman army.⁵²¹ Even when it did have to adopt a defensive posture, the Roman army was never static, like many of its enemies; instead, defense was used to ready itself for counter-attack. Using defense tactically to position oneself favorably for the next attack is again a very Clausewitzian notion before its time. As

⁵²⁰ Polybius I.17, 19.

⁵²¹ Eckstein 129; Polybius III.70, 75.

Goldsworthy concludes, the "emphasis on adopting and maintaining the offensive apparent in Roman strategy was thus equally dominant in the army's tactics."⁵²²

Concretely, the Romans' most frequent offensive battle goal was to break through the enemy line, coupled with a secondary attack on the flank to surround the enemy and simultaneously cut its fleeing options. The Roman army's action of choice was thus akin to a breakthrough and hold, an eminently modern tactic that is still the primary choice of many current armies, as Stephen Biddle mentions, and was only possible because the Romans always kept a large proportion of its troops in reserve positions, so that they always had troops left over for holding and exploiting once the breakthrough was successful.⁵²³ This resulted in very "fluid," fast-paced battles. In fact, I could find only one instance where the Roman army was deployed in one single, thin line: a battle in 46 B.C. in Africa, where Caesar was so grossly outnumbered that he had no other choice. Besides the political incentives for commanders, the Romans also likely favored offensive tactics because their army was often less numerous than that of the enemy, despite a general assumption of the contrary. Therefore they always preferred hills and higher ground that offered both protection and the advantage of charging downhill, with terrain not too rugged to allow optimal maneuver. The Roman army also always reinforced their flanks due to their generally smaller numbers, in order to avoid encirclement, and would often choose terrain that secured the flanks against an obstacle, such as defiles, or in-between woods, and if necessary the troops would add trenches on their sides to reinforce their natural positions. In open terrain the flanks were generally secured by the cavalry because it was more mobile. Overall, the specific arrangement of the troops varied considerably depending on the circumstances, and the

⁵²² Goldsworthy, *The Roman Army at War*, 141.

⁵²³ Biddle 40-44.

Roman army was trained to adapt to each tactical situation and evolve into new formations quickly, as mentioned above, in order to be least predictable to the enemy.⁵²⁴

The Romans' easy adaptation to different tactical situations is a "sign of the flexibility of the Roman system, that using essentially the same mixture of forces, they could develop a method of successfully opposing very different types of armies," Goldsworthy explains. In fact, the Romans managed to become proficient "in every scale and type of warfare" and went on to "beat the enemy at its own type of warfare." The Romans became victorious not only in the high-intensity wars they were used to conduct, but also against more atypical enemies engaging in low-intensity, guerilla-type warfare like the German, African, and Parthian tribes. For example, in 17-24 A.D. the Roman army was sent to quash a rebellion that raged in the provinces of Africa from Mauritania to Cirta in modern Algeria. The rebellion was serious because the rebel bands, under tribal leader Tacfarinas, were evidently organized militarily and had good equipment. They conducted razzias in both the Northern mountains and Southern oases and steppes, taking advantage of the terrain, appearing and disappearing with lightning speed in regions controlled by Rome.⁵²⁵

The Roman response was surprisingly close to today's counterinsurgency warfare. The Legion III stationed along the African limes fractioned into small units, with mobile columns for rapid response, and set up small camps all along the limes from where the guerillas came to serve as camouflaged bases of operations for the units and increase their mobility. The units selected for the job were mostly auxiliary troops with a reputation for mobility and a familiarity with the terrain: Thracians, Sardignians, Hungarians, Dalmatians, Maures, and Musulmanes, who were all light-infantrymen, lancers, archers, scouts, and expert riders, as well as Chaldeans from Syria

⁵²⁴ Goldsworthy, *The Roman Army at War*, 141-2.

⁵²⁵ Ferrill 8-10; Goldsworthy, *The Roman Army at War*, 135-7, 286; Baradez 28-29.

who possessed tailored equipment and steppes horses and were specialized in desert combat. As Baradez points out, "we can see with what care Rome picked the men defending its African provinces, choosing those it estimated best adapted to the type of enemies they had to combat and to the terrain they had to defend." In addition, the Romans relied on non-rebel local militias, directed by Roman officers, to support the auxiliary units. In a few cases the Romans even armed the local population so they could defend themselves against the rebels. All those tactics seem remarkably similar to those used in modern counterinsurgency warfare. And although they took some time to succeed, the Romans managed to bring the African rebellion under control.⁵²⁶

Rome's tactical skills proved so proficient that Rome frequently won despite being a technological underdog. One of the best examples is probably Rome's decisive naval defeat of Carthage at the 256 B.C. Battle of Ecnomus in the First Punic War. Aware of their technological inferiority, the Romans knew they had to think up a superior tactical plan to stand a chance. "Their enemies possessed the faster vessels, and they therefore took great pains to devise a formation that would remain unbroken and would be difficult to attack," Polybius explains. To limit its vulnerability, the Roman fleet spread as a triangle-shaped wedge in tight formation, facing Carthage's traditional single-file formation. Rome's tight, deep formation scored a massive victory, as the Roman ships were able break through the Carthaginian line and encircle a large part of the enemy fleet.⁵²⁷

Besides surmounting its technological weaknesses, tactical skills also helped Rome trick their opponents. The Romans recurrently used false attacks and withdrawals and other diversionary maneuvers with great skill to create surprise, confuse the adversary or lead it to waste supplies or send troops in the wrong place when Rome was in numerical inferiority, for

⁵²⁶ Baradez 31-33.

⁵²⁷ Polybius I.26-28.

example, and by diversifying those tricks the Romans were able to win numerous battles. Unlike the Romans, who took advantage of every possible tactical opportunity to surpass their adversaries, Rome's enemies sometimes made large tactical mistakes. One example is Carthage's decision, in the 255 B.C. Battle of Adys in North Africa, to position its troops on high ground, normally a stronger tactical position but in this case a very poor choice for Carthage since it was inaccessible to most of its force. Indeed, Carthage's main strength lied in its cavalry and elephants, neither of which could maneuver on rocky hills. Basically the Carthaginians just "show[ed] their enemies how to best attack them," Polybius writes. Their elephants and cavalry could not participate in the battle. Of course, the Romans went on the offensive as soon as they realized that "their enemies' most effective and formidable arm had been rendered useless to them by their choice of ground." They crushed the Carthaginian so severely that Carthage sued for peace following the battle.⁵²⁸

c. Proficient Technological Adaptation (IV5.1)

As hinted above, Rome's strength came primarily from their tactical and organizational skills, rather than from their technological prowess alone. Yet, the Romans never let themselves be surpassed by technology (IV5.1). What they could not invent themselves they just copied from their most proficient competitor, and most of the time were able to dramatically improve upon the weapons and techniques borrowed from others to make them even more successful than the originals. Weaponry is perhaps the best example of the Romans' technological inventiveness. The standard legionary uniform included a bronze helmet, a mail shirt, a sword, a shield, a spear, a heavy javelin (pilum), and a dagger. But those weapons and armor pieces evolved considerably

⁵²⁸ Ibid., I.30, 33-34, IX.4, X.11-16, XV.9.

throughout Rome's rise, as Rome sought to improve them. This evolution clearly shows influence and learning from the civilizations Rome conquered, as the Romans embraced better technology whenever they encountered it, and simultaneously innovated themselves—engaging, in other words, in creative borrowing. In the end, the Roman army was better equipped than all of its competitors. "There is little doubt that the Romans always showed a readiness to adopt and to borrow ideas from other people, a practice they were not always ready to admit," Webster acknowledges about Roman arms.⁵²⁹

In the 6th and 5th century B.C. already, the Romans copied the more advanced technology of the Central Italians, mostly of the Etruscans, and adopted their short sword, pilum, and rounded helmet. In the 4th and 3rd centuries, the Romans were more influenced by the Greeks and their armor styles, while at the same time Italic weapons persisted, "a manifestation of that double character, both conservative and innovative," distinctive of Rome. Couissin further explains that "Rome ... adopted the foreign arms it deemed most efficient, but retained, while perfecting them, many of its own national weapons." In the 2nd century B.C., as the rising empire grew, Roman technology fell under Gallic and Spanish influence. This was a major evolution, as the pilum became the Romans' main weapon, paired with a sword that was now much larger, based on the Iberic sword. The Romans adopted the Iberic swords after the Second Punic War, after realizing that their shorter, curve-bladed Greco-Italian sword could not parry the swords of the Spaniards and left them at a serious disadvantage. Simultaneously the Romans abandoned the Greeks' heavy bronze protections and adopted the lighter, more functional Celtic armor types. The mixture of indigenous, Italic, and Hellenistic styles created an increasingly unique equipment, specifically Roman. In the 1st century B.C., the Romans slowly abandoned Greek and primitive Italian material and borrowed progressively more from the Gauls and the Spaniards,

⁵²⁹ Webster 8; Ferrill 5-6; Goldsworthy, *The Roman Army at War*, 183-4, 246.

but at the same time they refined those weapons more and more, rendering them increasingly Roman, until they became standardized under Caesar as a much lighter equipment, designed to facilitate Rome's offensive tactical posture. Thus, Couissin concludes, "although it was composed primarily of borrowed foreign elements, [Roman armament] was eminently Roman ... by its 'utilitarian' character that made both its unity and originality."⁵³⁰

The Romans' engineering ingenuity enabled them to copy foreign weapons and techniques and improve them to surpass the originals, even though they mostly did not start out with the best technology. Within the army the Romans employed a number of engineers, craftsmen, and experts who specialized in perfecting their technology and its use. Thus, while Rome's neighbors, both friends and foes, might have made more innovations of their own, the Romans' tactical superiority came from their ability to gather the best technology and techniques observed from the neighbors and integrate an enhanced version in their own army. This is particularly evident with regard to siege warfare. Roman experts worked around the clock on perfecting siege engines, and from the simple devices used by their competitors, soon developed an array of engines tailored to different purposes: the scorpion, a big mounted crossbow; the catapult, a modified crossbow with horizontal launch used to throw missiles like javelins; the ballista, an engine with a 45-degree launch throwing balls, rocks, and wood; the onager, a vertical launch engine with only one arm which was one of the Romans' biggest improvement; the *turre ambulatoria*, a powerful moving tower on rollers invented by Agrippa's engineers in ca. 30 B.C., with one or more drawbridges, a lower strip with battering rams and upper stories with their own siege engines that could hold archers and slingers; the *vinea*, a movable shed on rollers used for shield workmen or soldiers during siege operations, covered with rawhide against fire;

⁵³⁰ Couissin 175-6, 213, 224-232, 271-2, 351-2, own translation; Polybius III.114, VI.23; Goldsworthy, *The Complete Roman Army*, 29; Harkness xxix-xxxi.

and most impressive perhaps, the agger, a mount built with tree trunks, wood, rocks, earth and brush, of generally up to 500 feet long, 50-80 feet or 8-10 stories high, like a highway for soldiers and siege engines to reach the walls of a fortified hill town, with built-in shields to protect the men building it and the soldiers advancing on it.⁵³¹

The Romans famously used the most extreme agger during the First Jewish-Roman War in 70-73 A.D. to take the mountain fortress of Masada. Instead of risking a certainly doomed direct assault or a lengthy strangulation, Roman general and governor of Judea Lucius Flavius Silva simply instructed his engineers to build an agger, a ramp 675 feet long and 275 feet high constructed all the way up to the fortress, with a stone road on top so the infantry and artillery could reach the fortress walls. Another major Roman innovation was a grapnel mounted on a catapult on top a ship to catch and restrain another ship. It was singlehandedly responsible for many of Rome's naval victories during the Punic Wars and beyond. Bridges became another one of Rome's engineering feats: Caesar's men reportedly built a large trestle bridge over the Rhine in just ten days, while Trajan's chief engineer Apollodorus constructed the largest permanent bridge in antiquity over the Danube, 150 feet high, 60 feet wide, with stone piers set 170 feet apart. As Ferrill concludes, the Romans took warfare as far technologically as it was possible without gunpowder and motors. Roman siege engines were as excellent as siege engines would ever become, with the light-projectile dart-throwers reaching as far as 600 to 800 yards, more than the 500 yards of modern-day catapults. The Romans' technological superiority against all of their opponents fed into their tactical prowess; it enabled their offensive posture by allowing forward deployment thanks to their unsurpassed ability to build roads, bridges, and state-of-the-art

⁵³¹ Harkness lx-lxiii, lxv.

artillery in a heartbeat. Technologically, the Romans were "more sophisticated than modern armies would be until the 19th century."⁵³²

The best example of Rome's technological learning skills resides in their dramatic development of naval power during the First Punic War. The Romans had virtually no naval capabilities prior to the First Punic War. In order to transport their troops to Sicily for that war, they even had to borrow ships from the Tarentines and other coastal allies. But during the early stages of the war they suffered greatly from the Carthaginian fleet, who held naval supremacy in the region and raided the Italian coast while Rome was forced to watch helplessly. This led Rome to develop its own sea capabilities, even though it had no natural affinity with naval power and, according to Adcock, "to the Romans the sea was something incalculable, treacherous." Using the model of a Carthaginian ship confiscated while crossing the Sicilian straits, Rome built 100 quinqueremes and 20 triremes—warships propelled by five and three banks of oars, respectively, which were essentially the start of a massive shipbuilding program. As Polybius concedes, "they faced great difficulties because their shipwrights were completely inexperienced, ... yet ... this fact illustrates better than any other the extraordinary spirit and audacity of the Romans' decision." Despite lacking the proper resources and knowledge, the Romans pulled it through and had soon recruited and trained crews that were able to sail off within months. After some trial and error, not only did they catch up to the Carthaginians on the seas, they even innovated to surpass their rivals. Most importantly, the Romans supplemented the Carthaginian-style ships they had built with the *corvus* (raven), an iron-spiked gangplank with a grapple that eased accosting and boarding of an enemy ship, which surprised Carthage and gave the Romans their first naval victory against the Carthaginian fleet at Mylae off the Northeastern coast of Sicily in the summer of 260 B.C. And although the Roman fleet initially suffered defeats due to

⁵³² Josephus VII.VIII.1-7; Ferrill 29-32.

their seafaring inexperience—it was repeatedly destroyed by storms—the Romans persevered and rebuilt the fleet despite the cost, using funds lent by private citizens who knew they would be repaid only in case of victory. The First Punic War transformed Rome into a naval giant as the Romans took regional naval supremacy away from the Carthaginians.⁵³³

The development of the mailed cavalry is another example of Rome's pragmatic innovation skills based on a concept borrowed and adapted from others. The Romans' specialty was infantry, while the biggest innovators in the field of cavalry in antiquity were the Assyrians of Mesopotamia, who invented the hybrid horseman armed with a bow and arrow and a pike and thus fast and effective both from a distance and in one-on-one combat. The Romans were isolated from Middle Eastern military developments during their early rise and were therefore not confronted with state-of-the-art mailed cavalry until their encounter with Antiochus III at the Battle of Magnesia in 190 B.C. But while Antiochus's 3,000 cataphracts, or heavy scaled-armored horsemen, were no match for them, "the Romans would quickly adopt the equipment and tactics of their defeated opponent," Eadie points out. Their defeat at Carrhae in 53 B.C. by the cavalry-proficient Parthians was even more of a revelation. "The disaster at Carrhae served to demonstrate ... the vulnerability of the legions to cavalry attack and the necessity of strengthening the Roman cavalry, which had been neglected since the military reforms of Marius" because of Rome's faith in the invulnerability of the legions. The Roman cavalry, partly thanks to its reliance on auxiliary troops with foreign experience, was then quicker to pick up the latest innovations in cavalry like the full coat of mail and invent its own counter-tactics against

⁵³³ Polybius I.20-25, 58-62; Adcock 30, 33-4, 37; Webster 9; Harkness lxx; Goldsworthy, *The Complete Roman Army*, 34-42.

enemy mailed cavalry, such as tripping charging horses and focusing on their few unprotected, vulnerable parts.⁵³⁴

The Romans were also quick to embrace the elephant, a new and confusing animal, as a part of their military force. After suffering greatly from the enemy's elephants on their first expedition to Carthage, where many Romans were trampled to death as they were attempting to retreat, the Romans first developed new ways to paralyze the elephants to mitigate the advantage they gave the Carthaginians. At the Battle of Panormus in Sicily in June 250 B.C., the Romans succeeded in defeating the Carthaginian elephants with volleys of arrows, spears, and javelins. The panicked elephants turned against their own troops, trampling them, breaking the ranks, resulting in deaths, injuries, and widespread chaos among the Carthaginians. The Romans then successfully rounded up the remaining elephants and their Indian trainers, which enabled them to learn from them—they devised various ways to stop enemy elephants—and even used them in their own campaigns afterwards.⁵³⁵

Thus, the Roman army, by borrowing foreign concepts and technology and further improving upon them, developed the instruments to turn its superior tactical skills into military victory. Coupled with its incremental, innovative strategic thinking, and its uniquely efficient and disciplined organization, the Roman military naturally surpassed all of its competitors in the field and brought about Rome's hegemonic success. The Roman empire arose mainly because of the "phenomenal achievements" of the Roman army, Lintott points out, which "won the empire in the first place." But military skills alone are insufficient to build an empire of the stature and length of Rome, particularly to generate a cohesive empire despite its plural ethnic, religious, cultural, and linguistic identities. It is the peaceful mechanisms of empire-building on which

⁵³⁴ John W. Eadie, "The Development of Roman Mailed Cavalry." *The Journal of Roman Studies* Vol. 57 No. 1/2 (1967), 161, 163-6, 173.

⁵³⁵ Polybius I.40.

Rome was rooted—its progressive, inclusive economic, political, and social institutions—that fostered the allegiance of the people taken over by Rome's military conquests. As Lintott puts it, it is "clear that the empire was not held down merely by military force ... Roman rule, even if at many points unjust and inefficient, was not only tolerated but appreciated by many of its subjects."⁵³⁶

2. Non-military Skills and Innovation (IV6)

To counterbalance its frequent use of military force to expend the territorial boundaries of the empire, Rome simultaneously sought to create a favorable living environment where its subjects would want to live, thus redressing the duress of conquest and minimizing their propensity to revolt. Starting with a moderate geographic advantage (IV6.1), Rome engaged in vast empire-building projects that paved the way for long-term control over the region. More specifically, Rome undertook three ambitious, novel programs, developed throughout its rise and designed to create hospitable conditions that would attract the support of all subjects: a growth-oriented economy (IV6.2), a non-arbitrary yet little interfering political rule (IV6.3), and an inclusive social setting (IV6.4). Its geographic advantage remained moderate at first, grew significantly with Rome's mastery of the seas (IV6.1).

a. Geography (IV6.1)

Geography played a moderate role in Rome's rise. While the diversity and fertility of the land on the Italian peninsula provided a sound economic basis for hegemonic growth, Rome's location proved more problematic. Its central position in the region was first a liability for Rome

⁵³⁶ Lintott 14-15.

and almost led to its downfall before Rome was able to exploit it and turn it into an instrument of its later success.

The Italian peninsula provided the ideal backdrop for Rome's development a strong economic foundation. The city of Rome and its surroundings were a microcosm of the peninsula itself, with a mild climate, fertile soil, and combination of plains and hills that spurred agricultural production and turned the city into a center of commercial activity. Once Rome expanded throughout Italy, it benefited from a similar diversity of land. The most fertile areas, the Po River valley in the north and the western side of the peninsula around Rome and along Etruria, were richly irrigated by a network of navigable rivers that not only facilitated cultivation but also enabled the circulation of goods. The western plain was rich in mineral deposits, while the Po valley remained the most agriculturally productive Roman area of the Roman Empire. Polybius notes, for instance, that "the fertility of this region is not easy to convey in words" and that "in respect of size and fertility this plain surpasses any other in Europe with which I am acquainted." It "yields ... an abundance of corn" as well as a "huge production of millet." In addition, "a great proportion of ... [the] very large number of pigs [that] are slaughtered in Italy every year both for domestic consumption and to feed the army ... are reared on this plain." The woods that widely covered the plain also supplied acorn to feed the swine population. The Italian peninsula also featured a number of mountain ranges, among them the Alps in the north and the Apennines that split the peninsula from north to south, and their slopes provided hilly territory that was also valuable, being "not too rocky and possess[ing] a certain depth of soil." With the sea running along the entire length of the peninsula on both sides and supporting a large fishing industry, the peninsula was richly doted to become an economic powerhouse.⁵³⁷

⁵³⁷ Polybius II.14-15.

Despite its advantageous geographic composition, the Italian peninsula's location in the region proved at first a challenge for the growing Roman Republic. On first look it may seem that Rome enjoyed a particularly secure position in the Mediterranean region. Not only did it share only one land border with its neighbors, but that border also coincided with the formidable natural barrier of the Alps, the highest mountain range in Europe. In reality Rome's location was not as advantageous as it appeared during its early rise. While the Roman Republic sat in the middle of the region and could therefore easily access and trade with both the Eastern and the Western Mediterranean, its central and protruding location also meant that it faced potential enemies on all sides. The Mediterranean region was compact and held a number of experienced sea-faring nations, so that the water did not shield Rome's long coastal borders from attack. Roman coasts were particularly vulnerable prior to the Punic Wars, when the Republic had not yet acquired advanced naval capabilities, and coastal areas were recurrently raided by Greeks, Carthaginians, and pirates, who at times undermined Rome's conquest of Italy.

Rome's land border was not secure either. While the Alps may have appeared forbidding to the Romans, they failed to discourage the Gauls, who lived in the mountains and used them as a sanctuary from which they could launch attacks on Roman territory and where they could safely withdraw. The Gallic tribes conducted regular incursions into Rome's northern plains and remained a major source of concern for the Roman Republic for several centuries until they were finally conquered. The Alps similarly failed to stop Hannibal from reaching Italy and nearly destroying the rising hegemon. His achievement, defeating an obstacle the Romans had deemed insurmountable, took the rising hegemon entirely by surprise.⁵³⁸

Once Rome had developed significant naval capabilities and had achieved naval dominance in the Mediterranean in the aftermath of the Punic Wars, its central geographic

⁵³⁸ Prevas 4-6.

location turned from a threat into a strategic advantage to further its military and economic growth into all four directions. However, it was Rome's acquisition of naval supremacy that made a difference rather than its location, so we must conclude that overall, geography played only a peripheral role in Rome's success.

b. Economic Development (IV6.2)

The rising Roman Republic and Principate made considerable efforts to reinforce the economic potential of Rome and boost its growth, not only to generate additional revenue for itself but also to attract more supporters. As a result of those efforts, the Romans had a high standard of living, far higher than in the Middle Ages to follow, which was appealing to the people living under their rule. Thanks to their successful economic policies, the Romans were able to keep defense costs—and thus, tax rates—relatively low, which further stimulated economic activity. Ferrill estimates the military budget at roughly 500 million sesterces a year in the early empire, or about 30% of the total budget.⁵³⁹ The emergence of Rome generated a number of economic changes not only for Rome itself, but also for the region, that improved the economic prospects of the empire as a whole.

First, the growing political and military involvement of Rome in Mediterranean affairs propelled it to the role of champion and protector of trade in the entire region. Rome's soldiers and diplomatic envoys did not just obtain additional territory for the growing empire, they also policed the region and secured the routes of commercial exchange: they took care of local bullies who constantly attacked their neighbors and disrupted trade relations, they got rid of the pirates

⁵³⁹ Ferrill 34. While this might see like a high proportion in comparison to current defense budgets, it remains considerably lower than the military expenditure level of Western countries during World War II, for example, even as those countries benefited from a much higher overall economic output (and thus even higher gross military spending).

that infested Mediterranean waters and the brigands that looted the countryside and hampered exchange, and they provided improved means of communication and commercial infrastructure by building roads and canals. The rising hegemon "fostered trade and commerce ... and so promoted the well-being and happiness of its citizens and subjects," as classical trade scholar Martin P. Charlesworth points out. Unlike many aspiring hegemons, Rome did not use force to generate and retain the allegiance of its people. It understood that tyranny would not enable it to last, but developing the right environment for subjects to enrich themselves would. With the exception of the destructive Civil War years between the Republic and Principate, Rome's economic policy focused on promoting peaceful trade and prosperous exchange among its provinces, client-states and beyond.

For example, Augustus curbed resurging piracy in the Mediterranean, improved irrigation channels in Egypt, one of the main agricultural providers of the empire, passed a fair-competition law inflicting strict penalties for actions hindering the grain trade or scheming to artificially inflate grain prices. Nero even proposed the establishment of a free-trade zone throughout the whole empire, though the project never came to fruition due to the opposition of some of his advisors. The Roman government also recurrently reached out to the faraway regions of Arabia, Asia, and Africa to open new trade routes and bring their riches to the West. Augustus and his successors instructed geographers to draw large maps of the world, demonstrating their awareness of the economic opportunity farther regions represented. Merchants and traders were encouraged to move quickly to the new provinces conquered by Rome to establish commercial relations. Soldiers on the periphery were also enticed to support trade, by bringing local goods back home and thus spurring demand for those goods. Markets were often held outside the camps set up by the frontier, and whole frontier towns were sometimes born out of that exchange

(Cologne and Augsburg in Germany, for example). Commerce was further stimulated by veterans who received land grants in the provinces for their retirement; many veterans became merchants themselves rather than farmers as commerce was highly profitable. In fact, Roman traders spread everywhere; there are reports of some reaching as far Aden and Ceylon.⁵⁴⁰

Second, the rise of Rome also propelled Rome's domestic economy to new heights. Before the conquest began, the Roman economy consisted mainly of small-scale, self-sufficient agriculture. But the conquest radically transformed that landscape. After the first military victories, retired consuls and praetors, who were forbidden by law to make carriers in commerce or become state contractors, began investing the wealth they accumulated in war into the land. Starting around the Second Punic War, they brought about a major revolution in Italy, buying out the small, independent, family-owned farms and developing massive centralized farms mostly run by slaves the former generals had brought back from their campaigns. The transformation was accelerated by the hardship resulting from long conscription during the war with Carthage, coupled with Hannibal's brutal destruction of the countryside that put many small farms out of business. By the mid-2nd century B.C., Rome was experiencing a major jump in productivity with vast plantations springing up all over Italy. The economic boom continued as the multiplying campaigns generated further wealth, and after an interruption during the Civil Wars, the boom was renewed thanks to the stability of the new imperial system of government, the business freedom and little involvement of the imperial and local administrations in business transaction, and the territorial expansion of the first century A.D. that offered new markets and opportunities within the empire.⁵⁴¹

⁵⁴⁰ Martin Percival Charlesworth, *Trade-Routes and Commerce of the Roman Empire* (Chicago: Ares Publishers Inc., 2nd Ed., 1926), xii-xiv, 227-8, 232-5.

⁵⁴¹ Starr 51-52, 116.

As its size expanded the empire developed its own productive geographical specialization and division of labor. By the end of the Republic, the Eastern part of the empire had become proficient in industrial and manufactured goods, mainly cloaks, garments, rugs, carpets, pottery, glass, perfumes, cosmetics, jewelry, spices, ornaments, etc., while the Western part had turned into the raw material warehouse of the empire, thanks to its mineral deposits, gold, silver, copper, lead, tin, and iron mines, as well as wheat fields. Yet the search for new profits kept prompting new opportunities, and by the 1st and 2nd centuries A.D., the West also developed its own, high-end manufacturing and services base as mine owners accumulated wealth of their own and began hiring Eastern scholars, doctors, artists, sculptors, musicians, potters, mosaic-makers, etc. to work for them. This pattern of economic development has a familiar modern outlook. It led to massive migration across the empire, which mixed backgrounds and diminished grievances, contributing to the development of an increasingly cosmopolitan, integrated society. As Charlesworth describes it:

In Gaul numerous Syrians are encountered; one enterprising Palmyrene had traveled as far away as Britain to do business; ... the quicksilver melters of Ephesus migrated to Rome, Alexandrians flocked westwards, and musicians from Egypt are found upon the banks of the Rhine... Thus over lands that had been pacified and secured by Roman arms and open seas that had been freed from piracy by Roman ships we see a stream of traffic flowing; merchants and traders were passing and repassing. Within the empire itself long voyages and journeys were made; ... men penetrated to far-distant lands, and the Roman name became known far and wide. The agents of Roman business had reached Ireland and ... the Baltic Sea, ... had met Chinese traders ... and bought and sold in the marts of India, and bartered goods with the Ethiopians.⁵⁴²

Rome's unusually open economic policy was thus largely responsible for spreading prosperity and growth to all corners of the empire and beyond, encouraging the acceptance of its rule.

⁵⁴² Charlesworth 237-240.

The gradual annexation of Rome's client-states as provinces was not only important for security reasons to secure control of the Rome's periphery, then, but also for economic reasons. Many of Rome's provinces were rich in resources. The province of Africa, for example, provided two-thirds of the grain consumption of metropolitan Rome during the Principate. Egypt also became one of the agricultural storehouses of the empire, producing a good part of the remaining grain. Spain produced most of the empire's raw metals and minerals. Syria and Judea represented another great economic potential because they provided access to the caravan routes leading to the Orient and thus gave the Romans control of the silk, ivory, and spice shipments they loved.⁵⁴³

While the Roman government only minimally regulated trade to ensure its free flow, it invested heavily in providing the necessary infrastructure to encourage its growth. The Roman roads' primary purpose remained military transport and military communication, but their secondary function was to facilitate economic activity and the transportation of goods. This was the intent of the Roman government, as clearly evidenced by the density of roads built around the richest areas of the empire, for example the Spanish mining regions of Asturia and Galacia or Northwestern Gaul's iron ore country. The Romans not only spent large amounts building the roads, but they also constantly maintained them, and generally each province was responsible for maintaining the sections of road that passed through its territory. Though most of the major roads were constructed by the 1st century B.C., whenever Rome acquired new provinces, the central government took charge of building new roads to access it, and it also directed the restructuring and update of itineraries. Hadrian built a new road in Egypt connecting to the Red Sea, for example, and built an alternative to Trajan's overcrowded Arabian road to Damascus. Domitian reorganized the road system in Asia Minor. Flavian and his successors improved the road between the Rhine, the Danube, and the Euphrates that connected Rome to the heartland of

⁵⁴³ Baradez 26; Webster 34-38.

Europe. Trajan engaged in a massive road improvement program throughout the empire that involved replacing porous surfaces with stone paving, adding drainage, building causeways over marches and bridges over rivers, filling in valleys, leveling gradients, and digging tunnels.⁵⁴⁴

Rome's massive public works did not focus solely on roads, though. Rome's building programs included some of the most modern infrastructure of the time that showcased the Romans' innovative engineering skills. In the African provinces, for instance, Rome developed massive hydraulic works, mainly for the purpose of "consolidation and protection of the land against rain waters in a brutally ... extreme climate," Baradez points out. The project was designed to force rain waters to infiltrate into the soil, and the water would then be preserved in a "network of colossal artificial sponges." To then bring those water surpluses to the agricultural basins lacking water, the Romans built a network of canals and reservoirs. Most of the work was conducted by Roman legionaries and auxiliary troops stationed in the province. Thanks to their efforts, Baradez concludes, the Romans were able to reclaim a 60 km-wide zone previously conquered by the Saharan desert steppes, and transform a previously arid and hostile lands into a cereal- and olive-producing region. The Romans also developed ports and coastal infrastructure. Claudius and Nero, among others, took a particular interest in naval trade, and invested in the widening of ports, at Ostia, for example, and the building of lighthouses, beacons, and breakwaters to protect ships. They even dug up new harbors on busy trade routes, such as Antium, to facilitate navigation. One of the most innovative projects the Romans took on was probably the canal running through the 4-mile-wide Isthmus of Corinth, to save time from rolling vessels to the other side or unloading and reloading cargo on each side. Caesar's engineers submitted plans for such a canal in the late Republic, but Caesar was assassinated before he could set the project in motion. Nero later renewed the project and began excavations, but the

⁵⁴⁴ Chevallier 148; Charlesworth 229-232.

project was stopped again at his death. The canal was not finished until the 1880s, when the currently used canal was built on the foundations of Nero's. Finally, the most expensive public works project undertaken by rising Rome was no doubt the Aqua Marcia, the longest aqueduct supplying the city of Rome in water. It was built between 144 and 140 B.C. and measure ca. 56 miles, most of them underground, with large arcades in the above-ground sections.⁵⁴⁵

In addition to promoting commerce, Roman infrastructure generated its own economic growth by creating a whole new branch of services. The best example is the Roman roads. The roads were used by everyone from troops and high dignitaries, its primary users, to commercial traffic like food supplies and cattle transport, pilgrims, tourists, artists, teachers, people traveling to the games or theatre, funeral processions, etc. Because of the massive traffic, the roads became a business in itself, with an array of road-side services offered: private inns for travelers, horse and vehicle rentals, blacksmiths, and farm stands with local foods and fresh produce. Thus, an infrastructure first built for the army became the cornerstone of the Roman economy.⁵⁴⁶

The financial infrastructure of the rising empire was the last innovative pillar of Rome's economic growth. By subtly spreading, though not imposing, Roman currency, the Romans developed a prime tool of commerce. Already in Republican times, Rome did not enforce uniform coinage through its lands. Roman coins, introduced in the 3rd century B.C. (the gold aureus, silver denarius, bronze sestertius, brass dupondius, and copper as) were accepted everywhere, but the Romans also allowed the use local coins, which carried their local emblems and were often the standard of taxation. Especially in Greek lands, local coinage was widespread. As the empire grew, however, the use of Roman coins became increasingly prevalent as a uniform, more convenient vehicle of exchange. Because the Roman currency was

⁵⁴⁵ Livy XXXIX.44.5, XLI.27.5-13; Harris 71-2; Baradez 34; Charlesworth 227-9.

⁵⁴⁶ Charlesworth 203-205, 232-233; Chevallier 189.

stable and widely available, it was trusted as safer than local currencies. Especially under the Principate, Starr mentions, the emperors "furnished a good coinage which ... inspire[d] business confidence." Its use even spread beyond Rome, further prodding commerce and playing a role of universal tender akin to the dollars' today. Archeologists found Roman coins as far as Sweden.⁵⁴⁷

Another sign of Rome's financial innovativeness and its flexibility in administrating its empire is that of its customs system. As mentioned before, the Roman government was minimally invasive in private business transactions, intending to impede exchange as little as possible. In line with this overarching goal, Rome came up with an new, effective customs system, the portorium. Whereas most states at the time, especially large states, levied a variety of customs taxes, rendering trade exchanges highly complicated, the portorium consisted of a simple, all-inclusive tax. Because the Roman government was more interested in raising money than in regulating commerce, it used one tax simultaneously as import/export taxes, town dues, and tolls. Thus, it was collected at the empire's and provinces' borders, at the gates of large cities, trading centers, and ports, and also on important roads, junctions, passes, bridges, and fords. Until the early Principate it was gathered by tax farmers, or private tax contractors, then replaced by civil servants. The tax was not overly invasive because it only applied to commercial goods. Animals, vehicles, all personal property, public goods, army supplies, and certain local goods used in festivities were exempt. Fraud was severely punished, but Roman authorities, again ahead of their time, established a special jurisdiction to settle disputes.⁵⁴⁸

The Roman taxation system was similarly pragmatic and incrementally improved when the circumstances required it. During the Republic taxes included a 1% (3% in wartime) wealth tax on property—including land and real estate, slaves, animals, and capital—and military

⁵⁴⁷ Lintott 48-48; Starr 116-119; Charlesworth 240.

⁵⁴⁸ Chevallier 195-7.

service. In 167 B.C. the Roman government discontinued the wealth tax on citizens because it had amassed considerable wealth through its conquests. Allies and client-states continued paying their traditional dues of troops and military material. Non-Romans in Italy paid the wealth tax until they became citizens after the Social War but all other provinces continued to submit to both the wealth tax and military service even after both were abrogated for citizens. The provinces thus became Rome's main tax base. As a result, accurate census was important in the provinces. But especially in the most remote provinces, where administrative services were not as well-developed, detailed reporting remained problematic. In those remote provinces, the taxes were then assessed on entire communities rather than on individuals. From the mid-Republic on, taxation was in the hands of tax farmers, the publicans. Under supervision of the provincial governors, those private contractors would advance the expected tax revenue to the state, then collect the money from the people. In other words, they were creditors that the population had to pay back. The Romans likely adopted the tax-farming system from the Greek city-states of Italy, in an effort to render tax collection more efficient. But it soon led to widespread abuse by the publicans, who sought to collect more than they were owed and frequently imposed high interest on the population, even in some cases resorting to extortion.⁵⁴⁹

Contracting out a task like tax collection seemed natural and uneventful when the system was first adopted because the Roman "state machinery [was] minimal," ancient historian Ernst Badian explains. Many administrative tasks were delegated during the early-to-mid Republic, such as the building of temples and secular buildings, the feeding of the sacred geese on the capitol, the summoning of the assemblies, and even the provision of military supplies for the army. All those contracts were handled by publicans, literally 'people who handle the public property of the Roman people.' But almost since the beginning, publicans were tempted by fraud

⁵⁴⁹ Lintott 70, 74-80.

because the legal profits from public contracts were small, oversight mechanisms were virtually inexistent, and the Senate and provincial governors were reluctant to punish them because they needed them. Army suppliers during the war in Spain against Hannibal, for example, would purposefully ship "worthless goods on unseaworthy ships" and later claim insurance compensation for more valuable goods and ships. The publicans' tax-collection function and power grew fast with the territorial expansion of the empire. Eventually the publicans "were the curse and scourge of conquered nations, largely ... responsible for the detestation of the Roman name among the subjects of Rome," a result largely impeding Rome's efforts to foster its subjects' allegiance. After provincials kept complaining of chronic unfair treatment, Augustus realized the subversive danger posed by the publicans' actions, and he ended the tax-farming system and replaced it with direct tax collection by civil servants. He restored the 1% wealth tax and added a flat poll tax per individual, possibly to finance the collection. Roman citizens remained exempt of taxes, unless they owned property within the city of Rome.⁵⁵⁰

Rome's awareness that economic prosperity and fair treatment would promote support for the empire is also evident in its uniquely generous welfare policies. From the early Republic on, the government considered it its duty to ensure that Rome was sufficiently supplied in grain, which provided the basis of the Roman diet. Livy explains that in periods of scarcity, the Roman government would purchase grain from neighboring countries and sell it to the population at a discounted price. With the boom of big agriculture and the rise of a new class impoverished, urbanized former small farmers after the Second Punic Wars had ravaged the Italian countryside, some Romans could not even afford the discounted government prices and demand grew for a more permanent welfare system to provide food for the poor. Up until then, the truly destitute

⁵⁵⁰ Ernst Badian, *Publicans and Sinners: Private Enterprise in the Service of the Roman Republic* (Ithaca, NY: Cornell University Press, 1972), 11, 14-18, 23-25, 32-33, 47; Starr 51.

were fed thanks to occasional charitable donations from the wealthy, but after the Punic Wars the number of destitute simply outgrew the episodic donations.⁵⁵¹

Therefore, in 123 B.C., tribune Gaius Sempronius Gracchus promoted the first Lex Frumentaria, which turned the episodic government discounts into a systematic program, increased the discounts, and set up a system of public granaries to keep reserves. According to the new law, each patriarch was entitled could request a certain amount of wheat per month to feed his family, at roughly half of the market price, Appian describes. While the entitlement was not restricted to the poor, the necessity to apply in person and in public most likely deterred all but the neediest. The Lex Frumentaria was later reinforced by other similar laws, and eventually 58 B.C., the Lex Clodia of tribune Clodius initiated the governments' free distributing of grain to needy citizens. The devastating events of the Civil Wars, however, continually increased the number of welfare recipients, and many non-citizens were even able to claim the benefit. As the Lex Clodia threatened to bankrupt the Roman stat, Julius Caesar, rather than repealing it, simply capped the number of recipients to 150,000, roughly diminishing the number by half. None of his successors dared to abolish the free distribution for fear of unrest, although the numbers fluctuated. The Roman Republic and Principate thus spent a very large amount to feed their poor, an expense that exceeded by far the welfare spending of other ancient states and complemented their growth-oriented economic policies.⁵⁵²

c. Stable Yet Flexible Political Rule (IV6.3)

In a further effort to generate allegiance and sustainability for the empire, Rome developed a uniquely flexible political rule, emphasizing local autonomy while at the same time

⁵⁵¹ Livy II.9, 34, IV.12, 52, X.11, XXVI.40; Harris 72-3.

⁵⁵² Appian I.21; William Smith, *A Dictionary of Greek and Roman Antiquity* (London: John Murray, 1875), 548-551.

preserving stability to enable expansion and limiting the arbitrary exercise of power—a seemingly impossible combination generated by Rome's pragmatism.

Non-interference and tolerance for local administrative, political, and cultural practices was a first cornerstone of Rome's novel style of imperial control. Focused mainly on establishing security, the Roman government left large autonomy to provincial communities, aware that less invasive government would be more likely to keep local populations content and satisfied with the regime. Rome's governing style was highly unusual for a rising hegemon, as growing empires more frequently try to exert maximum control and coerce populations into submission by standardizing behavior, worried that local identities might fuel nationalism and secession. In contrast, "Rome had discovered the secret which is still hid from many governing peoples, that an alien ruler can win the respect and even the affection of his subjects if in the affairs of everyday life he refrains from unnecessary interference and is content with the maintenance of peace, law, and order," Stevenson explains. Rome clearly strived to become a hegemonic, rather than simply a territorial, empire, and that included generating support rather than uniformity. The acceptance of local practices and traditions reached all dimensions of daily life. The preservation of local coinage, as mentioned above, is just one example of Rome's flexibility. The use of local languages is another. While Latin became the empire's administrative and military language for simplicity's sake, each province and even smaller entities were left free to keep their own language. Greek remained the official language in most of the Eastern provinces. The Romans did not press locals to embrace their religion or culture, either. The Gauls, Spaniards, and Britons kept their own gods, their own languages, their own loyalties, and their own tribal and political rifts.⁵⁵³

⁵⁵³ Stevenson 1-4.

The flexibility even extended to local administration, which largely kept its diversity. The Roman empire cleverly offset its imperialist aggression with political devolution. In general, Rome "interfered as little as possible with native institutions and made no attempt to impose homogeneity," as Stevenson points out. From early on, Rome recognized a large dose of self-government, beginning with Italy. Italian allies benefited from vast internal autonomy, though it was not absolute since Rome dictated their foreign policy moves and succession issues. Even when the Italian allies became incorporated into the Roman empire, towns largely continued to manage their own affairs. Most kept their own assemblies and magistrates, set up according to their own rules, to determine local issues. A prefect from Rome was generally appointed to supervise and deal with tasks specifically assigned to the central government, like maintaining order and security. The same was true for non-Italian states incorporated into the empire as provinces, voluntarily or after conquest. Just like in a federal type of government, everyday decisions were by default left to the individual cities and towns, while a certain number of specifically defined issues were the resort of the central government, like the military, foreign relations, the justice system, or taxation, for example. In other words, the Roman central government provided the structural framework of the empire, within which each province was largely free to evolve as it wished. Just like the Italian regions had prefects, the provinces were supervised by Roman governors or proconsuls, with the exception of certain unstable provinces during the Principate that were directly administered by the emperor. Beyond that, each local government at the level of province, city, and town, varied. In general it included a number of magistrates, usually elected and holding the executive power, an assembly of citizens and a nominated advisory body holding the legislative power.⁵⁵⁴

⁵⁵⁴ Starr 122; Baradez 26, 29; Stevenson 10-11, 15-16, 37.

Provincial governors, because they were the main link between the central government in Rome and the provinces, played a major role. Though their authority was limited to certain functions, they possessed vast powers. The governor was the chief military commander of the province, as well as its chief tax collector and chief accountant, in charge of the province's budget and finances. The governor was also the chief provincial judge, and had the general role of managing the province. But the governor could not possibly handle all those administrative tasks himself. In fact, Lintott recognizes, "Roman government [in the provinces] would have been impossible without immense delegation of government." Wherever there was a preexisting local—city, town, or communal—government, "Rome made use of it." And where it did not preexist, "Rome encouraged its development," Lintott adds. The governor kept firm control over his above-mentioned mandates, while delegating many tasks to the local governments. The vast local autonomy Rome granted to its provinces was thus not only a choice to generate support, but also a necessity to administer its vast territory.⁵⁵⁵

Because of their vast powers and often distant location, governors also presented a risk for the Roman empire. Their power needed to be balanced with sufficient oversight. Therefore, governors were appointed by the Roman Senate and later by the emperors and thus remained under their supervision and had to follow their instructions on how to run the provinces. The Senate was also in charge of providing the provinces' budgets and could control unruly governors that way. Although the provinces' conquest booty and later taxes were kept in the provincial treasuries and were thus available to the governors, the Senate always remained responsible for granting supplemental budgets, which the provinces often needed. The governors' tasks were legally circumscribed in several documents like the Lex Porcia of 101-100 B.C. and the series of Leges de Repetundis passed in the 1st century B.C., which defined legal and illegal

⁵⁵⁵ Lintott 54.

activities, particularly with regards to gifts and bribes. Governors were also barred from using their position to secure land acquisition or slaves. Though those laws did not eliminate the corruption of governors and other Roman provincial officials, they contributed to curb it.⁵⁵⁶

The Senate also had the power to remove and try corrupt governors and provincial officials who would take legality into their own hands. Normally the governor possessed the supreme judicial authority in the province, but in case an individual or a community had a dispute with their governor, to avoid arbitrariness they could take their case to the Senate. There are reports that as early as 171 B.C., the Senate was particularly attentive to "complaints of mistreated provincials," Harris writes. That year, a delegation of provincials from Spain declared before the Senate that they had been exploited by their local Roman officials, and the Senate appointed a special commission to investigate their allegations. To make matters even fairer, high provincial officials were in certain cases tried in front of a jury, specifically appointed after an elaborate selection process. If convicted, officials would be fired or fined. In 170 B.C., for example, two praetors were fined hefty sums for wrongfully sending a number of Greeks into slavery during the 3rd Macedonian War. A few years later in 149 B.C., the Senate enacted a law allowing provincials to recover stolen property taken by Roman officials. The penalty was the restitution of twice the value of the stolen goods.⁵⁵⁷ In fact, the Roman legal system, which was far more elaborate than any of its contemporary equivalents and provided the groundwork for many modern legal systems, acted as the common thread throughout the empire. One of the rare aspects of Roman administration to be applied uniformly throughout the empire, it gave the empire a sense of stability transcending the diverse local government practices and provided a bulwark against arbitrariness that further reinforced popular support for Rome's rule. "Its

⁵⁵⁶ Harris 159-160.

⁵⁵⁷ Lintott 44-45, 47-48, 97-102.

extension through the empire was the greatest benefit conferred by on her subjects," Stevenson argues.⁵⁵⁸

Rome's unique central government system, first republican, then imperial but with a remnant of republican institutions, was largely responsible for allowing Rome's provincial flexibility. The norm in antiquity was that empires were led by kings, not republican city-states—though some individual city-states, particularly in ancient Greece, managed to become powerful, all remained short of imperial proportions. Monarchical rule, however, was by definition more rigid and authoritarian and thus less appealing than republican rule since authority rested on the sole person of the monarch while a republic was ruled by a collective assembly. Besides monarchies, the only other larger powers in antiquity were collections of city-states or leagues, like the Aetolian or Achean Leagues, but their fragmented power limited their hopes of external growth. Thus Rome, an individual, republican city-state, benefited from a one-of-a-kind political system for an entity of its dimension and aspiration. It combined the advantage of the Leagues, having a republican form of government that limited the alienation of its subjects, with the advantage of monarchical rule, i.e., concentrated power that allowed for a unitary, more efficient foreign policy.⁵⁵⁹

Polybius is perhaps the best advocate explaining why Rome's unique form of government was a major factor contributing to its hegemonic success. "... The best and most useful aim of my work is to explain ... by ... virtue of what political institutions almost the whole world fell under the rule of one power, that of Rome, an event which is absolutely without parallel in earlier history," Polybius writes. "The principal factor which makes for success of failure [in hegemonic pursuits] is the form of a state's constitution: it is from this source ... that all designs

⁵⁵⁸ Stevenson 3-4.

⁵⁵⁹ Eckstein 245-6.

and plans of action not only originate but also reach their fulfillment," he further asserts. Polybius wrote during the height of the Republic and was a firm believer in the superiority of Rome's institutional design because its combination of different sources of authority gave the regime flexibility and a capacity to adapt to most circumstances. Rome's constitution was uniquely effective, Polybius argues, because it blended elements of monarchy, oligarchy, and democracy without suffering from the disadvantages each type presented on its own. Because it gave access to power in some form to all classes except for slaves, the Roman Republican constitution offered something to everyone and thus garnered the support of all. The consuls, and occasional temporary dictators, were the monarchical elements as supreme executive leaders and commanders-in-chief. The Senate constituted the oligarchic element, since membership was open only to aristocrats, and was the legislator, chief diplomat, head of the empire's finances, supreme judge for certain public crimes like treason or assassinations. Finally, the tribunes or assembly of the people represented the democratic element, since they were elected to public office by the population and were responsible for deliberating and approving the Senate's laws, declaring war and peace, ratifying treaties and alliances, and judging all other violations of the law.⁵⁶⁰

The strength and flexibility of Rome's institutional design came from its unprecedented system of checks and balances, which testifies to the innovative nature of Rome's constitution. Each branch was linked to the others in a way that made them interdependent and mutually balance each other out. In fact, the Roman institutions were so intertwined, Polybius writes, "that it was impossible even for the Romans themselves to declare with certainty whether the whole system was an aristocracy, a democracy, or a monarchy." The consuls technically retained absolute executive power, but in practice they needed approval of the people and the Senate for

⁵⁶⁰ Polybius VI.2- 3, 11-15.

many of their activities. For example, the Senate furnished the finances necessary to gather war supplies, so consuls could not send off to war without their approval. This explains, for example, why Scipio Africanus was forced to battle Carthage in North Africa only with volunteers—the Senate had denied him funding. Similarly, the people's assembly ratified peace and alliance treaties and thus consuls were highly dependent on their actions. "Under no circumstance is it safe for the consuls to neglect to cultivate the goodwill both of the Senate and of the people," Polybius points out. In addition, Rome's consuls were most frequently two, and thus had to share executive decision-making powers, which also limited them, and they were elected by the Senate for a short, one-year term only, which reinforced their accountability in front of the Senate. Similarly, when a dictator was elected in times of emergency to cumulate the power of the consuls, it was always for a limited tenure, generally one year also. In their constant effort to balance power, the Romans even appointed a co-dictator after their defeat at Lake Trasimene, when Hannibal was getting close to conquering the city of Rome, to compensate the perceived political deficiencies of the dictator they had elected.⁵⁶¹

The Senate, then, possessed the default legislative power, but the tribunes had a veto power on many of the Senate's decisions. Just one tribune's veto could block the Senate's will in many cases. And because the tribunes were bound to represent the people who elected them, "the Senate stands in awe of the masses and takes heed of the popular will," Polybius emphasizes. The People's assembly were also forced to take into account the Senate because it had the keys to Rome's revenue. Because the tribunes were charged with executing some of the legislation, such as the public works and state contracts, they needed the funding. And while the tribunes had the authority to set up courts for civil trials, judges were often drawn from the Senate. Thus, the three powers constantly needed to cooperate, particularly when facing a common threat, and

⁵⁶¹ Ibid., III.87, 103.

could be very powerful but the need to work in sync prevented the tyranny of any one power over the others. The constitution naturally corrected abuses of power. "The result is a union which is strong enough to withstand all emergencies," Polybius concludes. Rome was the only polity and aspiring hegemon in the ancient world to provide such an equilibrium of power. Athens did not have checks and balances and mixed sources of power, while the Spartan system did not allow for enough concentrated power in cases of emergency, and Carthage offered too much democracy to enable effective war-making, as evidenced by their difficulties to reach decisions on how to proceed at the end of the Punic Wars.⁵⁶²

When the empire grew in size, the Romans simply adapted their form of government to rule more efficiently over the larger territory. The rigid senatorial democracy praised by Polybius did not provide enough control options for the periphery once the empire reached such unruly areas as Germany. Because only certain aristocrats were eligible to become senators and thus provincial magistrates, there were in effect not enough statesmen to administer the empire. To each province created through 146 B.C., Rome sent each year only one elected, unsalaried governor and one assistant. The central government's oversight of the provinces was thus insufficient and lacked continuity since governors changed each year. Those problems contributed to the political turmoil of the last century B.C. and the end of the Republic, as more and more politicians in Rome were offering solutions to govern the empire more effectively without losing it. The tension grew between conservatives and reformists, and the empire reached the verge of collapse, as the institutions formatted for a small-to-medium power did not know how to deal with the rapid political, economic, and social changes accompanying the massive growth of the empire. Some reforms attempted to strengthen the Senate, but always fell through. As no reform was able to take root, the turmoil eventually engendered the Civil Wars,

⁵⁶² Ibid., VI.15-18.

which brought in the first emperor, Augustus, who singlehandedly "reorganized the ... structure of government built up in the Republic" to match its territorial growth. His reforms enabled Rome to reach its heights and lasted until 180 A.D., for almost two centuries.⁵⁶³

Augustus' imperial institutions had surprising continuity with the Republic, however. While one would expect the emperor to have removed all traces of Republic and installed an authoritarian state, perhaps best suited to effectively control such a vast territory, Augustus instead formally restored the Republic and the Senate. He thus acknowledged that he considered the ultimate power to belong to the people, but at the same time he substituted the authority of an emperor to the dual consuls, which enabled him to establish direct, concentrated authority over the executive and the provinces and bind together all the elements of the empire. "In one sense it is not correct to use the terms 'Empire' and 'Emperor' to describe the Augustan system and its director," Starr analyzes. Augustus, Nero, Hadrian, and their successors benefited from large powers but were not hereditary rulers; instead, they were "persons to whom the Roman people by solemn law gave a collection of purely legal powers." In fact Augustus did not call himself emperor but 'princeps,' or first citizen, and called his system of government 'Principate,' which seems more appropriate than Empire. Under the reformed government system, provincial posts were multiplied, remunerated, and closely supervised by the emperor, ensuring better, less corrupt provincial administration. The size of the bureaucracy also dramatically grew to accommodate the new size of the empire. And while over the long term the power of the emperors tended to grow at the expense of its counterbalance, the Senate, "a lot of power resided in the administration" and growing bureaucracy too, creating a new type of counterweight to the

⁵⁶³ Starr 49-50, 93.

emperor. Augustus's reforms, including the provincial government, central bureaucracy, senate and princeps, "all combined to produce a lasting peace for the empire," Starr concludes.⁵⁶⁴

Rome's political system was thus successful in retaining both the support of the population and control over a growing territory—a difficult balancing act, thanks to its innovative combination of central power, local flexibility, and limited arbitrariness.

d. Social Inclusiveness (IV6.4)

Finally, besides its innovative economic and political measures, the third part of Rome's empire-building efforts was to offer its subjects a socially inclusive environment. Rome's combined incentives for newly conquered populations to embrace Rome and subtle efforts at Romanizing were aimed to generate allegiance for the growing empire and forestall revolt against its expansion. Rome's bid for hegemony was successful in large part because it was a socializing, perhaps even civilizational enterprise: Rome was able to integrate both former enemies and friends into its empire and thus build a state and not, like many other aspiring hegemon, a collection of entities with incompatible interests. The loyalty it sought from its subjects was not the kind that came out of fear of reprisals, but rather out of the desire to be included.

The first hint to Rome's unique approach was its generous extension of citizenship and immigration rights to former enemies as well as friends. Instead of beheading or enslaving all defeated enemies as was customary in the ancient world, Rome offered many of them a place in Roman society. It imposed milder settlements on the losers than was standard practice at the time, which generated further gratefulness and preempted any revanchist feelings among the

⁵⁶⁴ Ibid., 67-74, 95-100, 102-103, 107.

losers. This practice started early in the Republic. In 381 B.C., for example, after Rome won over the Latin city of Tusculum, it offered its inhabitants Roman citizenship, while allowing them to retain their distinct identity and local government, a compromise called *municipium* and offered to several other cities around Rome. In the 380s B.C. also, the Romans granted its informal ally, the Etruscan city of Caere, citizenship without the right to vote in Roman elections (*civitas sine suffragio*), another common compromise. In 340 Rome gave full citizenship, including voting rights, to 1,600 Capuans to reward them for their support in the Latin War. Soon an increasing number of Italic people obtained Roman citizenship; by 300 B.C., Rome already had over 200,000 citizens. Eventually, by 88 B.C., all Italian allies South of the Po were granted some form of Roman citizenship. Such a practice was inconceivable in the Greek world, by contrast. At the peak of its power in the 5th century B.C., for example, Athens only counted 6,000 citizens, and it strictly segregated citizens from non-citizens. Members of the Achean and Aetolian Leagues also jealously kept their citizenships distinct, even though it hindered their unity.⁵⁶⁵

Rome's inclusiveness did not stop at the border of the Italian peninsula but was extended throughout the empire, even to populations with identities and cultures highly different from the Romans', creating a very fluid and cosmopolitan society loosely based on Roman culture. In fact, Eckstein asserts, Rome "replaced ethnicity and geographic location as the basis of membership in the polity with a ladder of legal status-groups not tied to either ethnicity or geography," including resident ally (the *socii*), citizen without suffrage (*cives sine suffragio*), and full citizen (*cives*). "And because the Romans were relatively generous in allowing non-Roman individuals and even (very occasionally) whole non-Roman polities to climb up this status hierarchy, Rome gained an enhanced capacity to win loyalty, or at least acquiescence," Eckstein stresses. An increasing number of provincials, though not all, became Roman citizens. Caesar, for example,

⁵⁶⁵ Eckstein 246-7, 251-5; Starr 65-66.

gave citizenship and suffrage even to the Cisalpine Gaul province that had been so troublesome to Rome in the past. Eventually, even some of Rome's most powerful emperors were of foreign origin. Emperor Claudius was a descendent of immigrants, and Emperor Trajan a provincial from Spain, demonstrating the amplitude of Rome's integrative reach.⁵⁶⁶

Rome's focus on integration may in fact have played as important a role as its military might in determining its hegemonic success. Starr argues that the reason Rome "worked" so well and was able to last as a hegemon was its openness and cosmopolitanism. Beyond its military activities and conquests, it was able to construct an empire which was a "fusion of Roman culture with Greek civilization." Rome essentially created an "urban, Mediterranean" identity, a "summation of ancient civilizations" that was appealing beyond the Italian peninsula. In other words, Rome's unique inclusiveness led the identity of the Roman citizen to become slowly divorced from ethnicity, location, and other divisive determinants and "it was [that] flexibility that helped make Rome so formidable," Eckstein writes. As a result of Rome's integrationist policies, the social structure of the Roman empire was remarkably stable. Despite a few class tensions between rich and poor and aristocrats and common people, there was never any mass insurrection like elsewhere; different classes and groups were used to compromise. In fact, the lower classes were even able to gain a number of rights by the 3rd century, B.C., increasing their legal protection and political representation. "Rarely in history has a state been able to carry out such great changes without violent explosions," Starr notes. The absence of domestic instability was a circumstance greatly favorable to the rise of Rome.⁵⁶⁷

But Rome also took active steps to promote its new elastic Roman identity. It coupled its integrationist policies like the granting of citizenship with subtle but effective attempts at

⁵⁶⁶ Eckstein 311-312.

⁵⁶⁷ Starr 2, 22-23, 74-75; Eckstein 250.

Romanization, or spreading Roman culture and practices throughout the empire, beyond the Italic and Hellenistic zones that were most receptive to Rome's civilizing role. As Webster underlines, "Roman governors and high-ranking officials, military and civilian, spread the Roman way of life [in the provinces]." While a few, like Tacitus, suggested that Rome's efforts to civilize the barbarians of the empire might be colonialist and oppressive, most Romans saw it as a noble enterprise to bring them much-needed amenities and refinement. Thus, local Roman authorities helped the provincials build towns, houses, temples, marketplaces, theatres, courthouses, and bathhouses, educated the sons of tribal chiefs to Latin and liberal arts, encouraged the people to wear clothes and attend the baths and local banquets, taught them commerce, communication, and entrepreneurship, spreading the Roman urban way and greatly improving living standards in the difficult-to-reach corners of the empire.⁵⁶⁸

There were multiple vectors of Romanization at work simultaneously. The massive number of Roman veterans who settled with their families in the provinces were key actors in helping local Roman authorities to Romanize the provinces, by setting the example of the Roman way of life for the natives. Veterans "played a considerable moral and material role" in developing provinces like Africa, as they became increasingly rooted to their new homes, Baradez points out. The provincials' service in the auxiliary forces was a second major factor in the spread of the Roman way, since auxiliary troops, often purposefully recruited from the barbarian districts as a way to discourage rebellion, were accustomed to Roman practices during service and adapted to Roman hygiene, food, etc., and would upon discharge return to their native districts and teach their families and tribes what they had learned. A third, crucial means of Romanization was Rome's tradition of hostage-taking. Rome regularly brought the sons of local tribal chiefs to Rome and provided them with a Roman education, and upon their return to

⁵⁶⁸ Webster 282-4.

their tribes they would not only have become sympathetic to Roman culture but they would also share and implement the Roman practices they had learned.⁵⁶⁹

Finally, Rome's economic efforts in themselves promoted Romanization. As Starr explains, "the material prosperity of the empire" provided the seed for the "rapid spread of Romanization throughout its lands and seas," as the new prospects for growth that Rome offered encouraged the people of the provinces to embrace and copy Roman culture and civilization.⁵⁷⁰ As a necessary chain in the material prosperity of the empire, the network of Roman roads was an additional instrument of Romanization. Charlesworth even suggests that the roads bore a "moral effect," as they contributed to spreading the Roman language, Roman myths and stories, and also Roman arts and crafts as it propelled the movement of artists and craftsmen and hence spread their methods throughout the empire. By the early Principate Rome had 372 main roads that covered a total of 53,638 miles. The civilizing effect of the road system is evident; it was the "line of advance" not only of commerce and army, but also "the binding force between races and cultural influence." The Roman roads, Charlesworth concludes, "served to unify the Roman world and so, at long remove, to create modern Europe."⁵⁷¹

Rome's subtle civilizing process worked beyond Rome's expectations. Social historian Ramsay MacMullen set out to assess the success of Romanization by looking at indigenous adoption of basic Roman practices, like the espousal of the Roman three-name custom, diets, clothing, cooking and building utensils, religion, and language. The mainstream historical approach, which focuses on legal, administrative, governmental, and infrastructural aspects as evidence for Romanization, is insufficient because it fails to show the voluntary embracement of Roman culture as does the adoption of everyday Roman practices. MacMullen's findings were

⁵⁶⁹ Baradez 26, 36.

⁵⁷⁰ Starr 120-121.

⁵⁷¹ Charlesworth 205-207.

that Romanization worked surprisingly well overall—Roman practices were accepted with very little resistance, primarily because of Rome's flexibility, which allowed local practices to continue and mix in with Roman practices. Romanization, however, was mostly a top-down phenomenon, McMullen adds, embraced first by the indigenous elites, the Romans' most important political allies in the provinces, before trickling down to the population. The three-part Roman names, Roman fashion, the Latin language, and everyday objects like cooking utensils were first adopted by the more urban and wealthy provincial upper classes and took much longer to reach the home of the poor, who were mostly rural and less educated and thus less exposed to new developments and also lacked the resources to be able to adhere to Roman fads. In addition, because Roman culture itself was more affluent and urban, it was naturally more attractive to those provincial populations most similar to it. Upper classes in the provinces were also motivated to embrace Roman ways for instrumental reasons, like ambition, status, self-esteem and prestige, and saw Romanization as a vehicle to further climb the social ladder.⁵⁷²

Romanization made its way down the social ladder and was eventually successful because the Romans remained tolerant of local customs, so that the lower indigenous classes remained appeased. Because the Romans were vastly tolerant of local practices, when revolts brewed in the provinces they were mainly instigated by the elites and not by the common people, and more frequently by the Romans in the provinces than by the provincials. The few uprisings started by locals were in fact most often directed against other locals, an expression of internal division rather than animosity toward Roman rule. Because they did not engage in forced acculturation, the Romans encountered little resistance to their cultural influence. They preserved local political practices and pre-existing currencies, as mentioned before, in such provinces as

⁵⁷² Ramsay MacMullen, "Notes on Romanization." *Bulletin of the American Society of Papyrologists* Vol. 21 Issue 1-4 (1984), 161-176.

Thrace, Gaul, Africa, and Spain. Since the ordinary man in the faraway provinces had little to gain from citizenship, name changes, and Roman fashion, the Romans often simply left him alone. Roman and non-Roman customs were even often mixed, another sign of Rome's flexibility. For example, modes of worship would combine Roman and Celtic influences, or art forms would blend Latin and native influences, highlighting the persistence of indigenism. While the conventional literature tends to regard Romanization as the massive and forceful imposition of Roman ways, it fails to take into account the unique tolerance and flexibility the Romans showed in their provinces. "A whole second world may have existed beneath the one familiar to us, containing an actual majority of the population, a wholly un-classical civilization," MacMullen concludes.⁵⁷³

Thanks to its subtle civilizing endeavor the rising hegemon was able to spread its cultural influence and generate a cohesive Roman society that transcended but did not suppress local identities and practices. By coupling this uniquely inclusive social policy with other innovative state-building efforts—a growth-oriented economic policy and flexible political institutions—Rome succeeded in obtaining the support of its growing number of subjects and building the basis of a lasting, popularly embraced hegemonic rule. Rome's state-building efforts were central in supplementing Rome's military conquests and counterbalancing its superior use of force, and were thus a necessary part of its success as a hegemon.

Conclusion

Rome was able to transcend balance of power and reach the status of hegemon because of a unique combination of factors. The first causal path originates with Rome's potential balancers,

⁵⁷³ Ibid., 166-7, 170-2, 177; Ramsay MacMullen, "How to Revolt in the Roman Empire." *Rivista Storica dell'Antichità* Vol. 15 (1985), 67, 70-1, 73-4.

who were unable to shut down to Rome's growing power. Strong balancing could have put serious hurdles in Rome's path and ultimately prevented its accession to hegemony, but Rome's Mediterranean neighbors failed to act. Though they experienced few physical communication hurdles (IV1.1), they were frequently deprived of the correct information about Rome's hegemonic rise by their own misperceptions of Roman power and intentions (IV1.2), in many ways reinforced by Rome's deceptive friendliness and cooperative posture (IV1.3; IV2.1; IV3.1). But even if potential balancers had not been deceived by Rome, it is unlikely they would have balanced effectively, because their constant power struggles caused them to focus on their immediate gains at the expense of the bigger threat, Rome, and prevented them from cooperating to stop Rome. Most often this in-fighting precluded alliances altogether (IV2.2); but even in the few instances where the potential balancers managed to act in concert against Rome, the distrust resulting from their frequent conflicts mostly destroyed all common efforts before they could reach fruition (IV3.2). In addition, many potential rivals bandwagoned with Rome, some out of fear (IV4.1) and others to extract profit (IV4.2), but mostly because the Romans enticed them to become its clients or friends instead of opposing it (IV4.3).

But the failings of the potential balancers alone cannot explain Rome's successful rise to hegemony. The second causal path derives from Rome's own superior skills, which enabled it to grow fast and consistently to surpass all of its competitors. Though the potential balancers' lack of strong opposition certainly facilitated Rome's task, it would not have been able to overtake its competitors and build such a vast empire had it not designed unique, innovative military and state systems. The superiority of the Roman military apparatus came largely from its professional, highly effective organization and strict discipline (IV5.3), reinforced by unique strategies and tactics always infused by state-of-the-art techniques and weapons borrowed from

competitors and further improved by the Romans (IV5.1; IV5.2). Finally, to counterbalance its reliance on military force to expand the territorial boundaries of the empire, Rome simultaneously sought to create a favorable living environment for its subjects, thus redressing the duress of conquest and gaining their trust and allegiance. Rome's empire-building projects included three ambitious, novel programs, developed throughout its rise and designed to create hospitable conditions that would attract the support of all subjects: a growth-oriented economy (IV6.2), flexible political institutions (IV6.3), and an inclusive social policy (IV6.4). Overall, Rome's military- and state-building enterprises, just like the empire itself, started as a small endeavor and gradually transformed through a series of reforms and adjustments into the most impressive instruments of power and control ever concentrated, demonstrating the Romans' learning skills and their incremental approach to hegemonic rise.

The Qin Unification of China

The rise of the Qin state to control China in the 4th and 3rd century B.C. and the growth of U.S. influence over the Caribbean and Latin America in the late 19th and early 20th century provide two further, yet very different, examples of successful paths toward regional hegemony. Both represent hegemony at a smaller scale than Rome's regional control. The Qin succeeded in dominating an exclusively continental, and thus more compact, region, which did not require them to develop naval skills like the Romans. The United States' hegemony in South America, representing the only case of regional hegemony in modern history, similarly occurred on a different scale, epitomizing the shift toward a modern hegemony that discards military conquest as a means of expansion in favor of a more informal and less easily quantifiable political, economic, and cultural influence.

Boundary of Inquiry and Timeline

The Qin kingdom rose to hegemony during a period of intense warfare between the seven main kingdoms of China that started in 403 B.C., known as the Warring States period, in which Qin successively conquered all other kingdoms, ultimately establishing the Qin Dynasty in 221 B.C. The Warring States represents the last period of the Zhou Dynasty, a monarchy that exercised indirect, feudal control over ancient China since 1111 B.C. and whose power had already begun waning during the historical period preceding the Warring States, known as the Spring and Autumn Period (770-481 B.C.). At the height of its power in 770 B.C., before a

barbarian invasion that forced them to retreat East, the Zhou dynasts directly oversaw roughly one thousand square miles of territory, and delegated the rest to kin-related feudal lords. Those lords exercised increasingly independent power and gradually consolidated their sovereignty over their respective lands during the Spring and Autumn Period. In 403 B.C., the clan leaders of Han, Chao and Wei forced the Zhou king to recognize them as independent entities after they had seized and divided up the Jin fiefdom, marking the formal demise of Zhou feudalism and the beginning of the post-feudal, multi-state period of the Warring States.⁵⁷⁴

During the Spring and Autumn Period, taking advantage of the weakening of the Zhou monarchs, several fiefdoms sought to expand and secure hegemonic control over the region, yet all were successfully checked by balance of power mechanisms. Chu's attempts at centralizing power and conquering its neighbors in central China resulted in a counter-attack by Jin, for example, which eventually defeated Chu in 632 B.C. Chu's own growing power and expansionist raids were then stopped by Wu, who defeated Chu and another contender, Qi, in 484 B.C. Wu's rise was consequently checked by Yue, which destroyed Wu in 473 B.C., and Jin's ambitions resulted in a civil war and its eventual partition into Han, Chao, and Wei.⁵⁷⁵

During that early period and the beginning of the Warring States, the kingdom of Qin remained in the background. Located on the western edge of China, it watched and pursued a predominantly defensive policy, mostly avoiding to take part in the successive expansion-balancing waves altogether. Qin "played no major role in interstate politics until the 360s

⁵⁷⁴ Tzong-Ho Bau, *The Stability of International Systems: A Study of the Warring States System of Ancient China* (Ph.D. diss., The University of Texas at Austin, 1986), 23-25; Ralph D. Sawyer, *The Seven Military Classics of Ancient China*, transl. by Ralph D. Sawyer (Boulder, CO: Westview Press, 1993), 13.

⁵⁷⁵ Victoria Tin-Bor Hui, *War and State Formation in Ancient China and Early Modern Europe* (New York: Cambridge University Press, 2005), 54-60.

B.C."⁵⁷⁶ Between 657 and 357 B.C., Qin started only 11 of the 160 major wars that shook the region, and was at best a "minor factor" in the wars it participated in. At the start of the Warring States period, Wei was the state to beat, led by the ambitious Marquis Wen (r. 445-396). It possessed the most strategically situated territory along the Yellow River, pursued military reform that dramatically enhanced its power, including the transition from small aristocratic armies to mass peasant armies that would later forge the success of Qin, and enacted a number of economic reforms that stimulated its growth, combined with a generous welfare policy. When Marquis Wen allied with the other two former Jin states, Han and Zhao, and won consecutive battles against the states of Qi, Qin, Chu, Song, and Zheng, Wei clearly became the "most powerful in the system." But its dominance did not last, as Qi undertook reforms of its own and started challenging Wei. By 341 B.C., after having defeated Wei twice, Qi, became the new regional leader.⁵⁷⁷

Simultaneously toward the middle of the 4th century B.C., however, the state of Qin began its own rise, engaging in innovative self-strengthening reforms that resulted in a dramatic increase in relative capabilities, thus marking the logical starting point of this inquiry. As a preamble, the Qin state began by moving against easier, strategically situated targets. It first turned toward the non-Chinese peoples bordering China, taking over the Rong tribes on its northern border, and expanded southwest by taking over the two small states of Shu and Ba in modern Sichuan, a fertile area that "became a major source of Qin economic and military power." Qin power grew spectacularly between roughly 366 and 293 B.C., and the Warring States multipolar system of seven major states—Qin, Qi, Wei, Han, Zhao, Yen, and Chu—lost

⁵⁷⁶ Mark Edward Lewis, "Warring States Political History," in Michael Loewe and Edward L. Shaughnessy, eds., *The Cambridge History of Ancient China – From the Origins of Civilization to 221 B.C.* (Cambridge, UK: Cambridge University Press, 1999), 617.

⁵⁷⁷ Hui 60-64.

members as they were gradually defeated by Qin. Qin and Qi, which was still the second strongest power of the region throughout the second half of the 4th century B.C., agreed to share the title of emperor of China in 288 B.C., but Qi was destroyed four years later by an anti-Qi alliance that Qin had engineered. Qin then turned against the only two remaining great powers, Chu and Zhao, which were already weakened by previous defeats, and crushed them in two wars in 280-276 B.C. and 262-257 B.C., respectively. By 257 B.C., "no single state could match it [Qin]."⁵⁷⁸ All the Warring States save Qin had lost their great power status and the power distribution in ancient China had essentially become unipolar. At that point Qin formally controlled about half of the region and just had to annex the rest to achieve hegemony, which it did during the final unification wars of 236-221 B.C. Han eventually fell in 230, Wei in 225, Chu in 223, Chao and Yen in 222, and Chi in 221, sealing the unification of China under the Qin, whose king became the First Emperor.⁵⁷⁹

The Qin was the first dynasty to unify China under centralized control and the scope of its territorial control was unprecedented. Chinese historian Dun Li described the Qin Dynasty's direct control as extending "from South Manchuria to North Vietnam and from the East China Sea to the Eastern slope of the Tibetan plateau, a territory that was one of the largest political units ever appearing on the face of the earth." The great extent of Qin's hegemonic achievement is further illustrated by a minister to the First Emperor, Shi Huang-ti, who claimed that the Qin Empire was "a great deed that can be achieved only once in thousands of years." Although perhaps exaggerated, this statement shows that people were already aware of the magnitude of the dynasty's power in the Qin times.⁵⁸⁰ Qin's hegemonic expansion was indeed impressive: from roughly 50,000 square miles at the start of its rise in the 360s B.C., the Qin state had grown

⁵⁷⁸ Lewis 587, 616, 641.

⁵⁷⁹ Bau 34-35; Hui 64-66.

⁵⁸⁰ Dun J. Li, *The Ageless Chinese: A History* (New York: Charles Scribner's Sons, 1978), 97

to encompass some 900,000 square miles by 221 B.C. at the advent of the Empire. And yet, the Qin's regional hegemony represents less than half of Rome's peak size of over two million square miles. Comparing early Roman and Qin rise, sociologists Edgar Kiser and Yong Cai highlight that "Rome was attempting to administer an area roughly 34 times as large as Qin," which may explain why it took Qin only a fraction of Rome's time to achieve hegemonic status. While therefore on a different scale than Rome, Qin nevertheless achieved one of the fastest and most extensive regional hegemonies recorded in history.⁵⁸¹

Surprisingly, however, the Qin leaders themselves, despite their great achievement, remained in power for a total of just fifteen years before being replaced in 206 B.C. by another lineage of monarchs that covered the same territory, the Han Dynasty. According to 1st century B.C. Han statesman Chia Yi, the Qin Dynasty was overthrown because its rule was oppressive and it was unable to turn its territorial conquest into political rule.⁵⁸² In other words, the Qin emperor abused his power and thereby spurred the rebellion that brought about his downfall. Yet while the Qin house did not survive, it set the tone for a lasting, unified control over China by the future dynasties. The ruling families changed, but the structure created by the Qin outlived them by far. As historian Robin D.S. Yates stresses, "the Qin succeeded in founding an imperial system that lasted, despite their own rapid demise, through many important social, economic, and political changes until the 1911 revolution of Sun Yatsen [which brought an end to monarchy in China]." One must thus ask, why did the Qin succeed at unifying China, despite facing balancing efforts and high expansion costs, and bringing about such groundbreaking transformation, when countless other Chinese fiefdoms and states before it had tried and failed?

⁵⁸¹ Edgar Kiser and Yong Cai, "War and Bureaucratization in Qin China: Exploring an Anomalous Case," in *American Sociological Review* 68.4 (Aug. 2003), 529.

⁵⁸² Chia Yi, "The Faults of Ch'in," transl. by Burton Watson, in Cyril Birch, ed., *The Anthology of Chinese Literature – Vol. 1: From Early Times to the Fourteenth Century* (New York: Grove Press, 1994), 48.

As Yates concludes, "the unification of China and the founding of the imperial order in 221 B.C. was by no means inevitable." Why did balance of power mechanisms not succeed where they had worked so well in the past?⁵⁸³

The question of how Qin was able to rise all the way to hegemony without being stopped is particularly puzzling because Qin was relative newcomers to the Chinese interstate community at the beginning of the Warring States period, and was long seen as backward and barbarian by the other Chinese states. Qin succeeded because of a combination of its own internal innovation and external collective action problems. The primary causes of the failure of balance of power in the Qin case were the refusal of the kingdoms of Han, Wei, Chao, Chu, Qi, and Yen to resolve their collective action problems, reinforced by Qin's clever divide-and-conquer strategies and deceptive incentives, and their inability to copy Qin's groundbreaking self-strengthening reforms or counter their political, economic, and organizational achievements and geographical advantages. Qin was thus able to close the gap with the other states, while also surmounting their relative power and preventing or shutting down countervailing alliances.

⁵⁸³ Robin D.S. Yates, "Siege Engines and Late Zhou Military Technology," in *Explorations in the History of Science and Technology in China*, "Special Number of the 'Collections of Essays on Chinese Literature and History' (1982), 409.

Map 7.1: China ca. 278 B.C.



Source: Ian Mladjov, University of Michigan, reprinted with the permission of Ian Mladjov.
Note: The small territory around the city of Luoyang, between Wei and Han, is the remnant of Zhou.

Map 7.2: China ca. 210 B.C.



Source: Ian Mladjov, University of Michigan, reprinted with the permission of Ian Mladjov.

1. The Balancers' Side: Collective Action Disasters and Vulnerability to Deception

Throughout its rise Qin faced six great powers that could have acted to stop it: Qi, Wei, Han, Zhao, Yen, and Chu. While there were a number of primarily nomadic, non-Chinese tribes bordering ancient China, as well as some secondary states within the former Zhou feudal network, all were too weak and/or disorganized to play a significant role against any of the seven main Warring States. The six great powers encountered overwhelming difficulties to coordinate their balancing efforts due to their deep-seated mistrust and immediate gain focus. When occasionally they managed to cooperate against Qin, their efforts were undermined by the same underlying issues. Their difficulties were reinforced by Qin's clever deceptive tactics and its ability to entice a number of the great powers into joining at its side.

a. Communication Problems (IV1)

Communication defections facilitated the growth of Qin, though just like with the rise of Rome, they played an auxiliary role and were not central in enabling hegemony. Once again, physical communication problems were not key, despite the primitiveness of communication technology in the 3rd and 4th centuries B.C. Instead, the other states' misperceptions and the Qin's deliberate deceptive efforts were responsible for the potential balancers' lack of accurate information about the Qin threat.

1. Physical Communication Problems (IV1.1)

Even though information circulated very slowly in ancient China, mostly via travelers, traders, or scholars who journeyed by horse and chariot or on foot, a number of factors led to an

unusually dense information exchange between the seven major Chinese states. The Qin's potential balancers can thus not have ignored their early rival's growth and expansion.

All seven Warring States covered a relatively compact geographical area. Unlike the Mongols, whose targets were often situated thousands of miles away from the Mongol home base, the Qin aimed primarily at their closest neighbors, so that information about their growing power did not need to travel halfway across the globe as it did for the Mongols. The Qin capital of Xianyang, near modern Xi'an in Shaanxi, was located only about 700 miles from the capitals of the two furthest Warring States, Qi and Yan. Not only were the Qin and their potential balancers geographically close, but except perhaps for the non-Chinese tribes that inhabited the perimeter of ancient China, they were also culturally similar, which further facilitated the exchange of information. The Warring States "shared a similar language, customs, and faith. The food, dressing, living and transportation of the seven great powers were quite close to each other... They shared similar music, handicraft, fine arts, dancing, exercise, education systems, literature and academic thoughts." This similarity suggests a high level of exchange between the states already before the Warring States period.⁵⁸⁴

In addition, because of their feudal relationship with the centrally-located Zhou during the period preceding the rise of Qin, the ancient Chinese states had developed an extensive political and diplomatic exchange system. Although the leadership of Zhou broke down in the 5th century B.C., these diplomatic channels were sufficiently well-established that they remained in use throughout the Warring States period. Diplomat and political scientist Samuel Shih-Tsai Chen confirms that ancient China displayed an advanced system of international law leading to frequent and numerous interactions among the states, giving each clear knowledge of developments abroad. Zhou feudality had introduced the *li*, a number of rules or protocols of

⁵⁸⁴ Bau 36-7.

international conduct "strikingly similar to, if not identical with, the rules of modern international law." Those rules prescribed the legal equality between states of similar status or power (the rules did not extend to the bordering non-Chinese tribal states), which meant equal sovereignty and equality in arbitration rights, in treaty negotiation and in consent, among other things. Disputes among states could be brought before a Zhou court and there are written records of rulings dating as far back as the late 7th century B.C.⁵⁸⁵

The prevalence of diplomatic exchanges is also evident in the frequent practice of sharing statesmen and political experts between states. Not only did states regularly share one same statesman to handle certain issues on behalf of several states, but statesmen originally working for one state also often moved on to serving in another state, after being sacked, for example, and were welcomed with open arms by their new employer in the hopes they would reveal the neighbor's strategies and secrets. All seven states engaged in these practices and thus learned about each other's power and progress, and Qin in particular benefited from foreign dignitaries at its employ. Shang Yang, the architect of Qin's monumental internal reforms, for instance, was a native of Wei, where he had witnessed Marquis Wen's innovative 4th century B.C. transformations, which served as his inspiration in engineering Qin's reforms.⁵⁸⁶ There is no doubt, therefore, that the seven Warring States were fully aware of each other's political, economic, and military developments, including the rise of Qin. Physical communication hurdles did not prevent them from acting.

⁵⁸⁵ Samuel Shih-Tsai Chen, "The Equality of States in Ancient China," in *The American Journal of International Law* 35.4 (Oct. 1941), 641-4.

⁵⁸⁶ Lewis 633.

2. Misperception (IV1.2)

It is much more likely that Qin's potential balancers drew wrong conclusions about its rise because they misperceived the Qin threat. Although a member of the Zhou feudal network, Qin was a relative newcomer to the ancient Chinese international scene, and as such, it was still considered by many of the more established states as a state of barbarians. Chen emphasizes that as recently as the Spring and Autumn Period, Qin and Chu, another state on the periphery of the Zhou Dynasty, were viewed as barbarian. Barbarian status was determined by behavior in ancient China and not by race or ethnicity. Although all ancient Chinese states share the same or a similar race, including those like Qin or Chu who stood at the edge of the Zhou feudal boundaries, their later economic and political development, different fighting customs, or relative cultural backwardness compared to the more established states sufficed to earn them a barbarian status. As a result, there were numerous cases of racially Chinese states considered "barbarous" by the feudal hierarchy, and several cases of racially non-Chinese, tribal states considered Chinese. Qin belonged to the former category, and such a reputation was hard to shake.⁵⁸⁷

We also know from Chen that those considered barbarian were set in a category apart, not recognized as serious contenders and even frequently dismissed altogether as inferior. They were generally denied the benefits of the "li" rules of international interaction. The Guliang Zhuan, one of the classic historical works of ancient China published during the Han Dynasty but passed down orally from the Warring States, eloquently exemplifies the dismissive, condescending attitude the Chinese states held toward those they considered barbarians: "the armed strifes between the Chinese and the barbarous states could not be termed as wars; in all cases it was simply that the Chinese defeated the barbarians." Because barbarian states did not act in conformity with the Chinese states' customs, the Guliang Zhuan adds, "there was neither right

⁵⁸⁷ Ibid., 650.

nor wrong as far as the barbarians were concerned." Legally a conflict between a barbarian and a Chinese state was therefore not a war and barbarian captives taken during such a conflict were for example not given the protections generally granted to prisoners of war from Chinese states.⁵⁸⁸ Given the extent of the prejudice, then, even after Qin became officially integrated into the Zhou family of states and interacted with its feudal neighbors following the rules of the "li", it continued to bear the stigma of its barbarous origins. It is therefore more than likely that Qin's threat potential was underestimated by the other states, and that Qin's rising power was not taken seriously despite the mounting evidence of its rapid growth, especially at the beginning of Qin's rise when it was not yet well accepted within the network of Chinese states.

3. Deliberate Deception (IV1.3)

The potential balancers' misperception of the Qin threat was reinforced by Qin's pervasive use of deception to dupe its potential balancers and conceal its real power. Trickery and ruses of all sorts were endemic to ancient Chinese international practices, both strategically and tactically: states did not hesitate to connive behind an ally's back or to kidnap and bribe foreign diplomats or military officers, for example. Yet Qin became particularly skillful in deceiving its potential enemies, and went to great lengths to wreak havoc in foreign states. Both government officials and administrators and military commanders were targets. The Zhan Guo Ce (Intrigues of the Warring States), a historical compilation of records from each of the Warring States written between the 3rd and 1st centuries B.C., provides numerous, often caustic, examples. T'ien Hsin, an advisor to King Hui of Qin, who ruled between 337 and 311 B.C., reportedly quoted to the king a passage from a Chinese classic, the Book of Chou, saying that "beautiful

⁵⁸⁸ *Guliang Zhuan (The Commentary of Guliang on the Spring and Autumn Annals)*, attr. Guliang Xi or his disciple Zi-Xia, ca. 3rd cent. A.D., in *Harvard-Yenching Institute Sinological Index Series* (Beijing: Harvard-Yenching Institute, 1937), VI.1 (146) and VIII.XII (232), qt. Chen 649.

women can tangle a tongue." He then suggested sending beautiful women as presents to enemy kings before a battle to distract their minds.⁵⁸⁹ After thus 'preparing' the enemies for battle, Qin leaders also used deception during warfare. They frequently feigned indirection, used surprise, manipulated the enemy toward disadvantageous terrain or positions, etc. They also recurrently generated disloyalty within the enemy's rank by offering bribes to military officers, often even revealing the bribe post-facto to their superior in order to force a change of commander that would disrupt the enemy's organization and morale. Qin employed this trick against Wei in 340 B.C., against Zhao in 260 and again in 229 B.C., and against the anti-Qin alliance in 247 B.C., each time forcing the opponent to sack an able general and leading to the enemy's collapse.

One of the most telling examples of Qin deception occurred against Qi. When Qi became Qin's strongest rival in 341 B.C., Qin managed to convince Qi of its benevolence and coax Qi into abandoning its ambitions by offering King Min of Qi the title of Eastern Emperor of China, while Qin would feign to content itself with the title of Western Emperor. The title and official recognition of power it embodied served to flatter Qi while the division of China into two spheres of influence showed Qin's apparent restraint and willingness to settle. King Min fell for the trick, ignoring that at the same time Qin was plotting an attack on Qi with other states. In order to win more time to fully develop their plan, Qin officials then bribed a number of Qi leaders so they would halt any further increases in military capabilities and refrain from supporting Qin enemies. As a result Qi suffered severe defeats and King Min was captured. Although the state survived as an independent unit until the later Qin unification wars, it never regained its power and was never a serious contender against Qin again.⁵⁹⁰

⁵⁸⁹ Chan-kuo Ts'e (*Intrigues of the Warring States*), transl. and ed. by J.I. Crump, *Michigan Monographs in Chinese Studies Series 77* (Ann Arbor, MI: The University of Michigan Press, 1996), Ch'in 53 (88).

⁵⁹⁰ Hui 72, 89.

Even though Qin repeated such deceptive schemes *ad libitum* throughout its rise, surprisingly few potential balancers learned from the ill fates of their colleagues or even their own. Even fewer attempted to give Qin a taste of its own medicine, despite the success that might entail. After its defeat by Qin in 260 B.C., Zhao saved itself from conquest by creating a rift between Qin's highest military commander, Bo Qi, and Qin's chief minister, Fan Sui, by convincing Fan Sui that the successful Bo Qi would fare better than him with honors and rankings after the war. To avoid another victory by Bo Qi, Fan Sui called off the conquest.⁵⁹¹ While Zhao was eventually conquered by Qin during the final wars of unification in 229 B.C., this example shows that had the Warring States been more proactive in first deciphering, then embracing and persistently using Qin's tricks, they would have enhanced their chances of survival and perhaps even defeated Qin at its own game. Instead, both misperception and Qin's deliberate deception contributed to their inability to react. But defective communication alone was not responsible for the potential balancers' lack of action. Rather, their recurrent collective action problems prevented them from acting in concert.

b. Collective Inaction (IV2)

The potential balancer's lack of trust in one another was the main root of their inability to act together against Qin (IV2.2). The ancient Chinese states' attraction to short-term gains at the expense of their neighbors generated severe trust issues that led them in many cases to forego common action against Qin altogether. I found little or no evidence of insufficient interest (IV2.3) or buckpassing (IV2.4) as a cause of inaction—Qin's neighbors were constantly engaged in conflicts and other interactions with each other, so they showed definite interest in the

⁵⁹¹ Ibid., 106-7.

region's developments and were certainly not hesitant in participating and bearing the cost of intervention when their interest was at stake. However, Qin clearly used its deceptive communication skills in indirect fashion (IV2.1) to heighten its potential balancers' already rampant lack of trust and prevent their alliance. Qin extensively used of divide-and-conquer strategies to tear potential alliances apart before they were even formed. As Hui points out, "Qin further weakened the balance-of-power mechanism by the divide-and-conquer strategy," and "as a result, anti-Qin alliances formed very slowly and infrequently." For example, when Qi became Qin's main competitor after the fall of Wei in 341 B.C., Qin used the jealousy of other states toward Qi to stir them against Qi and form a vast anti-Qi alliance in 284 B.C. that kept Qi at bay.⁵⁹²

The success of Qin's divide-and-conquer strategies stems from their clever exploitation of the underlying problem of trust, which without the intervention of Qin already poisoned most attempts at external balancing. Throughout the rise of Qin, the six major states of the region—Wei, Zhao, Han, Qi, Chu, and Yan—were constantly quarrelling amongst each other and focused on scoring points against one another while ignoring Qin's rising power. In fact, constant conflict seems to have dominated ancient Chinese history for centuries, in large part fueled by the region's layout as one compact continental bloc, which increased states' suspicion of their neighbors, shortened mobilization times, and hindered peaceful crisis diffusion. John Fairbank notes that during the Spring and Autumn Period already, at least 110 states disappeared or were conquered, and only 22 states survived the two-and-a-half century period.⁵⁹³ Kiser and Cai calculated that there were only 38 years of peace between 722 and 464 B.C. The Warring States period itself hardly bears a brighter record. Between 463 and 221 B.C. when Qin achieved

⁵⁹² Hui 72, 79.

⁵⁹³ John K. Fairbank, "Introduction: Varieties of the Chinese Military Experience," in John K. Fairbank and Frank A. Kiernan, *Chinese Ways in Warfare* (Cambridge, MA: Harvard University Press, 1974), 5.

hegemony, Kiser and Cai counted only 89 years of peace. Although it appears that wars occurred less frequently during the rise of Qin than during the preceding Spring and Autumn Period, "they lasted much longer and were more intense and larger in scale," thus pointing to a worsening of trust issues rather than an improvement. In all, during the five centuries preceding Qin's hegemonic success, the states of ancient China were at war 75% of the time. It is therefore hardly surprising that these states' suspicions of each other was strong enough to derail any effort to cooperate to stop the rise of Qin.⁵⁹⁴

The disenchanted tone of the original Chinese sources when they describe the utter chaos of international relations during Qin's rise is particularly telling. Famed Han Dynasty historian Sima Qian, who wrote a history of Qin at the turn of the 2nd century B.C., lamented that by the beginning of Qin's rise in the mid-4th century B.C., "the Zhou royal house had sunk into insignificance and the feudal lords ruled by force, wrangling with one another and annexing each other's lands."⁵⁹⁵ The Zhan Guo Ce similarly relates that SuCh'in, a military strategist of King Hui of Qin (r. 337 to 311 B.C.), enumerated a long list of recent wars during an audience with the king and concluded: "Indeed, who has not gone to war?"⁵⁹⁶ Despite such dire observations, political experts and rulers of the time were aware of the dangers of pursuing short term gains and ignoring larger threats. Han Feizi, a 3rd century B.C. philosopher from the state of Han who witnessed the events leading to Qin's hegemony and might have sought to warn his relative, the King of Han, through his writings, lists the ten biggest faults he believed a ruler could make. "To fix your eye on a petty gain and thereby lose a larger one" ranks second on his list. He cites the example of the Duke of Yü, who accepted jade and horses from Qin in exchange for their right

⁵⁹⁴ Kiser and Cai 519-20.

⁵⁹⁵ Sima Qian, *Records of the Grand Historian*, ca. 109-91 B.C., transl. Burton Watson (Hong Kong: The Chinese University of Hong Kong and Columbia University Press, 1993), 23.

⁵⁹⁶ *Chan-kuo Ts'ue*, Ch'in 47 (82).

of passage on his land and was conquered. The Duke's longing for a small gain destroyed everything he had. Han Feizi also tells the story of Chih Po, who tried to seize all of his allies' territories, but only succeeded in ganging them up against him. In retaliation his former allies killed him and took his possessions. Han Feizi concludes, "therefore I say, by fixing your eyes on a petty gain, you may deprive yourself of a much larger one." He adds that "to be greedy ... and too fond of profit opens the way to the destruction of the state and your own demise." But no Warring State paid attention to his warnings.⁵⁹⁷

Instead, the targets of aggression were multiple. The six main potential balancers all took over secondary powers to increase their territory and power vis-à-vis the others: Han conquered Zheng in 375 B.C.; Wei conquered Wey in 254 B.C.; Chu conquered Lu in 254 B.C.; Zhao conquered Zhongshan in 296 B.C.; and Qi conquered Song in 286; and these are only a few among many examples. The same six states also frequently seized part of each other's territory, including cities and sometimes whole counties. Out of 96 wars involving the major powers between 356 and 221 B.C., 27 "involved mutual attacks among the six states." This "systemic phenomenon of mutual aggression" played into the hands of Qin since its enemies exhausted each other and invaded states already at war with others, thus opening new fronts these states could not handle. The primacy of immediate gain in the potential balancers' strategies is particularly evidenced by their frequent tendency to take advantage of a neighbor's defeat by Qin by attacking it after Qin's withdrawal and gaining more territory out of the exhausted neighbor, rather than joining it to repel Qin from the area. For instance, Qi attacked Wei and Zhao in 317 B.C. after they had been defeated by Qin just months before. Yan similarly attacked Zhao in 251-252 B.C. after Qin had inflicted considerable damage on Zhao a few years prior. As a result

⁵⁹⁷ Han Fei Tzu, *Basic Writings*, ca. mid-3rd cent. B.C., transl. Burton Watson (New York: Columbia University Press, 1964), Section X, 49-51, 56-62.

of such behavior, Qin's potential balancers were constantly distrustful of the others' agenda and extremely reluctant to take the risk of cooperating to stop Qin.⁵⁹⁸

The improbable turmoil created by Chu and its neighbors Han and Wei illustrates why most were reluctant to commit to an alliance. Chu, which with Qi was one of the strongest powers after Qin's defeat of Wei in 341 B.C. and thus was an ideal candidate to lead a balancing alliance, was instead renowned for attacking friends and foes alike in a erratic quest for territorial aggrandizement that triggered a rapid and utterly unpredictable succession of wars in the decade around the turn of the 3rd century B.C. By then, a coalition of major states could still have "easily" defeated Qin if they acted in concert, even though Qin had defeated most of them separately before. Instead, Chu and its neighbors Han and Wei kept battling each other, which Qin made the most of by supporting whomever was at a disadvantage. Qin helped Chu when it first attacked Han and Wei, then it helped Han and Wei when they attacked Chu in return, and later it made peace with Chu and helped it invade Han in 304 B.C. A year later Han, with the support of Qi and Wei, attacked Chu to avenge itself. In one instance Wei even paid Qin with a district of land in exchange for not intervening on Chu's behalf, even though it thus directly increased the power of the rising hegemon.⁵⁹⁹

When Qin then invaded Chu during a long war that ended in 298 B.C., the other states watched idly, only too happy to see Chu punished for its previous aggressions. Qin turned against Han and Wei next, from 294-286 B.C., and though one would imagine Chu had understood the danger by then, Chu watched on without intervening in retaliation for Han and Wei's earlier failure to assist it against Qin. Unsurprisingly, Qin turned against Chu again in 280 B.C., and instead of helping Chu, Han and Wei made plans for joining Qin in its next round of

⁵⁹⁸ Hui 75-6.

⁵⁹⁹ Hui 71-2.

attacks against Chu. However, Chu managed to negotiate with Qin and forestall the planned attack, and Qin instead turned once more against Han and Wei. In hindsight, Han, Wei, and Chu's inability to realize that their squabbles played right into Qin's hegemonic strategy and that an alliance against Qin would have instead preserved them seems inexplicable. But the lack of trust ran so deep that it seems to have undermined even their basic survival instincts.⁶⁰⁰

Even more baffling is the fact that no one learned from those events of the turn of the 3rd century B.C. The same lack of trust and short term gain strategies continued to be pursued even as Qin was exponentially gaining in power. In the 280s, the other main contender against Qin's rise, Qi, managed to rapidly increase its relative power and occupy Song, a secondary state to its east. But "Qi's success ... immediately made it the target of collective fear and suspicion," which Qin exploited by encouraging anti-Qi movements. Even Yan, a traditional friend of Qi, abandoned Qi and instead attacked it by surprise in 284 B.C from the north. Simultaneously, Qin, Zhao, Han and Wei attacked Qi from the west. Chu was the only state to side with Qi, but it was already seriously weakened by then and did not actively help Qi but just reoccupied some land it had previously lost. In that year Qi's armies were totally destroyed, its king was killed, almost the entire state was occupied, and Qi stopped being a contender against Qin altogether. It would have probably been the best candidate for balancing Qin had it not been for the greed and jealousy of the other potential balancers.⁶⁰¹ After the fall of Qi, Zhao remained the "only great power still capable of resisting Qin" and putting together an anti-Qin alliance. However, unable to trust any of the remaining powers to assist it, Zhao resolved to face Qin alone, an unwise endeavor since by then, Qin had developed overwhelming power. Although Zhao came up with an ingenious plan of attack to reach Qin via its northern barbarian border where it would expect

⁶⁰⁰ *Chan-kuo Ts'ue*, Ch'in 49 (85-6).

⁶⁰¹ Lewis 637-8.

it less, Zhao never got a chance to put the plan into effect. Qin attacked promptly, and after fierce Zhao resistance, came back with a stronger force and defeated Zhao's armies in 260 B.C. at the Battle of Changping. After that date, Zhao too became unable to contend with Qin.⁶⁰² Lack of trust and focus on short-term gains were thus rampant in ancient China, preventing in many instances the forging of balancing alliances altogether.

c. Laggard Balancing (IV3)

Even when some states managed to cooperate against Qin, the same trust issues and immediate gain concerns prevented these alliances from success in stopping Qin (IV3.2). All collaborative efforts undertaken by the balancers eventually fell short, being too late, too weak, or simply insufficient to seriously undermine Qin's rise. As Hui suggests, "in the end, the balance of power merely slowed but did not check Qin's rise to universal domination." There is little evidence of insufficient interest (IV3.3) and none of buckpassing (IV3.4) as a cause of lethargic balancing—all major ancient Chinese states were involved in conflicts throughout the region and paid attention to developments in the neighboring states, including Qin. The rising hegemon's deceptive communication techniques (IV3.1), particularly its divide-and-conquer tactics, contributed to wreaking havoc in the potential balancers' cooperation attempts, but deception played an auxiliary role, succeeding primarily because of the underlying trust issues that were so rampant they would have most likely derailed alliances attempts regardless of Qin intervention. Hui captures the core of the trust problem when she writes that the six major states'

... balancing efforts were not effective because targets of domination shared the self-interest motivation, pursued their own optimistic expansion, and thus had conflicts of interests among themselves... As a result, ... anti-Qin alliances did not have enough

⁶⁰² Hui 72-3.

members to overpower Qin, they rarely had unified command, and they readily disintegrated.⁶⁰³

Lack of trust sabotaged otherwise promising countervailing alliances. As the six main neighbors of Qin became aware of the rising hegemon's growth in the late 4th century, several of them participated in a balancing strategy called *hezong* or vertical alliance because it sought to erect a vertical barrier to Qin's spread that would contain Qin to western China by linking Yan and Zhao in the north with Han, Wei and Qi in central China and Chu in the south. This strategy was pursued for half a century until the mid-3rd century B.C. The vertical alliance system was particularly promising because it was engineered by SuCh'in, the Qin military strategist who had counseled King Hui of Qin for years and was the architect of Qin's own expansion strategy. After SuCh'in failed to convince King Hui to take certain actions he recommended, he emigrated to Zhao and began advising the King of Zhao. The vertical balancing strategy was designed with specific insight into Qin's own strategies and tailored as a direct response to those strategies and thus should have produced great results. Indeed, the allies managed to put together eight collective endeavors within that half centuries: four offensive movements specifically meant to stop Qin's hegemonic expansion (318-317 B.C., 298-296 B.C., 287 B.C., and 241 B.C.) and four defensive responses to Qin attacks (294-286 B.C., 276-274 B.C., 259-257 B.C., and 247 B.C.).⁶⁰⁴

Out of the eight common efforts, however, five were either defeated by Qin or dissolved because of internal problems. During the 318-317 B.C. war, Wei took the lead in trying to get the allies together but faced much difficulty because its partners were still suspicious of its own intentions even though Wei had been weakened by its 341 B.C. defeat and therefore slowed down the process. This alliance attempt also showcases the only instances of insufficient

⁶⁰³ Ibid., 79.

⁶⁰⁴ Chan-kuo Ts'e, Ch'in 47 (84); Hui 67-68.

balancing because of a lack of interest. Once the alliance was formed, Qi, then one of the most powerful potential balancers, refused to actively participate because it did not believe Qin was yet a sufficiently threatening force. Similarly, Yan joined the alliance but did not provide troops because it failed to see Qin as a large threat, being the most distant state from Qin. Regardless of Qi and Yan's participation, the alliance, which regrouped Han, Wei, and Zhao, could have fared well because it was more numerous than Qin. However, because the allies did not set up a unified command, Qin was able to counter and defeat each army separately. In addition, the alliance broke down immediately after the defeat as Wei abandoned Han and Zhao to conclude a separate, more advantageous peace with Qin in 317 B.C. Qi then attacked the weakened Wei and Zhao, instead of lending them support, to score easy territorial gains. The 294-286 B.C. and 276-274 B.C. wars brought together only Wei and Han, in an attempt to fend off repeated Qin attacks. But again their armies were "uncoordinated," and were thus easy for Qin to divide up and defeat separately.⁶⁰⁵ At the Battle of Mount Yi-Ch'üeh in 293 B.C., for example, Qin scored a massive victory against the combined armies of Han and Wei. Qin general Po Ch'i explains:

In the fighting at Yi-Ch'üeh ... Han did not want to be the first to use her forces. Wei was relying on Han's best troops which she hoped to make her vanguard. Both armies bickered over petty advantage and their strength was not united. Because of this I could use diversionary forces to make Han keep to her formation while our main force and crack troops took Wei by surprise. With the Wei army defeated, the Han force scattered, and ... we drove them into retreat. This was why I was able to succeed there. I ... ma[d]e capital of [my enemies'] inherent tendencies.⁶⁰⁶

In 287 B.C., Han and Wei managed to convince Zhao, Qi, and Yan to join the alliance. But several "conflicts of interest" arose between the allies, and the partnership fell apart before

⁶⁰⁵ Hui 69.

⁶⁰⁶ *Chan-kuo Ts'üeh*, Ch'in 104 (135).

achieving any military action. Finally, in 241 B.C., Zhao, Chu, Han, Wei, and Yan came together in what was perhaps the most sweeping alliance to date, but Qi, then the most powerful potential hegemon, once again refused to participate, likely because it was on good terms with Qin at the time, and doomed the alliance's effort by depriving it of precious resources against the now overwhelming power of Qin. Moreover, the 241 B.C. again alliance lacked a unified command. Therefore, when Qin attacked the Chu camp at night, Chu decided to leave the alliance unexpectedly, which led the others to retreat, too, and dissolve the alliance.⁶⁰⁷

Trust issues and immediate gain motivations were so profound that they even derailed the allies' more successful efforts. In the end, although the allies managed to score three victories against Qin, their successes did not prove decisive. In 298-296 B.C., the collective efforts of Qi, Han, and Wei enabled them to take control of the strategic Hangu Pass in the mountains protecting Qin's eastern border, and even to seize some territory from the rising hegemon. But the three were unable to agree on how to divide the conquered land amongst themselves. Most of it went to Han and Wei, and Qi, as the biggest contributor to the war effort, felt cheated by its partners. As a result Qi decided to sign a separate peace with Qin and to pursue territorial expansion on its own, even though together, Han, Wei and Qi could continued their successful incursions against the rising hegemon. In 259-257 B.C., Chu and Wei helped Zhao break up Qin's siege of the Zhao capital of Handan. However, they did not go after Qin's routed army, so the victory was only partial. The third allied success, in 247 B.C., was especially promising. Wei managed to gather support from Chu, Han, Yan, and Zhao against Qin's siege of its capital, Daliang, and for the first time, the allies operated under the unified command of the Wei general who had been instrumental in winning the siege of Handan. This time, after breaking the siege, the Wei general pursued the retreating Qin army all the way to the banks of the Yellow River.

⁶⁰⁷ Hui 69.

But as they were cornered, Qin resorted to deception to reignite the allies' deep-seated trust issues. Qin bribed a Wei official to spread the rumor that the Wei general leading the allies, a brother of the King of Wei, sought the throne of Wei. The King of Wei promptly removed his brother from command and the alliance faltered as news of its uniting figure's alleged treason spread. Once again, a balancing alliance that could have undermined Qin's rise fell short because of its members' inability to transcend their mistrust.⁶⁰⁸

In reality, the bribery displayed in the 247 B.C. war played a large role in amplifying the alliance's trust problems, and Qin simply managed to exploit it. The Zhan Guo Ce notes a suggestion of King Hui of Qin to one of his advisers, Han-Ch'üan Tzu : "Chao [i.e., Zhao] depends on the number of states allied to her and so had sent SuCh'in with bribes to them to pay for treaties." But, the king adds, "uniting the lords [that way] is as unlikely as having several game cocks sharing a single perch—that much is certainly clear."⁶⁰⁹ Chia Yi also recognizes that vertical alliance participation was by and large enticed by bribes, hardly the way to secure reliable allies. After Qin began acquiring strategic lands west with Ba and Shu and south with Han-chung, Chia Yi writes, plus east in Chu and north in barbarian territories, "the other feudal lords in alarm came together in council to devise some plan to weaken Ch'in [i.e., Qin]" and they "spar[ed] nothing in gifts of precious objects and rich lands to induce men from all over the [former Zhou] empire to come and join them in the Vertical Alliance." Thus, short term gain rather than security was the main incentive for joining the coalition, guaranteeing the untrustworthiness of its members, who would not hesitate to follow alternative paths if those led to greater shorter gains. Chia Yi notes that some allies even turned to bandwagoning with Qin when it seemed more profitable: "the Vertical Alliance collapsed, its treaties came to naught and

⁶⁰⁸ Lewis 636-7; Hui 70; Sima Qian 29.

⁶⁰⁹ *Chan-kuo Ts'ue*, Ch'in 48 (88).

the various states hastened to present Ch'in [i.e., Qin] with parts of their territories as bribes..."⁶¹⁰

The prevalence of bribery within the alliances just serves to prove the primacy of the immediate gains mentality over alliance partners and over the existential threat posed by Qin. The problem of trust constantly recurred because the "symmetry" of the alliances was "illusory." The vertical alliance was a combination of independent states but the alliance partners never displayed equal power. Unavoidably, one state was always more powerful than the others and dominated the coalition, leading to rivalries that would undermine the common effort. Every time the alliance was resurrected, the same pattern emerged. Lewis adds, "this asymmetry meant that ... success for a vertical alliance invariably led to its breakup, because it strengthened the preeminent ally and led the others to turn against it. The logic of the practitioners of [the vertical] alliance made any protracted coalition against Qin impossible."⁶¹¹ As a consequence, the risk of being abandoned by one's allies or even destroyed by a reversal of alliances and bearing the brunt of the balancing costs rendered the dominant states hesitant to commit to leading a vertical alliance. Qi, for example, who was one of the most powerful states of the region during the rise of Qin, refused several times to be drawn into vertical alliances, and promptly left the alliance once after its participation was not rewarded at its just value. Han Feizi sums up the dilemma facing would-be vertical alliance leaders like Qi: "When you face a powerful enemy [like Qin], you cannot always be sure that your allies will remain loyal. And if your allies break with you, you will be at the mercy of the powerful state."⁶¹² Being a part of the vertical alliance system was a dangerous gamble for any member. Rather than one alliance system, it was rather an "interplay

⁶¹⁰ Chia Yi 46-7.

⁶¹¹ Lewis 634.

⁶¹² Han Fei Tzu 112-3.

of alliances in which each state sided with whatever ally proved useful and changed allies as soon as the balance of power shifted."⁶¹³

Despite regular diplomatic meetings between the allies to settle their borders and stabilize their relationships, the lure of short term gains led them to revert time and again to their habitual, individual profit-driven ventures and the resulting "chaos of everchanging coalitions." The instability of the vertical alliance allegiances might even have been encouraged by the diplomats despite their apparent good intentions. The vertical alliance system gave rise to a new class of career diplomats, the "masters of persuasion," for whom negotiation and intricate political deals represented a livelihood. Many of them traveled from state to state and often even represented several states at a time. Through "their mastery of stratagem and language," they guided rulers through the constant intrigue of shifting alliances and contributed to increased alliance fluidity because of their innate lack of a strong loyalty and practice of alliance making as a sport or art.⁶¹⁴ SuCh'in, for example, originally from Qin, alternatively or simultaneously advised Zhao, Yan, and Qi, even at times when their interests were contradictory, spurring their disputes. Kung T'o, similarly, advised King Nan, the last Zhou ruler who ruled from 314 to 256 B.C. and still controlled a small territory around the former Zhou Dynasty capital of Luoyang that for a secondary state like Zhou, it was unwise to depend upon his protectors Han and Wei although they had not harmed Zhou. "It would be best to send [diplomat] ChouTsui to arrange a secret alliance with Chao [i.e., Zhao] to prepare your state against Ch'u [i.e., Chu]" instead. Kung T'o thus directly encouraged the King of Zhou to spread trouble among the vertical alliance members for no apparent reason.⁶¹⁵

⁶¹³ Lewis 616-7, 619.

⁶¹⁴ Ibid., 633.

⁶¹⁵ *Chan-kuo Ts'e*, Chou 19 (66).

Several ancient Chinese leaders seemed bitterly aware of the problem, but were powerless in trying to change their state's and their partners' mentality. Even SuCh'in, the engineer of the vertical alliance on behalf of Zhao, lamented that although he succeeded in bringing the alliance together, "arms and armor were never put aside" between its members. Similarly, one of Zhao's delegates sent to negotiate with King Chiao of Qin in the second half of the 3rd century B.C. recognized that it would be impossible to sustain an alliance long enough to defeat Qin. "I know that they [i.e., the potential balancers] cannot ally themselves and oppose Ch'in [i.e., Qin] successfully," Su-Tzu reportedly said. "To unite [them] ... is the height of folly." The rising hegemon was similarly aware of its potential balancers' trust and collective action problems and found it therefore easy to exploit them to facilitate its rise. The Zhan Guo Ce highlights Qin minister Marquis Yin's confidence as he reassured Qin King Chiao (Zhao), who ruled from 306 to 251 B.C., about the low risk posed by the vertical alliance. "There is no need to worry over this, your Majesty ... They gather now to make plans for an attack against Ch'in [i.e., Qin] because each seeks wealth and fame for himself," Yin purportedly said. His prediction turned out right and soon, the Zhan Guo Ce concludes, "all the officers of the empire [i.e., leaders of the six potential balancers] were fighting each other."⁶¹⁶

The mid-3rd century witnessed the last attempts of the vertical alliance system, which all failed. "While all the six states were afraid of Ch'in [i.e., Qin], they hated each other even more," historian Dun Li writes, resulting in the final collapse of the vertical alliance.⁶¹⁷ By then, however, since "no single state could match it [Qin]" and it had become clear that because of their inherent trust problems and focus on immediate gains, "the issue was no longer in doubt"

⁶¹⁶ Ibid., Ch'in 47 (82), 106 (139-40), 101 (131).

⁶¹⁷ Li 58.

that potential balancers were incapable of holding a coalition together for a sufficient period of time to stop the rising Qin.⁶¹⁸

d. Bandwagoning (IV4)

The constant shifts of allegiances within the vertical alliance pattern and alliance members' suspicion of each other and attraction to gains over their partners proved highly beneficial to Qin not only by reducing their combined threat of successful balancing, but also by allowing Qin to repeatedly attract disgruntled vertical alliance members into its own orbit. Qin's response to the *hezong* or vertical alliance strategy of its balancers was to create the *lianheng* or horizontal alliance, an east-west bandwagoning strategy meant to break the vertical north-south line confining it to western China. Qin alternatively sought to obtain Han, Wei, and Qi's cooperation. Given the balancers' widespread focus on short-term gains, their attraction to Qin remained exclusively profit-driven (IV4.2). There seems to be no instance of a major state bandwagoning with Qin out of fear (IV4.1) or following a mixed fear-profit incentive (IV4.3) until the very end of the period and the final wars of annexation in the 230s B.C., when Qin was just about to reach hegemony and any countervailing alliance, given the weakness of the remaining states, was virtually doomed to fail. Then, several states, including Han and Wei, "followed a policy of appeasement" as they "became so frightened" and began giving up territory to Qin without fighting.⁶¹⁹

The prevalence of the constant quest for profit combined with the lack of trust among vertical alliance members made it very easy for Qin to coax disaffected balancers into its horizontal alliance by using a variety of deceptive techniques: divide-and-conquer tactics to play

⁶¹⁸ Lewis 641.

⁶¹⁹ Hui 87.

balancers off each other, threats, bribes of corrupt officials in the target state, and even displays of force to show the target state how much profit Qin could extract. Qin repeatedly resorted to lying in diplomatic meetings, engaged in espionage to learn the target's weaknesses and identify the officials amenable to bribery, fomented political and territorial conflicts between the balancers, and even used marriage as a means of rapprochement with targeted balancers. Qin's schemes resulted in balancers' switches of allegiance from the vertical to the horizontal alliance, each time weakening the balancing side and reinforcing the rising hegemon. As is generally the case, though, no bandwagoner obtained the profit it had hoped; quite to the contrary, bandwagoning with Qin proved very costly.

Wei, the most powerful balancer in the mid-4th century B.C. at the beginning of Qin's rise, owes its downfall to a bandwagoning temptation. When in 344 B.C., Wei was planning a preventive war against Qin, Qin's lead minister Shang Yang learned of the plans and obtained an audience with Marquis Hui, the ruler of Wei. Playing the classic divide-and-conquer card, Shang Yang convinced Marquis Hui that Qin meant no harm to Wei and that Qi and Chu posed a greater immediate danger, and encouraged Hui to renounce its attack and accommodate Qin. As a bait, Shang Yang offered Hui the title of feudal king in replacement of the weakened Zhou king. The Marquis jumped at the occasion to enhance his status. When the other states became alarmed by his new title and boycotted the fait-accomplis, Hui became convinced that his new friend Shang Yang had been right about his other neighbors being greater threats. Completely reversing his original plans, he turned against Han, the weakest of his neighbors, which prompted the powerful Qi to defend Han. Wei suffered a devastating defeat in 341 B.C. and lost its leading status among the potential balancers. Qin had thus masterfully managed to rid itself of

the threat of Wei without taking any military action itself, by simply swaying Wei towards its side and stirring trouble among the other potential balancers.⁶²⁰

The price of Wei's friendship with Qin proved very high. Once Wei was weakened by its defeat against Qi, Qin invaded it in 340 B.C. Still, Wei had not learnt the lesson of Qin's treachery. When Shang Yang offered his former friend Prince Ang, who led the Wei army, to sign a peace settlement rather than obliterate Wei, Ang agreed to come over to the Qin camp to negotiate. When he arrived, Shang Yang had him captured. Without its leader, the Wei army fell into disarray and was forced to retreat, handing over parts of Wei's territory to Qin. Had Marquis Hui not been tempted by the title offered to him by Qin, his preventive war plans against Qin would have had an excellent chance of nipping Qin's hegemonic ascent in the bud, since in the 340s Qin was still early in its growth. The importance of Wei's mistake cannot be understated, as "Wei's decline was a turning point leading to Ch'in's [Qin's] eventual unification of the other six states."⁶²¹

The other potential balancers did not learn from Wei's mishaps either, and Qin repeated the same techniques with the next challengers. In 312 B.C., Chu and Qi, the two strongest potential balancers since Wei's collapse in 341 B.C., stood on the verge of an alliance, which presented a clear danger to Qin. To prevent the alliance from solidifying, Qin's chief minister Zhang Yi offered Chu's King Huai 600 *li* of territory—slightly short of 200 miles—if Chu broke off the alliance with Qi. Unable to resist the territorial gain, King Huai accepted. Unsurprisingly, Qin failed to keep its part of the deal and delivered only an insulting 6 *li* of territory. Chu subsequently launched two campaigns against Qin, but on its own it did not fare well and in the end Qin moved 600 *li* into Chu and took over two strategically located Chu cities that opened up

⁶²⁰ Bau 26-7; Hui 67-8, 70.

⁶²¹ Bau 26.

the central plains to its reach. A few decades later, Qin administered weakened Chu its coup de grace by invading the entire western part of the kingdom, including the capital of Ying. Thereafter Chu only survived on a stump of its former territory in the east. Again, if Chu had not succumbed to the temptation of bandwagoning, a Chu-Qi alliance would have without doubt posed a serious hurdle on Qin's path toward hegemony.⁶²²

Once again, some experts at the time were well aware of the dangers of bandwagoning, yet no one seems to have heeded their advice. Han Feizi, the Han philosopher who lived through the rise of Qin, specifically warned states tempted to join the horizontal alliance with Qin, including his own state of Han: "Now when you enter the service of a powerful state, you cannot yet be certain of the practical advantages ... You gain no benefit by entering the Horizontal Alliance in the service of a powerful state, but merely lose territory and undermine the government." Han Feizi cautioned that bandwagoners not only risk territorial costs, but might also lose their sovereignty: "If you enter the service of a powerful state, it will dispatch its own men of authority and take over the offices in your government." He gave the example of the secondary state of Wey that left the vertical alliance to join Qin in 241 B.C. "and in half a year it was ruined."⁶²³ Han Feizi was right to point out the political consequences of joining the horizontal alliance. To participate in the horizontal alliance "meant to be under sway, if only temporarily, of Qin," and to have to submit to Qin's will, as Wei was eventually forced to do after its defeats in the mid-3rd century B.C. Horizontal allies were by no means equals of Qin. But while in the case of the vertical alliance, the asymmetry between allies meant that alliance

⁶²² Hui 71.

⁶²³ Han Fei Tzu 112-4.

success resulted in tension and fragmentation, the preeminence of Qin ensured that "success increased the stability of a horizontal alliance."⁶²⁴

In sum, distrust-induced collective inaction and laggard balancing, aggravated by bandwagoning and Qin's deceptive tactics, virtually immobilized the potential balancers and ruined a number of promising opportunities they had of stopping the rising hegemon. Trust problems and immediate gain concerns may have been particularly acute in ancient China because of the states' geographic proximity, which is a key element of threat perception. "States that are nearby pose a greater threat" to each other and are naturally more suspicious of each other's intentions because their reaction and mobilization times are shortened and their first-move advantages are heightened, which exacerbates the security dilemma.⁶²⁵ As Bau demonstrates, "the aggressive behavior of states is encouraged in a system with centralized political geography." In other words, the closeness of the potential balancers to each other may explain why trust-induced collective inaction and laggard balancing, though present and central in the Mongol and Roman cases, scored exponentially high in the Chinese case.⁶²⁶ The potential balancers' trust problems were also aggravated by the rising hegemon's clever exploitation of the issue, via various deceptive strategies that encouraged the breakup of alliances both before and after they were formed and attracted some potential balancers to make devastating bandwagoning choices. But the balancers and their susceptibility to internal conflict and deception are not alone responsible for Qin's hegemonic success. Had Qin not simultaneously

⁶²⁴ Lewis 634.

⁶²⁵ Quote is from Stephen Walt, "Keeping the World 'Off-Balance': Self-Restraint and U.S. Foreign Policy," in G. John Ikenberry, ed., *America Unrivaled: The Future of the Balance of Power* (Ithaca, NY: Cornell University Press, 2002), 136-7. On first-move advantages and instability, see Stephen Van Evera, *Causes of War: Power and the Roots of Conflict* (Ithaca, NY: Cornell University Press, 1999), chp. 3.

⁶²⁶ Bau 144.

developed a uniquely innovative military and administrative apparatus to complement its expansion plans, it would not have been able to propel its power as far ahead of its competitors as it did.

2. The Rising Hegemon's Side: Innovative Military and State Institutions

The success of Qin over the other ancient Chinese states is surprising because of Qin's status as a newcomer and former "barbarian" state. One would imagine that a more established state in the region, such as Wei or Qi, would be more likely to overcome the balance of power since such states possessed more mature civil and military infrastructures and thus had the foundation necessary to build an empire and support a bid for hegemony. It was Qin, however, starting from scratch, that not only managed to close its gap with the more established states but also went on to surpass and overcome them. In fact, Qin's "backward" background may have not only led the other Chinese states to underestimate it, but also stimulated Qin's creativity because "Ch'in's [i.e., Qin's] relative freedom from the cultural traditions of the more purely 'Chinese' states made it easier to institute radical innovations."⁶²⁷

Indeed, Qin's monumental and rapid growth required an unprecedented amount of military- and state-building and astute capitalization on Qin's assets, beginning with a vast centralization of power that enabled Qin leaders to maximize resource extraction and political control for a more efficient growth plan. Qin's power expansion strategy relied on implementing the political philosophy of legalism, a utilitarian, anti-Confucian philosophy which called for "power ... to be maintained and secured at any cost" and relied on enacting laws to "attain... efficiency in government and orderliness in society," requiring strong centralized government

⁶²⁷ Derk Bodde, *China's First Unifier—A Study of the Ch'in Dynasty as Seen in the Life of Li Ssu* (Leiden: E.J. Brill, 1938), 47.

control. Legalists' goal was to use the state to channel the country's resources and to rationalize military (IV5) as well as economic, social and political (IV6) institutions. Qin rulers were inspired by leading legalist thinker Han Feizi, and Duke Xian (ruled 385-362 B.C.), Duke Xiao (ruled 362-338 B.C.) and Lord Shang Yang (in office in 359-338 B.C.), the architects of the sweeping reforms that enabled Qin's rapid rise, were followers of legalism who practiced the doctrine to its fullest extent.⁶²⁸

a. Military Innovation (IV5)

Qin's military success was tied directly to its adoption of the legalist philosophy, which led Qin leaders to enact a new, unprecedented organization of the military (IV5.1), encourage members to embrace the new system and further contribute to its improvement with a strict disciplinary system (IV5.2), and adopt strategies, tactics, and technologies that capitalized on the new organization (IV5.3).

1. Military Organization (IV5.1)

Qin's novel organization of its military reflects the legalist goal of rationalization and centralization of power, and maximization of resource extraction. Its overarching, and eminently modern theme was that of civilian control over the military. "How to keep the military under control within the social order ... became an early focus" of the Qin, who sought to develop a powerful military instrument while avoiding allowing it to spiral out of control. A number of Qin reforms described in this section, like the household registration system, the collective responsibility system, the parallel military and civilian appointments, and the relocation of

⁶²⁸ Li 91-2.

powerful local families away from their centers of operation, for example, linked military and social responsibilities and were designed as safeguards to keep the increasingly powerful military "subordinate." The simultaneous development of a military bureaucracy, with permanent appointments and "subject to ... controls over personnel and resources through orders from the center," also worked to "contain ... militarism" and encourage growth through non-violent means like diplomacy.⁶²⁹

The first reform leading to the dramatic growth of the Qin military was its complete change of composition. Qin led the Warring States transformation from small, nobility armies to mass infantries recruited by universal conscription. It brought about the "end of the dominance of the warrior nobility and the shift to a state based on the service of the peasant household."⁶³⁰ Qin's new commoner infantry inaugurated the "decreasing role of the chariot which had been the symbol of power and authority, both religious and secular, of the aristocratic elite during the previous centuries." In addition to the peasant armies, Qin also kept a professional corps of elite troops, generally heavy-armored soldiers carrying crossbows, halberds and swords, as well as several days' supplies to increase their independence, and trained to travel fast. The archetype of the Qin elite troops are the famous terra cotta warriors interred with the First Emperor and discovered in the 1970s near Xi'an. Although the combination of a mass infantry of drafted soldiers and a professional army of elite soldiers was originally a Wei idea, Qin was the state that first embraced the idea and implemented it consistently and to its full extent. The other states merely followed and copied, but failed to develop the new system as fast and sweepingly as Qin. Besides the infantry and elite corps, the Qin army also contained a cavalry, which once more it did not invent but quickly adopted from observing its rivals, designed to fight against the

⁶²⁹ Fairbank 4, 9.

⁶³⁰ Lewis 620-1.

peripheral tribes people, who were more often horsemen, though the cavalry made up only a small part of the army. The blend of mass infantry, elite troops, and specialized cavalry rendered the Qin army adaptable to all sorts of enemy and terrain and enabled its commanders to vary strategies and tactics and use an element of surprise in their attacks that was frequently sufficient to secure victory.⁶³¹

The new, democratized mass army fell directly into the legalist goal of resource maximization. Instead of conducting wars with a small, aristocratic force of sometimes rebellious local personalities, the state of Qin could now fully mobilize its docile mass of peasants to engage in total wars. Although we do not know for certain what Qin's military service requirements were, we can guess they were roughly similar to Qin's successor, Han, which required two years of service in peacetime, including one year at home in training and one year in a distant garrison. The sheer number of men under arms rose sharply during the Warring States. During the Spring and Autumn Period, Qin's and most of its neighbors' noble armies counted roughly 10,000 to 30,000 men. As each state gradually adopted Qin-style mass armies during the Warring States Period, even the weaker, secondary states developed armies of 100,000 men. Statistics were kept dutifully in all states because the number of enemy fatalities determined rewards. Though the records are not always reliable because commanders had an incentive to inflate the numbers to collect more rewards (e.g., cutting off ears from one's own dead or local peasants), the change in the order of magnitude is striking. Qin was clearly among the front-runners in terms of numbers. Its accounts list a total capacity of close to one million armored infantry, 1,000 chariots and 10,000 horses, and the possibility of mobilizing as many as 600,000 men for one campaign alone, which are "astounding figures even after discounting for inaccuracy and exaggeration" frequent in ancient Chinese historical records. At their peaks, Qi

⁶³¹ Yates 410.

and Chu, the two most powerful balancers, listed close to one million men also, while Wei had roughly 500,000, Zhao 400,000, Han 300,000, and Yan 100,000.⁶³²

Mass mobilization and increased army sizes naturally led to soaring numbers of casualties during Qin's rise. The aristocratic armies of the Spring and Autumn Period were generally more obedient to chivalry rules and less likely to engage in mass killings. But the development of mass armies led by professional generals and Qin's quest for land acquisition for expansion at all costs, coupled with legalism's focus on material gain and disregard for ethics resulted in the conduct of total war with a dramatic increase in lethality. For instance, the 341 B.C. Battle of Ma-Ling at the beginning of Qin's rise led to 100,000 Wei deaths, but already in 293 B.C. the Battle of Yi-Ch'üeh resulted in 240,000 Han and Wei deaths. By 260 B.C., the Qin war against Zhao led to 450,000 Zhao casualties and the Battle of Ch'ang-p'ing left 450,000 Chu deaths. As the vast number of casualties meant that large manpower reserves were necessary to prevail, the "numerical strength became critical" for victory, further reinforcing the trend toward mass mobilization. In the 260 B.C. battle against Zhao, a majority of Qin males over 15 were drafted.⁶³³ Given the life expectancy of ancient China—roughly 20-30 years—and an overall population estimate for the Chinese Warring States of 20 to 30 million, we can guess that about one third of the Qin population was males of fighting age. "In a major war ... a Qin army of 500,000 to 1,000,000 is highly possible," they conclude. Their "most conservative" estimate of the Qin militarization ratio or percentage of men under arms is more than 8% and the high estimate is 20%.⁶³⁴

In line with the legalist philosophy of mass mobilization, the Qin leaders sought to draft anyone they could into the military effort, even individuals that the traditional Chinese culture of

⁶³² Quote is from Sawyer 10. See also Lewis 625-8.

⁶³³ Sawyer 11.

⁶³⁴ Kiser and Cai 520.

the other Warring States would frown upon. For example, Shang Yang believed that although pitched battles and offensive actions should be left to men, women should be enlisted to participate to a war's defensive strategies, like holding cities, for example. *The Book of Lord Shang*, a work generally attributed to Shang Yang recounting his philosophy and political accomplishments, instructs military leaders that in addition to men, they should use "all able-bodied women" to "make, at the approach of the invaders, earth works as obstruction, and traps, chevaux-de-frise and pitfalls, to pull down the supporting beams and to tear down the houses, to transport what is transportable, and to burn what is untransportable" so that the enemy cannot use those installations to their advantage. Commanders should even use the "old and feeble" for simple supporting tasks as guarding cattle, forage for food and gather water for the army. Shang Yang nevertheless warns against mingling men, women, and the feeble to avoid intrigues.⁶³⁵

Another unorthodox Qin way to maximize military mobilization was to force convicts to perform their penalties in government artisanal workshops, particularly in "government bronze and iron foundries where coins, agricultural implements and weapons were cast." Convicts thus played a key role in arming the soldiers and keeping the Qin arsenal stocked. Although it is not clear if other states adopted this Qin policy, it is highly unlikely because Qin archeological findings remain the only ones to bear inscriptions of convict statuses. As Yates concludes, the use of convicts in supplying the army "may have been one of the contributing factors to Qin's success in unifying China":

By controlling the quantity and quality of the articles in their workshops and forcing large numbers of skilled workers to provide virtually free labor, the Qin may have been able to provide their armies with better equipment and ... resupply them even after defeats and

⁶³⁵ *The Book of Lord Shang: A Classic of the Chinese School of Law*, attr. Shang Yang, ca. mid-4th cent. B.C., transl. and ed. by J.J.L. Duyvendak (Chicago: University of Chicago Press, 1963), 250-1.

loss of material and thus have been able to sustain their war efforts at a much higher level of intensity than their opponents.⁶³⁶

The mass mobilization effort required by Qin's switch to a mass infantry and by its use of the larger population in the war efforts necessitated a reorganization of the population in order to be able to count the peasants and enlist them on a rotational basis without disrupting the agricultural production, which was also vital to the country's growth strategy. To that effect Duke Xian, who ruled Qin from 385 to 362 B.C. following a period of exile in Wei, had observed Wei's system of household registration and upon returning to Qin immediately implemented a similar system on a larger, more systematic scale. Duke Xian divided the state into a number of administrative districts (*xian*) that served as the base for the levy, then arranged the Qin population within these districts in groups of five households, which served as the basic social organization, and the army in corresponding squads of five men that served as the basic battle organization. The simple linkage of social and military positions already observed in Rome and Mongolia served to simplify and systematize military mobilization. This linkage was not only implemented in the lower social strata of peasants but also involved military commanders, who often simultaneously held posts as civilian administrators, reflecting the legalists' aim to maintain control over the military by avoiding to make it a distinct profession. Managing masses of peasants like Qin's army also required much skill. For that purpose, the Qin kept a core of professional officers specially appointed to train and organize the conscripts, like military managers. A straightforward military hierarchy completed the new structure of the Qin army to enable the smooth attribution of logistical and military responsibility in Qin's massive

⁶³⁶ Yates 412.

campaigns: each five-man squad was led by a corporal; every fifty squads or hundred men were led by a centurion; and above the centurions were prefects and generals.⁶³⁷

2. Army Discipline (IV5.2)

Although the unprecedented scope of Qin's military reorganization gave the rising hegemon the means to defeat its rivals, it was the unusual disciplinary system accompanying it, inspired again by the legalist philosophy, that propelled Qin's new army far beyond that of its rivals. The legalists firmly believed that strict laws defining conduct and individual responsibilities should provide a framework to facilitate the most efficient channeling of the state's resources. In order to ensure that each individual followed its pre-defined role and performed as was expected to yield the most productive result, Shang Yang's reforms instituted a law enforcement system based on rewards and severe punishments, both in the civilian and military realm. While in the other Warring States, rewards and punishments were generally temporary and ad-hoc, Shang Yang's disciplinary system was highly institutionalized, rewarding more and punishing harsher than in any other state. Prowess on the battlefield, determined by the number of enemies slain by an individual or by one's unit, was rewarded generously with systematic promotion in rank. For those slain in battle, the reward could be inherited by a descendent. Shang Yang established a 17-rank honor system, and each rank had corresponding military and administrative roles: the four lowest ranks corresponded to soldiers and low-level bureaucrats, while ranks 5 and up corresponded to officer positions in the army and official levels in the administration. The four highest ranks were the equivalent of the former feudal lords. After reaching the rank of officer, promotions would come with land, houses, tax-exemption or slaves, and often included a high-level administrative appointment. Upon reaching

⁶³⁷ *The Book of Lord Shang* 295; Sawyer 11; Hui 65; Fairbank 9.

the 8th rank or higher, men received whole villages' tax revenues. The ranks also carried certain legal and religious privileges: they could be remitted against punishments or penalties, or used to obtain burial advantages, for example, and were thus a crucial indicator of one's place in society.⁶³⁸

Unlike the disciplinary systems of Qin's rivals, Shang Yang's system was eminently meritocratic. Ranks were awarded on the basis of performance only and were not hereditary. Even the lordship titles of the upper four ranks constituted a purely meritocratic nobility. *The Book of Lord Shang* emphasizes that "when ... rank is given according to military merit, ... victory is certain." Shang Yang believed that because rewards and punishments constituted motivators for productivity, promotion should depend on achievement rather than be automatic as in other states. Shang Yang analogized that "if a tube ... has no bottom, it can certainly not be filled; to confer office and to grant salaries without regard to merit is like having no bottom."⁶³⁹ Therefore, access to promotion was open to all in Qin, which helped attract the best candidates. In contrast, in the other Warring States, ranks were distributed not throughout the population but only among the elite. In Qin, punishments were similarly inflicted equally upon all. Even though Qin imposed the most severe punishments, including frequent use of torture and executions, high-ranking officers were not spared if they skirted their responsibilities. As Fairbank notes, "generals ... lost their heads at least as often as anyone else." Income was also set by rank level, even for officers, in an attempt to deter bribery. As a result of the quality of its leadership, Qin bred a number of excellent generals from within its ranks, including the architects of its wars of conquest, Bo Qi and Wang Jian.⁶⁴⁰

⁶³⁸ Lewis 612-3.

⁶³⁹ *The Book of Lord Shang* 205, 253, 275.

⁶⁴⁰ Fairbank 7; Hui 80-81; Lewis 632.

Another unique trait of Qin's military discipline system was the notion of collective responsibility, also an invention of the legalists that was first implemented by Shang Yang, both in the military organization and in its civilian equivalent. Each member of a squad of five (and each family in a group of five households) was held responsible for the actions of the other members. For example, if a member of a squad deserted battle, the other four members would be punished unless they caught the missing man or killed an enemy soldier to redeem themselves. In addition to easing conscription, thus, the household registration system was also introduced "for purposes of mutual responsibility and surveillance." Denouncing a fellow squad member's crime or a fellow family-member's crime was rewarded as highly as showing bravery on the battlefield, while collecting enemy heads in battle while failing to denounce a crime or concealing it was punished as harshly as deserting or surrendering in battle. Because of the parallelism between military and civilian/social organization, mutual responsibility worked both in peace and in battle. Thus, "members of units of mutual responsibility were neighbors in times of peace and fought side by side in combat."⁶⁴¹ *The Book of Lord Shang* emphasizes that as a result, each squad member bore the responsibility to encourage others to act their best, producing a more disciplined, more efficient military force.⁶⁴²

3. Military Strategy, Tactics, and Technology (IV5.3)

To complement its new disciplined and organized military force, Qin devised a military strategy of prudence and concealed its rise. In line with Qin's attempts at deceiving potential balancers as to its real power, Duke Xian (385-362 B.C.) and his successor Duke Xiao (362-338 B.C.) adopted a defensive posture during most of Qin's reform period, both to spare resources

⁶⁴¹ Lewis 612, 614-5.

⁶⁴² *The Book of Lord Shang* 295-8.

and to feign weakness while building power. Only after Duke Xiao became confident that Qin had caught up with its neighbors' capabilities and had generated a sufficient increase in its relative power did he switch to the offense. Qin's wars of conquest started in 356 B.C. The shift in strategy is clear when looking at Qin's long-term war participation. Qin commenced only 11 out of 160 major wars in ancient China between 656-357 B.C., during the Spring and Autumn Period and the early Warring States Period, or roughly 7% of the conflicts, but it initiated 52 out of 96 great power wars in the region between 356 B.C. and its achievement of hegemony in 221 B.C., or about 54% of the wars. In the latter period, Qin achieved 48 victories in the 52 offensive wars it started, a 92% success rate.⁶⁴³ Qin's impressive offensive record is most certainly due to the prudence of Duke Xiao and Shang Yang in their use of the offensive. *The Book of Lord Shang* points to the importance of avoiding reckless attack even when possessing overwhelming means and suggests the calculated use of defensive postures to better prepare for offense, particularly for a state like Qin that was landlocked with strong neighbors and for which a multi-front war would be risky.⁶⁴⁴

In fact, despite a grand-strategic switch to offensive warfare during its rise, Qin never neglected its defensive positions. It developed a system of walls, forts, and watch towers on its border with Zhao and on its outer border with its northern nomadic neighbors, who regularly raided Qin lands. The First Emperor rendered the latter famous by connecting them, forming the first version of the Great Wall, but "the interior walls were more significant in the Warring States Period" given the constant warfare between the major states. The walls often took advantage of natural boundaries and where possible were built along rivers, behind the dykes. While other Warring States also constructed protective walls between each other, Qin was once more at the

⁶⁴³ Hui 65.

⁶⁴⁴ *The Book of Lord Shang* 248-9.

forefront, with the "earliest record of such a wall [i.e., an interior wall]" dating back to 461 B.C. when Qin "walled the banks of the Yellow River." Qin extended the Yellow River wall in 417 B.C. and during the early Warring States Period, and also constructed walls of the Luo River, a tributary of the Yellow River, after it pushed its frontier eastward.⁶⁴⁵ When needs for border protection intensified during the early- to mid-Warring States Period, Qin added forts throughout the countryside and checkpoints at strategic passes and intersections. Cities also became fortified, leading to the development of new techniques and the acquisition of new weaponry such as siege engines, mobile shields and towers, battering rams, etc. to besiege them. While some classic Chinese military strategists like Sun Tzu condemned attacks on cities, Qin's legalist rulers saw no grounds to refrain from such attacks, while at the same time enhancing their own cities' defensive mechanisms.⁶⁴⁶

Just like its peaceful rise and concealed strength strategy, Qin's tactics also prominently featured deception and prudence. *The Book of Lord Shang* cites Sun Tzu's military principles, thus suggesting that Shang Yang was a disciple of Sun Tzu and adhered to his teachings. In particular, *The Book* affirms that military commanders should follow Sun Tzu's advice on prudence in battle, always measuring the strength of an enemy before combat and never engaging in battle without superior manpower and provisions. Commanders should not become overconfident in victory, and should forego pursuing enemies too far, for example, to avoid overstressing their troops. Similarly, commanders should keep quiet about a victory so as not to alert the enemy. "He who intends to direct the people ... and he who mounts a good horse cannot but be on his guard," *The Book* advises.⁶⁴⁷ At the same time, Qin leaders seldom refrained from using tricks to defeat their enemies. Deception was particularly crucial at the beginning of Qin's

⁶⁴⁵ Lewis 629-30.

⁶⁴⁶ Bodde 47; Sawyer 12.

⁶⁴⁷ *The Book of Lord Shang* 245-7.

rise, before the new mass peasant army was fully operational, because Qin was at a numerical disadvantage. Even once the draft was in place, Qin leaders still sought to use their soldiers more sparingly than their neighbors because their legalist leanings led them to maximize their economic force also to generate more growth, and thus tricking the adversary to avoid costly frontal battles continued to be a central Qin tactic. The bribing of enemy officials, already mentioned above, remained one of Qin's favorite deceptive tactic.

One of the classic examples occurred in 262 B.C. during Qin's war with Zhao, when Qin bribed a Zhao official to spread the rumor that Zhao's chief general was "avoiding combat" out of fear of being defeated. In reality the general was carefully planning his attack. Zhao's king bought into the trick, however, and dismissed the general. The officer that replaced him was ardent to prove himself and changed plans entirely. He rushed into battle at Changping (260 B.C.) and went on a reckless offensive against Qin's far more experienced general Bo Qi, who simply waited for Zhao's army to advance before encircling it, cutting off its supplies, and taking its rear guards. Zhao suffered a severe defeat; when it surrendered, Qin slaughtered all of its soldiers, and as a result Zhao, like Wei and Qi before it, lost its great power status in the region.⁶⁴⁸ Despite the success of its strategy and tactics, Qin was not really superior in weaponry. Nor did it show particularly innovative skills when it came to military technology. Instead, it is Qin's "superiority in statecraft" that enabled it to extract the most resources, both domestically and by taking advantage of its opponents' innovations. Qin was "a borrower rather than an innovator," and copied the best and latest techniques from its neighbors and capitalized on them.⁶⁴⁹ The crossbow, for example, one of the new weapons introduced in the middle of the Warring States Period, was invented either by Chu engineers or by non-Chinese natives south of

⁶⁴⁸ Lewis 639-41; Hui 73, 89.

⁶⁴⁹ Hui 95-7.

Chu. Cavalry techniques, such as riding astride, and accessories, like riding pants, were likely copied from barbarian tribes north of Qin.⁶⁵⁰

There is a vigorous debate among historians and anthropologists, however, about Qin's possible innovation with iron technology. Iron-based weapons were first developed in China in the mid- to late 4th century B.C., just as Qin began its rise, and Qin benefited from a clear geographical advantage in their development because it had taken control of the core of China's iron-producing basin around Shu and Ba.⁶⁵¹ Some experts "attribute Ch'in's [i.e., Qin's] military success to an advanced iron technology which ... enabled Ch'in [i.e., Qin] to equip its soldiers with wrought-iron swords superior to the bronze weapons generally used by its opponents."⁶⁵² Others contend that excavations reveal few iron swords but many bronze swords in ancient China in general and "no clustering of iron swords in the areas under Ch'in [i.e., Qin] domination," so that it cannot be concluded that Qin prevailed because of a metallurgic innovation in weaponry.⁶⁵³ The absence of archeological evidence does not necessarily disprove the theory of Qin's metallurgic superiority, however. As anthropologist William Trousdale points out, "iron survives more poorly than bronze" and they are more likely to have been reused, repossessed by the state, and recast for other purposes. The "true ratio of iron to bronze swords was, then, probably ... more heavily weighted toward iron than the physical evidence admits, or likely will ever allow."⁶⁵⁴ The metallurgic innovation hypothesis can thus not be discarded.

In the end, Qin's military strength came primarily from its ability to pioneer a groundbreaking, highly centralized and efficient military organization, reinforced by a strict disciplinary

⁶⁵⁰ Yates 410.

⁶⁵¹ Fairbank 5, Yates 414.

⁶⁵² Bodde 46-7.

⁶⁵³ David N. Keightley, "Where Have All the Swords Gone? Reflections on the Unification of China," *Early China* 2 (1976), 31-33.

⁶⁵⁴ William Trousdale, "Where All the Swords Have Gone," *Early China* 3 (1977), 65-66.

system reflecting the political philosophy of legalism and its focus on tight control and the maximization of resources. Qin's combination of prudent and deceptive strategies and tactics also stemmed from their leader's adherence to legalism, as did Qin's emulation of the best weaponry from its neighbors to compensate for its possible lack of technological innovation. Qin's military reforms propelled it from weak to quasi-hegemonic in just about a century. Aware of the growing advantage their reforms gave them, Qin leaders went to great lengths to prevent others from copying them, as their competitors also attempted internal reforms of their own. By the turn of the 3rd century, "all other Warring States had pursued some variants of self-strengthening reforms." As a result, Qin "engaged in massive brutality and massive bribery" to deter emulation of its own reforms. Defeated armies like the Zhao army in 260 B.C. were slaughtered en masse, for example.⁶⁵⁵ But Qin's revolutionary legalist reforms did not solely deal with the military. Instead, they were all-encompassing reforms that launched Qin into an unprecedented state-building effort that aimed at strengthening every aspect of the state, and played a crucial role in enabling its accession to hegemony.

b. Non-Military Skills and Innovation (IV6)

As Yates points out, "none of these changes in military techniques and [organization] can be divorced from the social, political, and economic changes that were also taking place." Capitalizing on Qin's advantageous geopolitical setting (IV6.1), the legalist reforms swept over Qin's economic (IV6.2), political (IV6.3), and social (IV6.4) order. It transformed the Qin state into a very powerful instrument of control. Qin's self-strengthening reforms were also directed toward boosting the state's non-military power—its economic, political, and administrative

⁶⁵⁵ Hui 84-7.

capabilities—by rationalizing its state structure and making resources easier to extract and mobilize. Qin then used these assets to reinforce its military capabilities and accelerate its rise. While other ancient Chinese states also implemented domestic reforms geared toward economic, social, and political strength, none reached the extent of Qin's legalist reforms, which were owed primarily to Shang Yang in the mid-4th century B.C. Just like his military reforms, Shang Yang's economic, administrative, and social reforms were mostly based on learning and improving already existing policies rather than conceiving them from scratch. Shang Yang "applied old models to changing circumstances" and implemented them "more systematically" than Qin's rival states.⁶⁵⁶

Unlike the Romans and the Mongols, however, the Qin leaders' empire-building strategy focused primarily on maximizing resources and control in order to generate military victory, and failed to pay sufficient attention to winning the support of the population. In his introduction to *The Book of Lord Shang*, Duyvendak explains that "the methods which served Ch'in [i.e., Qin] to reach its aims were not altered when the whole empire had been unified under its sway." After achieving hegemony, Qin continued to exercise strict, tyrannical control and "thereby estranged all sincere people." As a result, the Qin Dynasty was unable to retain power under the pressure of popular revolt. The succeeding dynasty, Han, took over Qin's military, economic, and administrative achievements while avoiding its excesses and was therefore able, unlike Qin, to sustain its rule.⁶⁵⁷

⁶⁵⁶ Yates 410-11; Hui 79-80.

⁶⁵⁷ *The Book of Lord Shang* 2.

1. Geography and Geopolitical Assets (IV6.1)

The Qin state's geographic layout had a lot to do with its initial burst of power in the mid-3rd century B.C, allowing Qin a safe early rise, isolated from the bustle of central China, yet providing it with the best raw material for internal growth. Though it began its rise small and weak, Qin was located in the best possible environment to grow, and its leaders exploited that situational advantage. In the words of Han historian Sima Qian, Qin's original territory "was situated far out on the border [of China] in the old province of Yong," around the Wei River valley in modern Gansu Province, the capital area of the former Western Zhou Dynasty.⁶⁵⁸ Its core was a fertile, irrigated plain surrounded by high mountains, producing "a splendid geographic situation that combined productivity and security." Qin was the furthest west of all the Warring States and approaches from the central China were blocked by the mountain chains, pierced only by a few strategic passes like the forbidding Hangu Pass in the east and the Wu Pass in the southeast. In order to increase its safety, the Qin army permanently occupied both passes, as well as other, smaller key points of access from the east. Its only threat was therefore from the north, but there was no great power there, only Rong tribes which occasionally raided Qin land but "posed no sustained menace" and were conquered by 314 B.C., ending all threats from the north and west. South of Qin lay the small powers of Ba and Shu, in the current Sichuan Province, which Qin conquered between 441 and 316 B.C. and which "provided Qin with a new, highly productive 'grain basket' that supplied its wars of conquest in the 3rd century B.C.." By the early 4th century B.C., in other words, "Qin had made itself virtually impervious to attack."⁶⁵⁹

⁶⁵⁸ Sima Qian 22.

⁶⁵⁹ Lewis 596.

"Behind these barriers Ch'in [i.e., Qin] could build up its strength before launching attacks on the other states."⁶⁶⁰ This proved particularly easy since Qin controlled one of the richest areas in China, and its very first conquests added other particularly rich areas to its territory: Ba and Shu in 316 B.C., followed by the Hanzhong region in Western Chu in 312 B.C. that linked Ba and Shu to Qin's capital region around Xianyang. In 310 B.C., Qin built a new regional capital in Shu at Chengdu, modeled after Xianyang, which served as a base to develop the Sichuan basin around Shu, Ba, and Hanzhong as a key agricultural center. But Qin land was not only fertile and very productive agriculturally, it was also particularly rich in minerals and precious metals.⁶⁶¹ Sinologist Donald B. Wagner found mentions of Qin's riches in a Han document dating from around 117 B.C., the *Shanhaijing*, which Wagner believes is based on an original but lost document from the 4th century B.C. listing gold, iron, salt, copper, tin, and silver deposits in a number of areas. As Wagner discovered, "these areas correspond roughly to the territory controlled by the state of Qin in the mid-3rd century B.C., before its final conquest of the east and northeast." The 4th century list was likely compiled by Qin for administrative purposes. The "centre" areas on the list seem to correspond to Ba, Shu, and Hanzhong, and contained a particularly high concentration of iron and other metals, probably a reason Qin conquered these areas early. Incidentally, Wagner's finding seems to corroborate the theory that Qin's military superiority may have been due to its technological advances in the production of iron weaponry.⁶⁶²

Qin strategically embraced its geographical isolation and outsider position during its early rise at the beginning of the Warring States Period. The state "did not participate in the

⁶⁶⁰ Bodde 46.

⁶⁶¹ Lewis 635.

⁶⁶² Donald B. Wagner, *Iron and Steel in Ancient China* (Leiden: E.J. Brill, 1996), 247-50.

alliances of the feudal lords of the central states, being treated instead like a barbarian."⁶⁶³ Qin leaders were well aware of the state's natural strengths, as a discussion between SuCh'in, the Qin military strategist, and King Hui (r. 337 to 311 B.C.) related in the *Zhan Guo Ce*, demonstrates. SuCh'in reportedly told the king:

Your majesty's state has the wealth of Pa [i.e., Ba], Shu, and Han-chung [i.e., Hanzhong] on its west and the steeds of Tai and the furs of Hu in the north. To the south it is bounded by Mount Wu and the lands of Ch'ien-chung and to the east it is sealed by the peaks of Yao and the Canyon of Han-ku [i.e., the Hanku Pass]. Fat fields, flourishing people, ten thousand chariots, and a million mettlesome troops; a thousand miles of rich fallow-land and an abundance laid up with defensible borders—truly an arsenal of nature, the most awesome state in the world!⁶⁶⁴

Thus, of all the Warring States, Qin was the best endowed geographically, granting it unique security and economic growth prospects. This natural predisposition contributed to the success of the pioneering economic reforms introduced by Shang Yang.

2. Economic Advantage (IV6.2)

Not all of Shang Yang's 356-350 B.C. economic reforms, like his military reforms, were unique to Qin, and some were implemented, at least partially, by other ancient Chinese states. However, "he [i.e., Shang Yang] put them into practice more systematically than had any of his precursors." The sweeping scope of his reforms is clear when reading later dynasties' annals and histories, which all portray Shang Yang as singlehandedly responsible for ending the feudal system and introducing private landownership, for example, as if "the entire economic history of five centuries was the work of a single evil genius." Though exaggerated, those later accounts

⁶⁶³ Sima Qian xi, 23.

⁶⁶⁴ *Chan-kuo Ts'e*, Ch'in 47 (82).

show how revolutionary Shang Yang's reforms were in that they gave Qin the tools to generate powerful economic growth to sustain its military efforts toward hegemony.⁶⁶⁵

Qin's economic reforms, like its military reforms, were inspired by the legalist philosophy embraced by Shang Yang and the Qin kings. In the economic realm, legalism dictated that all efforts be put towards intensive agricultural in industrial production so as to maximize economic growth and generate the surpluses necessary to support a powerful military. Shang Yang thus made individual households the basic units of both production and property or land ownership, a stark break from the other states' exclusive aristocratic ownership that helped propel ancient Qin, and later China as a whole, into post-feudalism. Just like with the military, law was used to regulate and encourage intensive production, along with a system of rewards and severe punishments to trigger obedience. *The Book of Lord Shang* underscores the key role of extensive farming in generating power, claiming that "the means, whereby a country is made prosperous, are agriculture and war." *The Book* later argues that "where a hundred men farm and one is idle, the state will attain supremacy; where ten men farm and one is idle, the state will be strong; where half farm and half is idle, the state will be in peril. That is why those, who govern the country well, wish the people to take to agriculture." Idleness promotes "quarrels over authority," which stirs trouble among leaders and can cause the country to become "dismembered." Accordingly, Shang Yang introduced a system of intensive agriculture designed to extract more war resources, based on the new small-scale land ownership. In exchange for granting land ownership to individual families, helping farmers with subsidies to purchase farm tools, and educating farmers on land techniques to improve their productivity, the new Qin system required farmers to reach certain farming quotas. Exceeding the quotas was rewarded by

⁶⁶⁵ Lewis 611.

tax exemptions, while failing to reach them was severely punished, sometimes even by slavery.⁶⁶⁶

In addition, because Qin was a frontier state with a relatively small population and a lot of undeveloped land, Shang Yang found innovative ways to encourage people to migrate to the state's less populated areas in order to maximize land exploitation. In 348 B.C., for example, he issued a pro-agriculture measure meant to develop the frontier and wasteland areas of the state. Qin constructed a "grid of pathways across the fields," dividing land throughout the empire into squares of a pre-set size corresponding to the cultivation capacity of one male farmer's household. Families with several adult males were required to divide into several households, with each male taking up his own square, if they wanted to avoid heavier taxation. As a result, the measure forced families to spread out and move to new patches of farmland further along the outskirts of Qin territory.⁶⁶⁷ Another unusual measure to exploit more land was to enlist criminals to colonize remote or newly acquired territory. According to Sima Qian, for instance, King Zhao (rule 307-251 B.C.) repeatedly pardoned criminals and in exchange sent them to populate distant areas and begin the agricultural exploitation of those lands. In 280 B.C., the king "pardoned criminals and moved them to the region of Rang." The following year he did the same, and "moved them to the region of Nanyang." In 278 B.C., Qin took over Yan and Deng from Chu, and "criminals were pardoned and transported to those areas."⁶⁶⁸

To further offset Qin's lack of manpower, Shang Yang allowed and supported labor-based immigration from the neighboring states of Han, Wei, and Zhao. "The territory of Chin [i.e., Qin] ... [is] incompletely utilized," *The Book of Lord Shang* explains, as only parts of it are exploited for agriculture and "the population is insufficient to fill the territory." To draw

⁶⁶⁶ *The Book of Lord Shang* 185, 191-2; Hui 81-2.

⁶⁶⁷ Lewis 613.

⁶⁶⁸ Sima Qian 30.

immigrants to Qin, Shang Yang offered them generous terms: free houses and free land, as well as a three-generation tax exemption and a ten year military service exemption. *The Book* expected everyone to desert the neighboring states to join Qin because of the enticing offer: "if your majesty follows this policy, then within ten years the various feudal lords will have no people," and everyone will have emigrated to Qin. The policy proved highly beneficial to Qin. Immigrants engaged in farm work, provided food and kept the economy growing while Qin farmers went to war, enabling Qin to pursue a heavy draft without loss of economic production. In addition, the policy hired potential soldiers away from its rivals Han, Wei, and Zhao. Qin rulers were well aware of this additional benefit, as *The Book* shows. "By this plan, two birds will be hit by one stone" as the enemy will be defeated from within, it argues.⁶⁶⁹ Qin did not hesitate to resort to immigration to attract foreign workers, even beyond unskilled laborers, and was willing to "recruit talent wherever it could be found." And although some other states also offered advantageous immigration policies, Qin did this "to an extent unequaled by any other state."⁶⁷⁰ Shang Yang, Lü Buwei, Li Si, and numerous other high-ranking Qin officials were of foreign origin. Qin just excluded its most vital security posts from foreign access. Thus, Qin's top officers and generals, like Po Ch'i, Wang Chien, and Meng Tian, were all of Qin origin.⁶⁷¹

In addition to agricultural production, Qin also sought to pioneer industrial development. Qin's innovative experiments with iron, for example, led the state to improvements not just limited to weaponry. Qin's developed the "technique of casting iron into agricultural implements, ... an innovation which profoundly increased agricultural production."⁶⁷² In fact, Qin under Shang Yang quickly became the leading industrial power in China thanks to the creation of a

⁶⁶⁹ *The Book of Lord Shang* 266-72.

⁶⁷⁰ Bodde 48.

⁶⁷¹ Hui 82.

⁶⁷² Yates 409.

new class of capitalists. In an effort to diminish the political power and potentially disruptive influence of Qin's remaining powerful and wealthy aristocratic families, Shang Yang deported those families from their original areas to new regions not yet fully exploited, generating power vacuums in the regions from which they were removed that the state quickly filled with loyal Qin officials. This predominantly political move triggered a positive externality for the Qin economy, however, since "in the places where they were resettled [like Shu, Ba, and Hanzhong], their wealth and their technical and organizational skills contributed to economic development." Those families took control of the production of iron and other raw materials, minerals, and precious metals, and engaged in trade with neighboring barbarians, boosting the productivity of Qin's border regions and newly conquered areas. "Qin's [legalist] ideology and ... policies seem to have favored industrial development," with a weakening of the traditional landed aristocracy and the growth of those "proto-capitalists, ... wealthy men who invested in industry rather than land and therefore helped to build up the political and military strength of the state." Wagner notes a parallel here between the rise of Qin and the rise of Germany in the 19th century stimulated by Bismarck's use of financiers and industrialists. Qin's "industrial development policy," therefore, with its focus on the natural resources of newly conquered regions and their development by resettling wealthy industrialists was very modern in essence. Production was controlled not by the state but by "private persons" focused on their own profits but working in a way that benefited the state. There was still some direct involvement of the Qin state to guarantee the convergence of goals: the state controlled the retail of iron and salt (perhaps through price controls) and imposed severe penalties if producers failed to reach mandated production quotas.⁶⁷³

⁶⁷³ Wagner 253-6, 259-61.

While Qin thus promoted agricultural and industrial production, since such activities directly contributed to increasing Qin's power, Shang Yang and fellow legalists remained very suspicious of other professions that failed to yield the same immediate quantitative results and from which it was harder to extract taxes. Merchants, artisans, and scholars in particular were considered idle. *The Book of Lord Shang* claims that "studious people hate law, merchants are clever in bartering, and artisans are useless," and if they are numerous "then the state will be poor and in a dangerous condition." The Qin reforms therefore found innovative ways to discourage their activities, for example by officially classifying them as secondary and non-important since they did not contribute to feeding the population and army and increasing the material wealth of Qin. As Han Feizi points out:

An enlightened ruler will administer his state in such a way as to decrease the number of merchants, artisans, and other men who make their living by wandering from place to place, and will see to it that such men are looked down upon. In this was he lessens the number of people who abandon primary pursuits [i.e., farming or industry] to take up secondary occupation.⁶⁷⁴

Artists and artisans were only tolerated if they produced things of practical use. "Music and fine clothing" distract farmers, make them "licentious" and turn their attention away from farming, *The Book of Lord Shang* says. The same was true with servants. Having servants encourages laziness and it is therefore better not to have them. "If lazy people cannot be inactive, and hirelings do not find a livelihood, there will certainly be agriculture." Hotels and good dining were considered distracting too because they encouraged leisure and travels and were favored by artisans and merchants. "Criminals, agitators, conspirators, and those who unsettle the minds of the farmers [should] not travel." Therefore, *The Book* recommends abolishing hotels and their

⁶⁷⁴ Han Fei Tzu 116; *The Book of Lord Shang* 191.

keepers, in the hopes that having no clients, they would turn to farming. It also suggests tax increases on wine and meat so people would not get inebriated or overweight and would "not neglect agriculture." Scholars other than legalists were also despised for their uselessness and ability to inciting dissent by debating ideas and policies—an idiosyncrasy since the Spring and Autumn Period and the early Warring States Period are known as the time of the "Hundred Schools of Thought," where philosophers and schools proliferated throughout China. The legalists' suspicion of intellectuals was so great that after Qin united the Warring States, the First Emperor and his chancellor, Li Si, engaged in massive book burnings and the live burial of hundreds of alchemists between 213 and 206 B.C., leading to monumental losses in the cultural and intellectual history of China.⁶⁷⁵

Merchants were frowned upon because of the belief that they took advantage of prices and made profit by "waiting for the best time to sell" rather than by adding value to a product, which amounted to "exploiting the farmers" by inflating food prices, and thus made the "the vermin of the state," according to Han Feizi.⁶⁷⁶ Shang Yang's policy goal was therefore to turn every non-farmer into a farmer. He imposed heavy tolls and taxes on market transactions and the movement of goods, as well as burdensome regulations on merchants, in the belief that if merchants could not make profit over the price of products, they would be "fearful, and being fearful, they turn to farmers" and increase the agricultural production.⁶⁷⁷ Merchants had to perform corvée duties, for example, and were subject to drastic registration procedures. They were required to register as "inferior people," and as a result were prohibited from wearing silk or riding horses and were sent on long garrison duties in isolated frontier lands where no trade

⁶⁷⁵ *The Book of Lord Shang* 177-9.

⁶⁷⁶ Han Fei Tzu 117.

⁶⁷⁷ *The Book of Lord Shang* 181.

was possible.⁶⁷⁸ Since their clients were also forced to register under the same status, as were servants, their business was severely curtailed. The only acceptable occupation for merchants was to become army suppliers, with strict monitoring so they would not have the "opportunity to hatch up corrupt schemes," *The Book* explains. But even army suppliers faced many restrictions, such as the compulsory registration of carts, oxen, carriages, and baggage wagons. Public contracts were also exclusive, and those merchants engaged in public transport were prevented from simultaneously carrying "private cargo."⁶⁷⁹

Besides aggressively favoring farmers and industrialists over other professions, Shang Yang also pursued a number of structural reforms indirectly aimed at increasing the state's agricultural and industrial productivity. Qin started a vast program of public works, which increased in scope as the state increased in size, as "it became possible to organize labor services on a vastly larger scale than ever before" and more and more workers were available. Just like men were required to serve in the military, they were also recruited to serve roughly one month per year of labor service to contribute to public construction projects. Qin developed a large network of roads throughout the state and then the empire, converging at Xianyang, the capital, and connecting all parts of Qin's territory. By the time of the empire, Qin highways covered roughly 4,250 miles. The Han Dynasty later expanded them to 22,000 miles. Though Qin's road system paled in comparison to the 48,500 miles of Roman roads, it was for the most part because the Qin empire covered much less distance than Rome's two million square miles.⁶⁸⁰

In addition to the roads, Qin also worked hard to enhance its main natural assets, strategic isolation and fertile soil, by building uniquely extensive irrigation systems, though these projects occurred later in Qin's rise. In 256 B.C., Qin built the intricate Dujianyan irrigation system in the

⁶⁷⁸ Lewis 613.

⁶⁷⁹ *The Book of Lord Shang* 183-4.

⁶⁸⁰ Kiser and Cai 529.

southwestern plain near the regional Shu capital of Chengdu, which is still used today and provides water to an area over 2,000 sq. miles. The system diverts the yearly excess floods of the nearby Min River in a channel through a mountain toward the dry plain. Cutting through rigid mountain rocks was a major feat prior to the invention of explosives. "Its economic importance must have been great," since "down to the present day it has continued to supply a never-failing flow of water to some five million people." A decade later, Qin completed the Zhengguo canal that irrigated an area of over 10,000 sq. miles in the Guangzhong plain north of Xianyang by connecting the Jing and Luo Rivers, both tributaries of the Wei River. The canal also played a key economic role, as Sima Qian points out, since it brought water to 465,000 acres of previously infertile land. The First Emperor also constructed the Lingqu Canal in 214 B.C., connecting the Xiang River to the Lijiang River and creating a direct waterway between the Yangtze in Central China, into which the Xiang flows, and the South China Sea since the Lijiang flows into two other rivers ending up by Macao. The 22.6 mile long canal, equipped with 36 canal locks, is the world's oldest contour canal, meandering canals espousing the natural environment in order to avoid costly engineering project. Because of these three projects date relatively late in Qin's rise, however, Bodde concludes that although they certainly reinforced Qin's economic power at a crucial point and perhaps accelerated its hegemonic accession, they "did not determine the course of Ch'in [i.e., Qin] history."⁶⁸¹

Finally, complementing its infrastructural reforms, Qin also implemented a number of standardization reforms in order to facilitate economic growth. Qin legal documents discovered in 1975 in a former Qin official's tomb near Yün-meng show that Qin imposed fixed, standard units for the measurement of weights and volumes, and that failing to use those official units resulted in heavy fines. This focus on standardization demonstrates that Qin was concerned with

⁶⁸¹ Bodde 46.

accounting, record-keeping and statistical assessment, which were crucial to its administrative control and military and economic performance. Indeed, Qin officials needed records and statistics of state contracts, population registration, tax collection, etc., to evaluate how its different geographic areas performed and to determine how to improve such performance. Paradoxically, despite its attempts at impeding commercial activity, Qin's focus on boosting its economic productivity while promoting individual land, property ownership and responsibility spurred the emergence of a "new economic system" that Yates describes as closely resembling a market economy. This was partly rendered possible through the introduction of a universal metal coinage system that gradually replaced the gift or barter exchange system of the Spring and Autumn Period and that Qin not only standardized but spread to all of China as it expanded, reinforcing its own economic appeal.⁶⁸²

3. Political Organization (IV6.3)

Although Shang Yang's economic reforms triggered considerable growth that contributed to the rising hegemon's victorious military endeavors, Qin's self-strengthening reforms would not have succeeded without an innovative overhaul of its administration that provided the ultimate central control that allowed the Qin state to complete and implement the array of legalist reforms. The administrative reforms were initiated by Duke Xian in 384 B.C. and continued by Shang Yang in the 350s B.C.

Qin's administrative reforms were both ground-breaking and far-reaching. Qin developed a modern, "partially bureaucratized administrative system ... about two millennia before ... European states," and it went further than other contemporary administrative rationalization

⁶⁸² *Yün-meng Manuscripts*, in Anthony François Paulus Hulsewé, ed., *Remnants of Ch'in Law: An Annotated Translation of the Ch'in Legal and Administrative Rules of the 3rd Century B.C., Discovered in Yün-meng Prefecture, Hupei Province, in 1975* (Leiden: E.J. Brill, 1985); Lewis 609-610; Yates 409.

attempts, like Rome's tax farming and local authority, that were progressive for their time yet primitive compared to the Qin state. Such reforms contributed to the rising hegemon's ability to develop a strong, centralized empire, but also transcended the eventual Qin Dynasty itself, as "many of its administrative innovations lasted more than two millennia" into contemporary China. Qin led the initial bureaucratization of China, once more with a legalist purpose in mind—maximizing control and revenue and thus power. Qin's administrative reforms were therefore closely intertwined with the state's other, military and economic reforms. The development of mass peasant armies, for instance, required increased central oversight of the military. Such "partial bureaucratization of the military" facilitated the development of a parallel civil bureaucracy as military officers were rewarded with civilian positions and became state bureaucrats. Those officers had already honed their bureaucratic skills in the military thanks to the complex logistical requirements of the mass infantry armies, which taught them how to handle a "large-scale hierarchical organization." The wars and the promotion of intensive agriculture and industry simultaneously led to large-scale infrastructural developments necessary for troop movement, such as roads and bridges, that also became functional for the emerging civil administration, creating self-reinforcing Weberian bureaucratic structures. Indeed, the growing bureaucracy kept the already dwindling aristocracy weak since it put strong central control in the hands of a new, professional class of administrators who depended on the state and thus had a vested interest in preserving and expanding the bureaucratic system. Qin rulers further embraced and encouraged bureaucratic development because the new structure helped them keep the aristocracy in check.⁶⁸³

Besides the standardizations of measures and currency, Qin's reforms also introduced a number of set procedures to administer the state and evaluate government actions: seals and

⁶⁸³ Kiser and Cai 511-2, 516-7, 526, 530.

tallies to bestow authorities, census and accounting procedures to count resources and optimize tax extraction, strict appointment rules for both military and administrative personnel, evaluation mechanisms for officials to rank performances and assess rewards or punishments, and requirements for all officials to submit annual reports including forecasts to appraise productivity are just a few examples. The household registration system used as the basis for military conscription and farming duties was standardized and systematized to serve also as the basis for taxation and corvée and soon became the state's main database. Shang Yang required all population members to register with the state once they reached the height of 4 ft. 11 in. and to share the tax burden from their registration until age 60. The 348 B.C. pro-agriculture measure dividing land into pre-set squares and dividing families with several adult males was part of the same reform. Not only did the measure dramatically increase the state's tax revenue, it also led to the breakup of traditional clan kinship structures and their transformation into nuclear families that facilitated centralized control of the population. The same year Shang Yang completed the reform by introducing a poll or capitation tax based on data from the household registration system, along with a tax administration designed to collect it: all registered males were required to pay a fixed per-head tax, vastly increased if several males lived in the same household. Naturally, all household splits or moves had to be reported to authorities. When land ownership data became available a few years later, Qin also added a land tax.⁶⁸⁴

Shang Yang completed the centralization the Qin state by implementing a four-tier administrative system of commanderies or prefectures (*jin*), counties (*xian*), townships (*xiang*), and villages (*li*), enabling direct hierarchical rule from Xianyang " to the village level."⁶⁸⁵ The state was split into a number of commanderies, each divided into a number of counties, etc.

⁶⁸⁴ Lewis 615; Hui 83-4.

⁶⁸⁵ Hui 83.

Originally Qin started with 30 counties in 350 B.C. around its central Wei River basin. By 328 B.C. it added a new 15-county commandery in former Wei territory, and it also spread the system to Ba and Shu, used as a laboratory to test the county organization. By 246 B.C., Qin had a total of 12 commanderies, and under the empire in 221 B.C. the number rose to 36. Each commandery had a civil governor and a military commander, as well as a number of magistrates in each county and lower-level centrally appointed officials in each town and village responsible for daily administrative tasks. Once Qin started expanding, it simply extended this hierarchy to the newly acquired territory, generating a "second stage of significant bureaucratization" that "produced a centralized control of the state that may well have been more systematic and effective than any that had previously existed elsewhere."⁶⁸⁶ Since the new system put local power into the hands of centrally appointed professionals, it virtually eliminated the role of the old feudal, landed aristocracy, already weakened by the privatization of land ownership. The new bureaucratic apparatus abolished independent fiefs altogether, placing all Qin territory under direct central administrative control. To ensure the permanence of the new hierarchy, Shang Yang and his successors displaced large numbers of aristocratic families from their local homes to the capital where they could be monitored, while simultaneously seizing or destroying much of the armaments they possessed from the time of the noble armies.⁶⁸⁷

The reformed Qin local organization was as centralized as possible because the legalists considered the unity of political control crucial to maximize the state's resources. As Han Feizi explains, "authority should never reside in two places; the power of decree should never be open to joint use."⁶⁸⁸ Thus, the local administration must be uniform and dependent upon the central source of authority. *The Book of Lord Shang* advises that "if the administration of all the districts

⁶⁸⁶ Kiser and Cai 528-9.

⁶⁸⁷ Lewis 613-4.

⁶⁸⁸ Han Fei Tzu 27-8.

is of one pattern, then [people] will be obedient." Unlike other rising hegemony, Qin therefore did not tolerate any form of local autonomy or local rules, instead keeping a firm grip on all local administration. "In administering a country, the trouble is when the people are scattered and when it is impossible to consolidate them. That is why a sage tried to bring about uniformity and consolidation."⁶⁸⁹

Shang Yang distinguished himself not only by subjecting all of Qin's existing territory to the new division by districts, generating uniform administration on a scale never before achieved in China, but also by insisting that newly acquired territory be immediately incorporated and subject to the same central control. In contrast, the traditional Chinese practice was to break up conquered regions as fiefs and give them to various leaders or royal family members as rewards, turning the new regions into uncontrollable "patchwork[s] of tiny political entities," Burton Watson explains. Shang Yang grasped the shortcomings of the traditional practice and decided instead to "exercise direct jurisdictional control" over newly conquered areas.⁶⁹⁰ This was already done with Shu and Ba, which were among Qin's first new possessions. Although a few local tribal chiefs were kept in place in Shu, this was a purely cosmetic move since the actual administration was immediately taken over by Qin—Qin laws, land division and property rules, as well as military service and taxation at once applied to Shu and Ba. After the 221 B.C. unification of China, the First Emperor similarly refused to put his sons in charge of even the most distant new territories, as some advisors had suggested, instead agreeing with chancellor Li Si that the old practice had been disastrous in the past. He expanded the commanderies throughout the empire and thus completed the expansion of centralized control begun by Shang Yang. To ensure the homogeneity of new, distant districts, the First Emperor did not hesitate to

⁶⁸⁹ *The Book of Lord Shang* 182, 193.

⁶⁹⁰

forcefully relocate numbers of Qin citizens there to spearhead assimilation. Lewis explains that the emperor also moved local nobles from the conquered states to the imperial capital region (as many as 120,000 families according to Chia Yi) in order to "reenact the centripetal flow of talent and tribute" already successfully achieved before with the Qin nobles. In addition, the emperor had all weapons collected throughout the empire, melting them into bells and statues to decorate imperial palaces, had city walls and other fortifications within the empire destroyed for easy access to the provinces, and built regional capitals and palaces in conquered lands modeled after his own capital to show his control over local officials. In the end, Qin's centralized system of commanderies provided the basis for China's modern provincial division.⁶⁹¹

The replacement of status by merit in the new Qin administration was the last blow striking down all aristocratic power and replacing it by central bureaucratic control, firmly anchoring Qin out of feudalism. All local officials were appointed and promoted primarily on the basis of merit and competence, though Qin was not yet using the famous Chinese examination system, which was pioneered by its Han successors. All aristocratic privileges and hereditary control over state positions were first restricted, then eliminated altogether. Merit became the rationale for hiring not only army officers, but also state officials, with the creation of a "hierarchy of social/honorary ranks based on meritorious service ... to replace the Zhou system of aristocratic ranks," parallel to the new military ranks. In addition, all local officials and magistrates were paid a set salary to curb corruption.⁶⁹² During Qin's rise "bureaucratic office became ... open to men of talent rather than just to those allied by blood to the chief families and lineages of the state," improving performance. In contrast, in all the other Warring States, local positions or fiefs were given to royal family members and therefore the ruling family held large

⁶⁹¹ Sima Qian xi-xii, 24; Lewis 635, 650; Bodde 54, Chia Yi 47.

⁶⁹² Hui 102; Kiser and Cai 527.

fiefs through which the power and control of the royal line was sustained.⁶⁹³ As a result of Shang Yang's reforms, which required "members of the royal lineage had to earn their position and status through service to the state," Qin stood out as an exception to status-based hereditary appointments.⁶⁹⁴ The barring of aristocratic government positions was firmly rooted in legalist thinking and its focus on performance. Han Feizi writes, for example, that "if appointments to office are controlled by cliques, then men will work only to establish profitable connections and ... official posts will never be filled by able men and the state will fall into disorder." Rulers must be careful to appoint those with abilities rather than those with enough money or influence to buy their post because the latter will not do their job right and may turn against the ruler, and the practice will discourage the able from applying for the position.⁶⁹⁵

Besides merit appointments, another novel means employed by Qin to ensure efficient administrative control and avoid corrupt practices was to introduce strict monitoring systems. All local officials had to submit yearly reports to the central government —about everything from crops and land use to weather-related problems, insects, grain stored in public granaries to feed officials, soldiers, and prisoners—and were subject to automatic sanctions for failing to do so. Local officials were also subject to strict codes of conduct with similar penalties. Once more the Qin reforms were based on Han Feizi's legalist theories: "the enlightened ruler controls his ministers by means of two handles alone: ... punishment and favor... those who act as ministers fear the penalties and hope to profit by the rewards." Han Feizi also advised rulers to limit officials to one mandate at a time, keep rewards low, diversify responsibilities among several officials, never let local officials become too powerful, and to that effect "prune their trees" every

⁶⁹³ Yates 409.

⁶⁹⁴ Lewis 608.

⁶⁹⁵ Han Fei Tzu 22-4, 48.

once in a while.⁶⁹⁶ The influence of Han Feizi on Qin's administrative practices is evident. For example, the Yün-meng documents contain legal codes "intended ... to constrain the actions of officials," translator Hulsewé explains. One section lists rules of behavior for officials, as well as record-keeping and inspection guidelines to keep track of officials. Another section focuses entirely on official stores (granaries etc.) and their records. A further section consists of guidelines instructing local officials how to interpret Qin laws correctly, and another is a guide to educating officials on how to conduct investigations and transmit results properly. Thus, Lewis concludes, "the primary focus of these actual legal documents was the rigorous control of the actions of those legal officials charged with overseeing the people." Another document found in the same Yün-meng tomb, titled "On the Way of Being an Official," reinforces Lewis's point. The text describes the ideal official behavior as representing the central government in specific localities "without interposing one's own will or ideas."⁶⁹⁷

In addition to these rules of conduct, there was also constant hierarchical supervision of local administrators by their superiors, as well as "ad-hoc" monitoring by special commissioners. Eventually the role of the commissioners was rendered permanent. To the civil governor and military commander responsible for each commandery, the First Emperor added an imperial inspector directly responsible in front of the emperor. The First Emperor also became personally involved in the monitoring process. Starting in 211 B.C., he went on five "national inspection tours." One goal of such tours was to scrutinize the work of lower-level state officials, and we can assume that previous Qin rulers engaged in similar, though perhaps less comprehensive, tours.⁶⁹⁸ Moreover, to limit the likelihood that local officials might attempt to build their own local fiefdoms and take over the former role of the nobility, Qin hired a number of foreign

⁶⁹⁶ Ibid., 30.

⁶⁹⁷ *Yün-meng Manuscripts* A1, 12, 31-33, 77, 92-3, C8; Lewis 610.

⁶⁹⁸ Bodde 55.

experts to hold local administrative positions. Duke Xiao (ruled 361-338 B.C.) reportedly issued a decree "inviting the best minds from all over the world to help him."⁶⁹⁹ For example, Qin created an office of prime minister in 328 B.C., and until the end of Qin in 206 B.C., "most people holding this position were foreigners."⁷⁰⁰

By abruptly seizing state offices from the aristocracy, the bulwark of feudal power, and replacing it with a corps of obedient, well-supervised bureaucrats, Qin was therefore able to rapidly develop a powerful, highly centralized state apparatus that gave it the tools necessary to implement military and economic reforms and mobilize the Qin state's resources toward hegemony. Bureaucratic control, centralization, and the professionalization of government power soared in the Qin state because due to its late development and "barbarian" background compared to the other ancient Chinese states, it had fewer nobles hanging onto the traditional, feudal organization of power to begin with, and the radical opening of the military to the masses and introduction of merit-based appointments chipped away their remaining roles. The social origin of Chinese state officials is telling: only 26% of all state officials came from non-aristocratic backgrounds during the Spring and Autumn Period, but the number rose to 55% during the Warring States Period under the impulse of Qin.⁷⁰¹ The espousal of legalism to its fullest extent served as the cement that brought all of Qin's reforms together by giving the state the legal tools of central control.

In addition to strict political, economic, and military oversight, the legalist philosophy also encouraged Qin leaders to enact rigid social controls that contributed to Qin's rising power and thus hegemonic success but may have also planted the seeds of the Dynasty's rapid downfall.

⁶⁹⁹ Hui 65.

⁷⁰⁰ Kiser and Cai 528-9.

⁷⁰¹ Ibid., 521.

4. Social Control and Organization (IV6.4)

Legalism not only provided Shang Yang and other Qin reformers the intellectual underpinnings to bring down Zhou feudalism and develop a strong, centralized state, military apparatus, and growth-oriented economy, it also suggested the rulers establish strict and effective control over its population. Only by regulating people's behavior and their natural tendency toward disorderliness and excesses can a state channel its resources in the most effective way and generate the most power, legalists preached. Han Feizi states that good governance had nothing to do with benevolence and righteousness, but with clear laws and regulations, enforced with rewards and punishments, which amount to "applying rouge, powder, and paint [i.e., makeup] to the state" and keeping it attractive. Indeed, "if the ruler does not apply the proper laws and procedures within his state, ... his state will never become powerful and well-ordered," Han Feizi adds.⁷⁰² *The Book of Lord Shang* echoes the same lesson. "Benevolence and righteousness are not sufficient for governing an empire." A ruler must "also ha[ve] a law by which he compels the whole empire to have good faith." *The Book* adds that "the basis of the people is the law ... if the basis is not solid, then people are like flying birds or animals... Therefore, a good ruler instruct[s] the people by means of the law." Legalists argued that while individuals may naturally exhibit some strengths and qualities, these qualities are often inherent to the person and thus not shared by others. The only way to create cohesion and coordinate individuals' qualities and strengths to exploit them to their fullest extent is to institute a clear set of laws to be followed by all, complete with rewards and punishments to prevent deviance. Even the most military powerful state could not sustain its power and prevail in the long run without a backbone of strong laws regulating people's behavior. "To have a numerous population and a

⁷⁰² Han Fei Tzu 118-22, 127.

strong army is the great capital of an emperor or king, but if he does not have clear laws by which to keep them, he is next door to peril and ruin," *The Book* concludes.⁷⁰³

Shang Yang thus developed a legal code for Qin that was the most far-reaching of any of the Warring States. While most other ancient Chinese states were traditionally governed by a large number of intertwined, unwritten customs and rites, Shang Yang promulgated a written code that distinguished itself by its highly detailed rules of behavior, particularly harsh penalties, and equality of application regardless of social rank. By institutionalizing those rules in writing, Shang Yang essentially made them invulnerable to easy reversal. Sima Qian explains that in 357 B.C., "Wei Yang [i.e., Shang Yang] spoke to the Duke [i.e., Xiao, then-ruler of Qin], urging him to change the laws [and] impose penalties... Duke Xiao approved these proposals." Other officials disagreed with the formalization of a legal code "but in the end the Duke put Wei Yang's law in effect," Sima Qian adds.⁷⁰⁴

Although China likely had instances of partially codified law prior to Shang Yang, Qin was the first state for which we have physical remnants of legal documents. The Yün-meng original documents indicate that Shang Yang's laws were much more elaborate than previous laws we have secondary accounts of. The Yün-meng papers include not only the administrative rules mentioned above, but also an array of penal laws. However, they did not comprise the entire Qin code but only the rules which were used daily in the local district of the prefectural clerk in whose tomb they were found. The clerk, who was born in 262 B.C., likely tried criminal cases in addition to administering government granaries and dealing with public labor and conscription issues, explaining why both administrative and penal laws were found in his possession. Just like the administrative rules detailed in the previous section, criminal matters

⁷⁰³ *The Book of Lord Shang* 243, 285, 294, 309, 311.

⁷⁰⁴ Sima Qian xii, 24; Hui 103.

followed strict procedures, which included an investigation with detailed reports, including body location, if any, clothes worn by the victim, measurements, etc, and the interrogation of suspects. In an unusually modern twist, the Yün-meng papers note that the use of torture for interrogation was acceptable only in a restricted number of cases and was considered inferior evidence due to the risk of false confessions. The documents differentiate between manslaughter or involuntary killing, and murder, which includes premeditation, and provide for different sentences for each. Punishment for robbery was gradual depending on the value of the items stolen and the number of thieves involved, and accessory to robbery was punished as severely as robbery itself. Another modern measure provides for the right of appeals after a trial.⁷⁰⁵

Despite those modern measures, the Qin penal code provided disproportionately harsh punishment for criminal offenses, as the Yün-meng documents suggest. The death penalty was a frequent punishment, generally in the form of a beheading in the town center. Other severe punishments included dismemberment, and drowning for criminal lepers. Lesser crimes were punished with hard labor sentences, often combined with physical mutilation like shaving, tattooing, amputation or even castration in the most severe cases. Hard labor consisted of construction work for men, as well as wood gathering and various guard duties, and grain pounding and rice sifting for women. Surprisingly the Yün-meng papers do not list sentence lengths, leading some authors to speculate that one could be sentenced to hard labor for life, thus becoming a sort of "government slave." Other punishments included banishment or exile, again with no duration mentioned but the papers provide the example of someone sentenced to exile in a newly conquered area. Fines were primarily a punishment for officials. Amounts were not in cash but in armaments (armor, shields, scale for suit of armor, etc.) and it is difficult to evaluate their value. Officials could also be fined with extra years of military service. Qin also used the

⁷⁰⁵ *Yün-meng Manuscripts* D1, 15-16, 60, 70-2, 95, E1, 2, 20, 22.

collective responsibility system already implemented in the military to ensure popular subjection to the criminal laws. Groups of five households shared the responsibility for the behavior of its members collectively "unless they reported the crime to officials." Not reporting it was a crime itself, as was leaving one's town without proper authorization. The result in terms of central control was drastic, as Kiser and Cai explain:

This kind of social control mechanisms held everyone under fairly constant local supervision and thus made tax collection and military recruitment much easier. The use of collective responsibility increases efficiency by encouraging self-monitoring, shifting the costs of monitoring from the state to local subject.⁷⁰⁶

In the end Shang Yang's legal reforms served to "break down the barriers that intervened between the individual and the state." They created strict "legal responsibility both horizontally among family members and five-family units and vertically" between individuals and state officials, maximizing social control.⁷⁰⁷

Just as Qin abolished privileges for job accession, it also applied legal rules equally to all, as prompted by the legalist philosophy. "The law no more makes exceptions for men of high station than the plumb line bends to accommodate a crooked place in the wood ... When faults are to be punished, the highest minister cannot escape; when good is to be rewarded, the lowest peasant must not be passed over," writes Han Feizi. The Zhan Guo Ce thus said of Shang Yang's legal reforms, "everywhere his laws were carried out to the letter and there was justice without favor. Punishment applied equally to the great and the powerful and reward was not limited to the favored and the well-born." *The Book of Lord Shang* similarly praises several historical examples where the law was applied equally to all, regardless of rank or ability, such as when a

⁷⁰⁶ Kiser and Cai 530; *Yün-meng Manuscripts* C8, D6, 27-8, 56-7, 62, 71, 93, 99, 101, 146, 151, 154, 162, E17.

⁷⁰⁷ Yates 410.

duke punished his favorite officer, or his brothers, for breaking a rule.⁷⁰⁸ In most instances the equal application of the law was duly followed. Sima Qian mentions several cases where high officers were punished, even with death. In October 257, for example, high ranking officer Bai Qi, Lord Wuan "was ... reduced to the rank of common soldier, and exiled to Yinmi" after having committed a crime. But he repeated the offense, so in December he was "put to death." It must be noted, however, that equal application of the law did not seem to reach the ruler himself, although in theory it should have. Sima Qian describes, for example, that after Duke Xiao died in 338 B.C., his son Huiwen took power without following the procedures and should have been punished for his violation. Huiwen refused and had his tutor penalized instead. In addition, the Yün-meng documents mention a "redemption" system that clearly favored the wealthy. Qin law allowed convicts to pay off their sentence, even a death sentence, with a fine. Sima Qian himself was sentenced in 99 B.C. for criticizing the Han emperor's brother-in-law, and, unable to afford a redemption fine, he was jailed and castrated.⁷⁰⁹

Despite the relative judicial equality of Qin citizens, the harshness and severity of Shang Yang's law and the strict social control it engendered caused much grief among the population and was in large part responsible for the rapid downfall of the Qin Dynasty. It took the population years to become used to the new laws, Sima Qian points out, as "the people at first complained bitterly." The Qin code reflected the legalist idea the people could not be trusted to make enlightened decisions and that political rule should be strictly applied from the top down. In *The Book of Lord Shang*, a minister to Duke Xiao tellingly reminds the Qin ruler that "he, who accomplishes a great work, does not take counsel with the multitude." Unlike other rising hegemonies such as Rome and the Mongols, successive Qin rulers failed to enact policies

⁷⁰⁸ Han Fei Tzu 28; *The Book of Lord Shang* 280-2; *Chan-kuo Ts'e Ch'in* 46 (81).

⁷⁰⁹ Sima Qian 25, 31-2; *Yün-meng Manuscripts* A68, 72, C20, D3, 25, 52, 94, 136, 164.

promoting the education, health, and welfare of their subjects, or the role of women, and were instead focused exclusively on raw power and productivity. The First Emperor, who applied legalism to the letter, "seems to have seriously underestimated the amount of bullying and oppression his people would bear," Burton Watson writes. Eventually his successor was toppled by a popular revolt that started in 209 B.C., and Qin's successor Han criticized the social oppressiveness of the Qin rule as the cause of its lack of popular support. While keeping large parts of Shang Yang's legal code intact, Han consequently proceeded to soften some of its most controversial aspects.⁷¹⁰ Chia Yi, the Han historian, captures the essence of Qin's excessive social control:

Yet after it [i.e., Qin] had become master of the whole empire ..., a single commoner opposed it and its ancestral temples toppled, its ruler died by the hands of men, and it became the laughingstock of the world. Why? Because it failed to rule with humanity and righteousness and to realize that the power to attack and the power to retain what one has thereby won are not the same.⁷¹¹

Conclusion

Nevertheless, the originally small, backward state of Qin achieved an incredible deed by subverting the balance of power of ancient China and becoming the region's hegemon in the record time of just over a century. Its spectacular growth was shielded by the ineptitude of the more established regional powers to cooperate and put together a balancing coalition, an obstacle that Qin leaders did not fail to cleverly exploit and reinforce with deceptive, divide-and-conquer techniques. Qin's six fellow Warring States mostly found cooperation unattainable altogether because their constant quest for immediate gains at the expense of their neighbors led to bitter

⁷¹⁰ Sima Qian xiii, 24; Han Fei Tzu 12; *The Book of Lord Shang* 169.

⁷¹¹ Chia Yi 48.

internecine feuds and profound distrust. The few times Qin's potential balancers succeeded in establishing a countervailing alliance, their efforts were soon derailed by the same collective action problems that prevented most alliances from forming in the first place. In addition, frequent bandwagoning with Qin, again encouraged by the rising hegemon's tricks, also plagued Qin's neighbors and reinforced their suspicion of each other.

Simultaneously, Qin's ground-breaking legalist-inspired reforms equipped the rising hegemon with the tools it needed to grow and maximize its power, enabling it to defeat all of its rival Warring States one after the other. Its military reforms provided it with a large reservoir of soldiers and the most tightly-organized and well-disciplined army in the region, while its economic reforms engendered widespread agricultural and industrial growth to support Qin's ambitious expansion plan. The rising hegemon's administrative reforms completed the picture by providing the Qin state a firm grip over its entire territory. But these successful reforms came at a price. Although the tight centralized control generated by the reforms helped Qin win its empire, it also created an oppressive social environment that prevented the hegemon from building a lasting imperial rule, instead making it a fragile empire. The Qin Dynasty, a hierarchic, top-down rule without the anchor of popular support, quickly crumbled after unifying China, as grassroots revolt spread like wildfire.

U.S. Expansion into the Western Hemisphere, 1865-1914

The United States represents an entirely different type of hegemonic rise and rule than the Qin's, although both share the same regional geographic coverage. From the end of the Civil War until the beginning of World War I, the United States grew from a local power to a regional hegemon, expelling all rival great powers from the region and developing an economic, political, and military monopoly of influence over the entire Western Hemisphere—the American continents and adjacent island territories in the Caribbean and Pacific.

The U.S. case draws particular significance because it represents the first and only instance of hegemony in recent history. As John Mearsheimer points out, "the United States is the only state in modern times to have gained regional hegemony."⁷¹² As such, the American hegemon developed unique characteristics. Unlike all previous cases, America's control over its region was characterized not by territorial conquest and foreign invasion, but by economic and political influence over its target states, with occasional military coercion. While the United States expanded its immediate borders during the first half of the 19th century and subsequently annexed some territory in the Western Hemisphere, it mostly took control of the region informally via diplomatic, economic, and occasionally military pressure on local governments and rival great powers. This informal type of hegemonic control, as opposed to the traditional, conquest-based hegemony, is a new development of the 19th century, according to John

⁷¹² John J. Mearsheimer, *The Tragedy of Great Power Politics* (New York: W.W. Norton & Company, 2001), 236.

Gallagher and Ronald Robinson, who analyzed the shift toward an 'informal empire'. "The conventional interpretation of the nineteenth-century empire continues to rest upon study of the formal empire alone, which is rather like judging the size and character of icebergs solely from the parts above the water-line," Gallagher and Robinson deplore. Instead, particularly after 1880, which corresponds to the height of America's rise, there has been a "qualitative change in the nature of ... expansion," where formal growth still occurred—primarily in the form of base acquisition—but was increasingly supplanted by "indirect methods."⁷¹³

Those methods included, among other things, creating political bonds or unofficial protectorates with the target countries, occasionally providing military assistance to secure their political stability and independence, promoting trade and economic exchange with preferential tariffs and interest rates, and financing development projects such as railway lines in the target countries. Such exchanges generated an informal dependency upon the United States, the provider of markets, funds, and protection. "The economic importance—even preeminence—of informal empire in this period has been stressed often enough," Gallagher and Robinson add, but "what was overlooked was the inter-relation of its economic and political arms; how political action aided the growth of commercial supremacy, and how this supremacy in turn strengthened political influence." Only when informal means failed to secure American interests did the United States resort to direct rule—the principle was to "extend control informally if possible and formally if necessary." Informal control carried great benefits. For example, it had "the advantage of saddling foreign governments with the liability of rule whilst allowing [the rising hegemon] the commercial [and strategic] interest," in essence allowing for "imperialism on the cheap," as Gallagher and Robinson point out. Because of advances in transportation and

⁷¹³ John Gallagher and Ronald Robinson, "The Imperialism of Free Trade," *The Economic History Review* VI.1, 2nd Series (1953), 1-3.

communication technology, in addition, costly direct rule—which required prolonged troop presence, scores of administrative personnel, and long-term financial investment—was in most cases no longer necessary to generate control.⁷¹⁴ Moreover, because the rise of the nation-state throughout the 18th and 19th century reinforced the notions of sovereignty and self-determination as a basis for political rule, and much of the countries in the region had already gained their independence from (mainly Spanish) colonial rule and governed themselves, subtler informal control was more readily acceptable and thus provided a safer strategy for a hegemon in infancy. Outright conquest would certainly have precipitated generalized condemnation and a perception of more imminent threat, perhaps even sufficient for potential balancers to overcome their cooperation issues and engage in fiercer balancing.

The shift toward informal hegemony led the United States to clash not only with the various independent states of the continent, mostly medium to small powers at times resenting the increasing American influence, but also with rival European great powers who also possessed or sought their own economic and political influence over the same areas. Unlike other regional hegemonies like China or Rome, the United States' most serious challengers for hegemony were not the local countries, which were all secondary powers wielding too little power to be able to make a difference in the outcome, but instead originated and operated from another region. The United States' primary rivals were European great powers—principally Britain, Germany, Russia, and Spain—and, unlike the United States at the time, they contended for global influence and used the American region as one among several theatres. Several European great powers had outposts on the American continents: while most regional states had gained independence from the continent's prime colonizer Spain by the time of the U.S. rise to power, a number of colonies remained. Spain retained control over Cuba, Puerto Rico, the Philippines, Guam, Ceuta and

⁷¹⁴ Ibid., 4-7, 13.

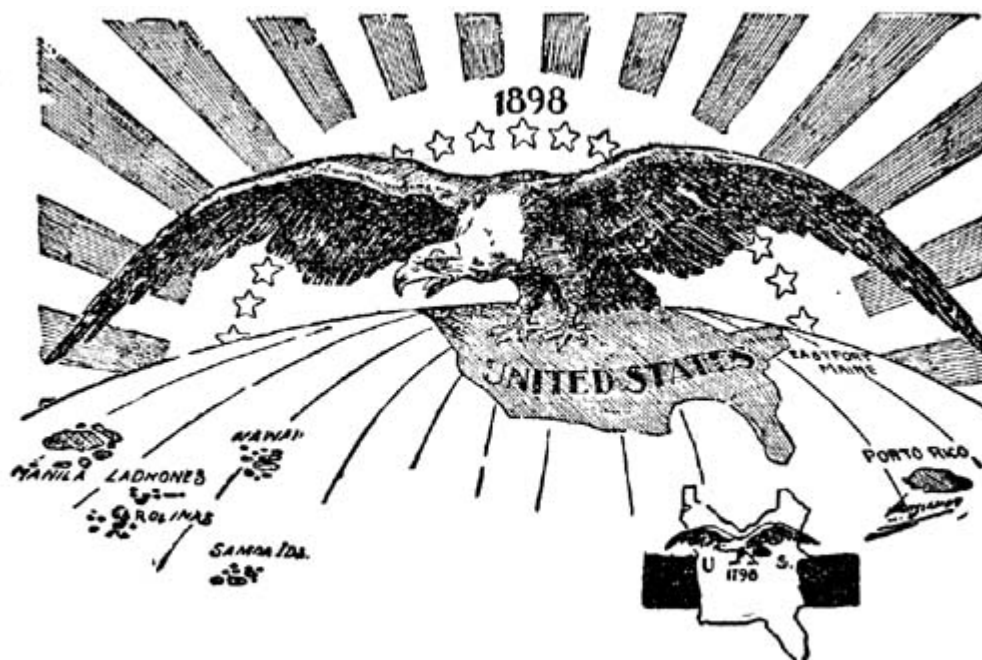
Melilla and a few other Pacific islands. A handful of states in Central and Latin America also remained colonies: Guyana was British, Suriname was Dutch, French Guiana was French, and British Honduras (now Belize) was leased by Spain to Britain. The story is a little different with the Caribbean, which for the most part remained in colonial hands throughout the 19th century with the exception of Haiti and San Domingo: Britain owned a number of islands and archipelagos there, including the British West Indies, the British Virgin Islands, Jamaica, and the Bahamas; there was also the French West Indies, the Dutch West Indies, and the Danish West Indies, which eventually became the U.S. Virgin Islands after Denmark sold the islands to the United States in 1917. This intertwining of regional and global systems denotes the evolution toward a less segregated, increasingly globalized world already hinted at in the decline of conquest as the means of hegemony. The involvement of extra-systemic great powers was readily embraced by actors in the United States, where editorialists and lobbyists played an increasing role in the 1880s and 1890s in stirring up the population to follow European foreign policy developments and inciting popular sentiment against the main rivals Britain and Spain.⁷¹⁵

Thus, this case is unique because the United States achieved regional hegemony by simultaneously developing informal control over the local, secondary powers, largely without conquests or military defeats, and evicting its extra-systemic great powers rivals from the region so they would not be able to compete for influence. Given these unusual characteristics, do the twenty-one factors of balance of power failure fare differently than in the previous cases? In reality, the United States succeeded in its hegemonic bid fundamentally for the same reasons that its ancient counterparts did, though some variables seem more or less potent in this case. The potential great power balancers were unable to cooperate to put an end to the American spread of

⁷¹⁵ Ernest R. May, *Imperial Democracy: The Emergence of America as a Great Power* (Chicago, IL: Imprint Publications, 1961, repr. 1991), 10.

influence, but it was due as much to their buckpassing and lack of sufficient interest as to their lack of trust, and their problems were reinforced by frequent bandwagoning, both by the secondary powers and by at least one great power. The United States, on the other hand, possessed the advantage of geography, as well as the strongest economy and the appeal of its informal tools of influence, and while it did not per se engage in dramatic military innovations, its military growth—particularly its naval growth—was exponential both in time and scope.

Illustration 8.1: "Ten Thousand Miles from Tip to Tip"



Ten thousand miles from tip to tip.—Philadelphia Press.

Source: Originally published in the *Philadelphia Press* (1898) and reprinted in William McKinley and Marshall Everett, *Exciting Experiences in our Wars with Spain and the Filipinos* (Chicago: The Educational Co., 1900), 272.

Boundary of Inquiry

The U.S. rise is also unusual because it followed a two-step growth process: a first, long period focused on economic development (1865-1880s) paved the way for military growth and

expansion of influence abroad (mid-1880s-onward). The United States reached hegemonic status in the Caribbean between 1895 and 1906, after imposing its will over Britain and Spain, the two other great powers present in the area, and establishing a free hand to intervene in the local states' internal affairs. It extended its hegemonic reach to Central America between 1910 and 1913, by turning most of the countries there into quasi-American protectorates. The exact date of U.S. hegemony over South America is harder to pinpoint because the distance made it more difficult for the United States to control. By the beginning of World War I, however, the European great powers became exclusively focused on their own continent and left the Western Hemisphere entirely in the hands the United States. With dramatic improvements in power projection but no serious rival left, the United States had soon finished transforming the Western Hemisphere into one vast American lake.

The rise of the United States to dominate the Western Hemisphere began with industrialization and the United States' dramatic post-Civil War economic boom, which set the stage and built up the country's resources for its later economic, political and military involvement abroad. Without this first phase, the rising U.S. hegemon would have been short in capabilities to succeed in the second growth phase and drive the already established European powers off the American continent. While most rising hegemons have no choice but to simultaneously combine economic and military growth as well as expansion to minimize balancing risks, the United States was able to become an economic powerhouse while keeping the interests and military capabilities of a small power because its security was safeguarded by two oceans that separated it from great powers in Europe and Asia, and the naval supremacy of the British Navy, whose control of the oceans ensured that no other threat could approach American shores.

Fareed Zakaria explains that throughout most of the 19th century, the United States was "protected from the vagaries of the world by the virtually omnipotent Royal Navy" and essentially had a "tacit," de-facto alliance with Britain. Even though American statesmen remained cautious in their dealings with Britain, reminded of the war of 1812 and British hostility with the North during the Civil War, and anti-British sentiment was rife among the American public, there were few actual flare-ups between the two states until the 1880s. Overall, "the United States enjoyed tremendous economic growth in an environment devoid of threats to its security." In other words it was able to develop its economic might "on the cheap," as Mearsheimer writes. Its post-Civil War growth was "meteoric," and "it clearly possessed the economic wherewithal to become a great power and compete around the globe with Europe's major powers." Yet, its military remained "tiny." The U.S. army was 14th in the world, after Bulgaria, with a force of only 25,000 in 1880. Given the expanse of coastal territory in the United States, the size and state of the American navy was "ridiculous," the smallest among all great powers, behind that of Italy. During this early phase, the only U.S. territorial acquisition was that of the Midway Islands, annexed in 1867, but the islands were uninhabited and did not belong to any European rival. Although the United States considered taking over Santo Domingo, Cuba, and Hawaii several times in the 1870s, and was seeking an American-controlled isthmian canal, it never acted on it. Things changed in the 1880s, when the United States initiated a massive naval buildup and became significantly more assertive abroad, initiating the second, proactive phase of its rise.⁷¹⁶

The causes of the United States' change toward a more active foreign policy are a source of debate. Some see the origin in U.S. domestic politics. Zakaria argues that the end of the

⁷¹⁶ Fareed Zakaria, *From Wealth to Power: The Unusual Origins of America's World Role* (Princeton, NJ: Princeton University Press, 1998), 43, 67-75, 82, 88-91, 137; Mearsheimer 235.

structural hostility between executive and legislative branch and the rise of executive power in the 1880s enabled the president and cabinet to bring foreign policy plans to fruition without congressional obstruction. John Grenville and George Young suggest that partisan politics became tamer and individual leaders' personalities more conciliatory in the 1880s. For others, such as Nicholas Spykman and Walter Lippmann, the 1880s witnessed a change in the international environment, which became adverse to the United States with increasing threats of European, particularly German, intrusion, and the retrenchment of Britain's protective Navy as the overstretched British empire was forced to compromise and focus on its key strategic possessions in the face of growing great power hostility. As a consequence, Robert Art explains, "the United States found itself forced to become more active militarily so as to reshape the environment in ways that better suited its interests." Grenville and Young point out that U.S. leaders feared that European powers "would establish bases [on the American continent] for future expansion in the Western Hemisphere," citing "rumors" that Germany sought to acquire bases in Central America and Cuba in the mid-1880s, for example. America's declining security in the face of those threats led to its massive military development and increased foreign activism. Art concurs, stressing that between 1895 and 1905 Germany, still a power "on the make," was "seeking ... footholds" on the American continent, turning the United States from a "consumer to a provider of security" on the continent.⁷¹⁷ Henry Cabot Lodge and Theodore Roosevelt, for instance, believed that the growing German power and interest outside Germany was a threat, especially after the Germans tried to take over Samoa in 1889. Lodge was worried

⁷¹⁷ Robert J. Art, *A Grand Strategy for America* (Ithaca, NY: The Century Foundation, Inc. and Cornell University Press, 2003), 181-2, 185; John A.S. Grenville and George Berkeley Young, *Politics, Strategy, and American Diplomacy: Studies in Foreign Policy 1873-1917* (New Haven, CT: Yale University Press, 1966), xi-xv, xvii-xviii, 7, 39-40, 77; Nicholas John Spykman, *America's Strategy in World Politics: The United States and the Balance of Power* (Archon Books, 1942, repr.1970); Walter Lippmann, *U.S. Foreign Policy: Shield of the Republic* (Boston, MA: Little, Brown, and Company, 1943); Zakaria 87-8.

that the German Kaiser might act with even less restraint a year later after he dismissed his seasoned chancellor Bismarck:

What if Germany then [at the time of the Samoan crisis] had been under her present guidance, passion and not reason might easily have ruled, and we might have found ourselves plunged suddenly into a naval war, and have seen our great coast cities laid in ruins before we could gather means to defend them.⁷¹⁸

The declassification of German and British archives shows that American fears of European meddling in America were "exaggerated." Britain did not seek aggression and Germany, while it may have wished to, "did not possess the means to sustain [an expansionist] policy" across the Atlantic. But German leaders may have thought they did. The German archives confirm that there were military studies of an invasion of the United States. And in December 1899 "the Kaiser personally instructed Admiral Otto von Diederichs, Chief of Staff of the Navy, and ... Count Alfred von Schlieffen, Chief of Staff of the Army, to prepare a war plan against the United States." The Kaiser believed Cuba should be the first target of an expeditionary force, while Schlieffen preferred Cape Cod, and the German naval representative in Washington was "instructed to reconnoiter Massachusetts and especially Cape Cod." His findings were not favorable to a landing there, and the Germans eventually adopted Puerto Rico instead as a first landing target. A further German provocation off of Venezuela in 1902 certainly seemed to support the view of Germany as a threat even though it would not have had the capabilities to sustain its efforts.⁷¹⁹

Regardless of its causes, the first outward signs of growing U.S. expansion occurred in the early 1880s, when the United States developed a mounting interest in happenings around the

⁷¹⁸ Henry Cabot Lodge, *Congressional Record*, 51st Congress, 1st Session (April 10, 1890), 21, 3268, qt. in Grenville and Young 212-215.

⁷¹⁹ Grenville and Young 236, 304-6, 309.

Hemisphere and began building up its capabilities. During this second period of rise, the United States sought consistent diplomatic and economic involvement with its neighbors and a curtailment of European involvement around the continent. For this purpose it set up a number of stations and naval bases in the Caribbean and Pacific as a bulwark against European encroachment. By the early 1890s its involvement abroad became even more palpable, both economically and diplomatically, and it turned decisive when the United States provoked Britain in a crisis over Venezuela in 1895-6 and forced it to back down, eliminating its most powerful great power rival in the region, then went to war with Spain over Cuba in 1898 and removed another great power, and finally in 1904 established its monopoly over European debt collection in Latin America and the Caribbean. Those interventions marked a "milestone in the emergence of the United States as a world power."⁷²⁰

By 1904-1905, the United States had essentially supplanted the European great powers' influence and established hegemony over the Caribbean, and at least political and economic hegemony over the rest of the Hemisphere. But the United States had not yet reached full hegemony besides the Caribbean, and it continued to pursue an "activist" foreign policy under Presidents Theodore Roosevelt and Taft, Zakaria argues. Indeed, at the onset of the 20th century cooperative balancing efforts by the European great powers could hypothetically still have curtailed America's spread over the region if they had not instead spiraled into increasing hostility as events in Europe moved toward World War I. When the Great War started, the European great powers were drawn into a fight for their survival in Europe and forced to withdraw their quest for influence across the Atlantic for good, leaving the Western Hemisphere devoid of any great power challenger to the United States. At the same time the U.S. Navy

⁷²⁰ Ibid., 94, 157.

became unchallenged. By 1914, thus, the United States' military hegemony over the Western Hemisphere was unequivocal.⁷²¹

Timeline of U.S. Expansion

Although the United States did not spring into action until the 1880s, its yearning for control over the Western Hemisphere is obvious from much earlier in the century. The Monroe Doctrine, proclaimed by President Monroe in 1823, was perhaps the first acknowledgment of such an intent, at a time when the American continent was threatened with incursions both from Asia via the Pacific (Russian expansion into the Aleutian Islands, Alaska, and northern California) and from Europe via the Atlantic (primarily France and Spain at the time, and economic competition with Britain). The Doctrine established the United States as the protector of independent Latin American political entities against the colonial encroachments of European powers. At the time, however, the United States still lacked the means of competing with European colonialists. But to American statesmen during the United States' early rise, "the maintenance of the Monroe Doctrine was a principle that the United States now, more than ever, could not afford to abandon" because they feared that "the European powers would parcel out Latin America as they were already partitioning Africa."⁷²² Throughout the United States' rise and in parallel with the growing U.S. involvement in the continent's affairs, then, the Doctrine's meaning gradually shifted from its original negative purpose of preventing European intervention to a larger, positive role for the U.S. as the sole policing authority over the continent—by the time of the Platt Amendment of 1901 and the Roosevelt Corollary of 1904, the Doctrine justified U.S. intervention in the domestic politics of its independent neighbors. It

⁷²¹ Mearsheimer 239; Zakaria 154, 180.

⁷²² Grenville and Young 223.

transformed, in essence, from a guarantee of independence and non-intervention to a promotion of intervention and interference with independence.⁷²³

The chronology of U.S. attempts at securing economic and political influence over the continent clearly shows an escalation in claims and control. In the 1880s U.S. foreign policy became more vocal both toward local secondary powers and especially toward rival great powers, although the United States still remained cautious and stopped short of provocation that might trigger a direct conflict as its military buildup was still in infancy. In the mid-1880s there was a large movement in the United States in favor of the free, unlimited coinage of silver as the U.S. gold reserve had been waning and gold supplies were largely controlled by Britain. American politicians like Cabot Lodge denounced the great powers, particularly Britain, "for refusing to depart from the gold standard," creating hostile sentiments towards Britain. In the mid-1880s resentment also grew with China, after violence spread against Asian immigrants along the West Coast and the U.S. Congress unilaterally froze Chinese immigration and travels in 1888. At the same time, American relations grew antagonistic with Canada, a self-governing British colony since 1867, after incidents erupted over respective fishing waters. The United States rejected a British-brokered treaty in 1888 and the President encouraged Congress to launch an economic war against Canada. Congress did not oblige, but the two countries remained at odds, particularly since some statesmen like Cabot Lodge believed "in the absorption of Canada" into the United States. Finally, when in 1889 the Germans, who shared some "ill-defined rights" over Samoa, sought to drive the United States and Britain out of the islands, the United States engaged in some saber-rattling, hinting it was ready to go to war over the issue, successfully causing the Germans to back down and agree to a tripartite protectorate over the

⁷²³ Dexter Perkins, *The Monroe Doctrine 1867-1907* (Gloucester, MA: Peter Smith, 1966), 396-7; Spykman 68, 352.

islands. Eventually the United States and Germany each annexed parts of the archipelago ten years later.⁷²⁴

In the 1890s as its capabilities soared, the United States grew increasingly assertive. After a riot caused the death of American sailors in Chile in the winter of 1891-2 and the Chilean government dragged its feet in issuing an apology to the United States, the U.S. government "almost precipitated a war with Chile" over the issue, which was only prevented by swift Chileans regrets and reparation offers. Throughout the 1890s the United States also became more involved in Asia, attempting to secure markets and contracts, amid fears it would be left out as the European great powers were "carving out spheres of influence" and annexing or leasing coastal ports, taking advantage of a weakened Chinese government. When Russia and Germany considered closing their ports to other great powers in 1899, U.S. Secretary of State John Hay issued the famous Open Door policy declaration, demanding and obtaining equal commercial rights for all powers within the respective spheres of influence and a guarantee of Chinese territorial integrity. When the Boxer Rebellion erupted in 1900 in China in response to the de facto foreign occupation, President McKinley sent 5,000 troops to China, and Hay was again able to force the European powers to maintain open trade and Chinese integrity with a second Open Door declaration. The limit of U.S. power projection outside the American hemisphere became clear a year later, however, when the United States rescinded its claim for its own naval base at Samsah Bay after Japan's vehement opposition. The 1890s also witnessed the annexation of Hawaii, an issue that had already come up a few times in previous decades but the United States had never acted upon. As May explains, because of the large U.S. population living on the islands and a naval base and preferential tariffs obtained in 1887, by the early 1890s Hawaii was

⁷²⁴ U.S. Dept. of State, *Papers Relating to the Foreign Relations of the United States – Transmitted to Congress Dec. 2, 1902* (Washington, D.C.: Government Printing Office, 1903); McCormick to Hay, U.S. Delegation in Vienna, March 7, 1902), 444-7; May 9-10, 33; Zakaria 160; Grenville and Young 45-7, 51-69, 224.

"already Americanized and virtually a protectorate." When in January 1893, American residents in Hawaii deposed the native Hawaiian queen, the U.S. government started working toward an annexation treaty, which was eventually passed and led to the takeover of the islands in 1898.⁷²⁵

A number of incidents throughout the 1890s also saw the United States progressively hold its ground more firmly against its most powerful great power rival, Great Britain. The first incident occurred in Nicaragua, where the local troops were attempting to subdue the Mosquito Coast, a semi-independent Indian reservation that appealed to Britain for help. After Britain sent a battleship and landed marines on the reservation in 1893, the United States, which had significant economic and strategic interests as the area constituted the eastern end of a possible isthmian canal, threatened intervention, leading Britain to withdraw. In 1894 American troops occupied the area and turned Nicaragua into a "de facto protectorate" of the United States. That same year a revolution erupted in Brazil that threatened to bring down the American-supported government. When Britain sided with the revolutionaries, the United States sent five warships to Rio de Janeiro, forcing the rebels to back down. Another incident occurred in 1895, when the Nicaraguan government arrested a British diplomat and Britain retaliated by landing troops and taking possession of a Nicaraguan customs house. Because the United States had been negotiating of a deal with Nicaragua to build an isthmian canal through the country, the British intervention sparked outrage in the United States, and as a result "Britain promptly withdrew the landing party."⁷²⁶

The escalating incidents with Britain finally culminated in a 1895-6 near-war crisis over Venezuela. The crisis started with a long-standing border dispute between British Guiana, a

⁷²⁵ Walter LaFeber, *The New Empire: An Interpretation of American Expansion 1860-1898* (Ithaca, NY: Cornell University Press, 1963), 380-1; May 10, 13, 25-6, 54; Zakaria 79, 140-2, 161-4; Grenville and Young 96-7, 221-2, 267-8.

⁷²⁶ Perkins 42-3; May 33; Zakaria 146-8.

colony of Britain, and Venezuela, which appealed for U.S. support, over a poorly defined section of border, an area rich in gold that could have potentially helped the United States secure enough gold to "ease the shortage of specie that had given rise to demands for silver coinage." The United States demanded that Britain submit to international arbitration to solve the dispute, and argued that Venezuela was in the Monroe Doctrine's sphere and that the United States would intervene if it deemed it necessary. When Britain refused arbitration and replied that the Monroe Doctrine did not apply, President Cleveland delivered a December 1895 address to Congress in which he "threatened Britain with war if she refused to acknowledge American supremacy ... [in] the Western hemisphere." Britain backed down once more and agreed to arbitration and the broad definition of the Monroe Doctrine. Finally in 1899, an American commission drew the boundary in the disputed area, somewhat favorably to Britain.⁷²⁷

The yielding of Britain was closely followed by that of Spain in 1898 in a dispute over Cuba. Spain was a great power on the decline, but its demise as a colonial power required American military intervention. Cuba, with Puerto Rico, were the only Spanish colonies on the American continent that had not gained independence by the late 19th century. A revolutionary movement erupted there in February 1895, seeking independence from Spain, and soon turned into a bloody civil war as the Spanish authorities repressed the rebels with particular brutality. From the onset the United States was sympathetic to the revolutionaries' quest to establish their own rule and a number of basic freedoms; moreover, many Americans had investments in Cuba, making the civil war costly for the U.S. economy. Spain refused U.S. mediation and was disinclined to grant the rebels any sort of autonomy. After riots broke out in January 1898 in Havana, the United States sent its warship *Maine* to Cuba to keep an eye on developments there, but just a month later it was sunk, killing all U.S. sailors on board, in a mysterious explosion. In

⁷²⁷ May 34-42; Zakaria 148-52; Grenville and Young 176-8.

April, the United States delivered an ultimatum to Spain demanding an armistice and Cuban independence, and when Spain rejected it, war ensued. The United States first defeated a stronger Spanish fleet off Manila and took the Philippines and Guam, Spain's main colony in the Pacific. Then it defeated Spanish colonial forces in Cuba and Puerto Rico. As a result of the war, Cuba gained independence in 1902, after a transitional U.S. military government, while the United States annexed the Philippines, Guam and Puerto Rico (1899), as well as the uninhabited atoll of Wake Islands in the vicinity of Guam.⁷²⁸

In the aftermath of the Venezuela and Cuba conflicts, a few more crises occurred between the United States and the extra-systemic great powers at the onset of the 20th century, but the tide was already changing and the other great powers were much less willing to stand up to the United States than in the 1880s and 1890s. In 1902, Germany provoked the United States by staging a show of force against Venezuela in the form of naval maneuvers off the Venezuelan coast to force the Latin American country to repay debts borrowed from Germany. Britain, which also had a number of outstanding loans, joined the German display, but things went awry and some Venezuelan ships were sunk, prompting the Venezuelan government to request American assistance. The British accepted American arbitration but not the Germans, at least not until the United States threatened to conduct a large naval exercise of its own off Venezuela, prompting the Germans to back down. Another dispute flared up between Canada and the United States over a contested thirty-mile stretch of Alaskan coast which had become highly valuable after gold was found there. Canada asked Britain for its support and an arbitration commission made up of three Americans, two Canadians, and one Briton was set up in 1903. President Roosevelt nominated three biased Americans, so that the final say was up to the British

⁷²⁸ Dana Gardner Munro, *The United States and the Caribbean Area* (Boston, MA: World Peace Foundation, 1934), 3-6; Zakaria 152-4, 156-9; Grenville and Young 181-200; May 69-75, 83-4, 163-5, 177.

representative, Lord Alverstone. After Secretary of State Hay hinted that the United States might go to war over the issue, Alverstone sided with the Americans, depriving Canada of most of the disputed territory. The only other incidents with outside powers involved Denmark, which rejected a 1902 deal selling the Danish West Indies to the United States (President Roosevelt was later able to purchase the islands in 1917 for \$25 million), and an American display of power with Japan 1907, when President Roosevelt sent the whole U.S. fleet of twenty-two battleship on a visit to Japan to obtain Japanese cessation of immigration to the United States.⁷²⁹

With outside powers gradually stepping out of the picture, the United States accelerated its hegemonic rise in the early 1900s by increasingly intervening in Caribbean and then Central American domestic affairs and multiplying its commercial interactions with South America. While granting independence to Cuba in 1901, the U.S. simultaneously obtained a number of naval stations on the island and passed the famous Platt Amendment, granting itself the right to intervene in Cuba's domestic politics should its government fail to protect "life, property, and individual liberty," thus turning Cuba into a virtual protectorate of the United States. In the years following Cuba's independence, a number of incidents occurred during which the United States directly intervened to restore order. In 1906, after widespread electoral fraud led to revolts, the United States established a provisional government that operated for three years, pacifying the country and setting up new elections. The United States also imposed its settlement in at least five other domestic disputes and revolts between 1911 and 1917.⁷³⁰

The United States' most strategic intervention—in Panama—resulted in securing an American-controlled isthmian canal joining the Atlantic and Pacific coasts of the American

⁷²⁹ Zakaria 168-9, 173.

⁷³⁰ U.S. Dept. of State, *Papers Relating to the Foreign Relations of the United States – Transmitted to Congress Dec. 2, 1902* (Washington, D.C.: Government Printing Office, 1903); McCormick to Hay, U.S. Delegation in Vienna, March 7, 1902), 320-1; Munro 11-12, 17, 21-4.

continent, a long-time pet project of various U.S. governments primarily for purposes of national defense since without it all U.S. ships were required to make a long trip around South America to reach the other side. Private companies, both American and French, had unsuccessfully tried to build a canal, the former through Nicaragua and the latter through Panama, and had eventually abandoned the project due to the heavy financial burden. In 1899 Congress appointed a commission to investigate the feasibility of the project for the U.S. government and in 1902 settled on Panama, agreeing to buy the rights and excavations already begun by the French company. In January 1903 the United States signed the Hay-Herran Treaty with Columbia, of which Panama was a province at the time, selling the United States the right to build and administer a canal. When at the last minute the Columbian Congress refused to ratify the treaty despite popular support for it in Panama, a revolt broke out in Panama that soon turned into a revolution seeking the independence of the province. U.S. forces, sent to defend the American-operated Panama Railroad, prevented Columbian troops from landing in Panama and suppressing the insurrection, essentially helping Panama proclaim its independence in November 1903. The new state obtained U.S. recognition, and immediately signed a treaty granting the United States permanent rights over the future neutral canal and a 10-mile zone around it in exchange for U.S. protection of Panama's independence. Despite some initial disagreement between Panama and the United States over sovereignty and tariff collection, the Canal opened in 1914 under very favorable terms for the United States, and just as Cuba, the United States also continued to intervene in Panama's domestic politics, helping to broker the boundary with Columbia in 1909 and with Costa Rica in 1914, stepping in to maintain order when plots to overthrow the Panama government were unearthed or riots erupted, and supervising elections in 1906, 1912, and 1918

to prevent fraud. By 1913, Panama had also become a de facto protectorate of the United States.⁷³¹

Virtually the same outcome occurred with the Dominican Republic, Haiti, and much of Central America. The Dominican Republic, independent in 1844, suffered from constant civil strife and a massive public debt accumulated by a brutal dictator with foreign lenders, mostly Dutch, Belgian, French, British, and American, until his death in 1899. After the Dominican Republic defaulted on its loans and refinancing efforts all failed, European governments threatened to intervene to recover their funds. The United States preempted their actions by taking temporary control of the Dominican custom houses and its duties in 1904, the main source of revenue of the Dominican government. In doing so, President Roosevelt issued his Roosevelt Corollary to the Monroe Doctrine, granting the United States "international police power" in the region against not only the incursion of Europeans but also against the excesses and instability of local countries. A 1907 treaty ratified the temporary arrangement, and the United States became responsible for ensuring repayment of the Dominican debt. In the following years the United States found itself repeatedly intervening in Dominican domestic politics to secure stable governments and forestall revolutions, at times sending in the marines—in 1911, 1913, and 1914. From 1916 to 1924 the U.S. military even formally occupied and ruled the Dominican Republic after a coup. Much of the same happened in Haiti. After gaining independence in 1804, Haiti became inflamed with a vicious cycle of revolutions, upheavals, and spiraling foreign debt, primarily with French, German, and American bankers, and was at high risk of default by the early 1900s. When the German government sought to seize coaling stations in Haiti as a means of compensation in 1914 and a new revolution toppled the Haitian government, the Wilson administration occupied Port-au-Prince and seized the Haitian customs. After a 10-year U.S.

⁷³¹ Grenville and Young 309-310; Munro 63-8; Zakaria 167.

control of Haitian finances, U.S. forces withdrew, but the United States found itself intervening again afterward to quell revolts.⁷³²

The five Central American states (Guatemala, Costa Rica, El Salvador, Honduras and Nicaragua) all suffered from similar internal political disorder, finally, and the frequent revolutions and injuries to foreigners that threatened to trigger intervention by the European great powers. In addition, Lake Nicaragua and the San Juan River valley "offered the only practicable route other than that at Panama for the construction of a canal from the Caribbean to the Pacific," and U.S. diplomats spent considerable efforts trying to prevent other European powers like Britain and France from gaining control over a territory that would carry the canal. Even after the Panama Canal was constructed, the possibility of a second canal via Nicaragua kept the United States interested in preserving the region's stability whenever necessary, for example in 1906-1907 when Guatemala and El Salvador, and then Honduras and Nicaragua, went to war. In 1910 the United States also negotiated that American bankers take charge of Honduras's and Nicaragua's defaulted foreign debts, and intervened several times, both diplomatically and militarily, to end a series of revolutions in Nicaragua between 1910 and 1914. In 1914 the United States signed the Bryan-Chamarro Treaty with Nicaragua, obtaining the right to construct and operate a canal across Nicaragua, as well as a 99-year lease over the Great and Little Corn Islands in the Caribbean and a 99-year lease to establish a naval base in the Gulf of Fonseca, and it did not hesitate to pressure Central American states to conduct internal reforms to throughout the period.⁷³³

By the onset of World War I the United States had thus gradually expelled its rival great power from the region and extended its will over the Caribbean and Central America, both in

⁷³² Munro 101-103, 110-1, 116-9, 149-56, 163, 166-7; Zakaria 170.

⁷³³ Munro 200-201, 229-31, 235, 242.

foreign and domestic politics. It had also established naval supremacy and acquired a formidable Navy and power projection capabilities that extended its hegemony over the entire Hemisphere. Why it succeeded so fast in overturning the balance of power had much to do with both its potential balancers' failings and its own unusual traits.

1. The Balancers' Side: Absence of Collective Action and Bandwagoning

The United States' potential balancers belonged to two categories: extra-systemic, mostly European, great powers possessing and/or seeking footholds on the American continents; and local secondary powers, mostly independent but too weak politically and militarily to individually alter the distribution of power on the continent. All potential balancers failed to stop or even slow down the United States' rise, however, because of their inability to cooperate. While the local powers were hindered by lack of trust, the great powers passed the buck amongst themselves and the distance between their home base and the American continent encouraged them to remain idle. In addition, bandwagoning both by local powers and by the most prominent great power, served to reinforce the United States' growing monopoly over the continent. Communication might also have played a role, though perhaps not a major one.

a. Communication Problems (IV1)

Communication problems hindered the potential balancers during the United States' rise, but they played out differently than in the pre-modern cases. Although physical communication barriers were predictably absent (IV1.1), the potential balancers' misperception of America's rise

was unusually prominent (IV1.2). There were very few deceptive attempts by the rising hegemon to conceal its rise or hamper the potential balancers' collaboration (IV1.3).

1. Physical Communication Problems (IV1.1)

Since primitive communication means did not dramatically hamper potential balancers' awareness of a rising hegemon even in most ancient cases, one can only assume that in a modern case, with improved communications technology and significantly shortened information travel times, physical issues did not hinder the potential balancers' knowledge of America's advances. Even though the great powers in this case were based across the Atlantic on a different continent, all had some presence on American soil, either in the form of a territorial possession or protectorate and/or in the form of diplomatic missions. Indeed, all European great powers had a permanent representative in Washington, as well as in other local countries, so that no great power could claim to ignore developments across the Atlantic. While the bulk of transatlantic messages was still transported by boat and thus took several weeks, urgent communications could be delivered instantly via the Atlantic telegraph, which was operational since 1858. In addition, European great powers were themselves in constant diplomatic contact and quickly passed news of transatlantic developments amongst each other. For instance, when at the onset of the 1898 Spanish-American War, the German consul in Hong Kong got word that American ships were being readied for a raid on Manilla, he passed the information along to the Kaiser at once, who immediately relayed it to the Spanish Queen Regent.⁷³⁴

There is clear evidence that America's rival great powers were well aware of the soaring U.S. economic power. "Certainly it cannot be said that Europeans were ignorant of America," May stresses. May emphasizes that in the late 19th century hundred of books and comparative

⁷³⁴ Grenville and Young 282-3.

studies were published in Europe on the United States, and a large number focused on the American economy, aiming to inform European businessmen about the competition. They stressed the business-friendly political climate and practical-oriented education system conducive to growth. European governments were conscious of America's economic power since several installed tariffs in the 1880s and 1890s to protect their own agricultural production. As early as 1881, a German pamphlet prophesized that Europe could soon "succumb to the United States in the battle for supremacy over world trade." Many publications also debated the superiority of American republicanism and liberalism over European monarchies. In addition, Europeans also knew about the United States' growing military might, particularly following the Spanish-American War, when except for Spain the great powers could still have acted to stop the spread of American influence. Writings in 1899 spoke of an "American peril." The French paper *Correspondant* even claimed in March 1899 that "the threat from America was as real and immediate as that from Germany." The Italian *Corriere della Sera* exclaimed that if the United States went after Italy's colonies, the Italians might well lose them. Many European pundits wrote about America's imperialist rise. In Germany, Otto von Moltke and others wrote books warning that the United States sought to take over the world. Publications also increasingly focused on comparative studies of European and American military capabilities, particularly naval power.⁷³⁵

Most of the United States' continental neighbors were similarly aware of the U.S. rise, also informed by their diplomatic and commercial representatives and by the growing number of Americans traveling to their lands on business or leisure, although a few of the most remote areas of the Western Hemisphere might have had much less contact with the outside world. While most coastal areas in Central and Latin America were in constant contact through trade

⁷³⁵ May 181-2, 263-4.

and diplomacy with the rest of the continent and thus aware of American growth, the Latin American hinterlands may have been more isolated, but since they mostly were not the center of political and economic activity, their possible lack of up-to-date information cannot be considered crucial toward affecting Latin American states' responses to the U.S. rise. The five small republics of Central America, for example, were particularly isolated because they were accessible and populated mostly on the Pacific side and cut off from the Caribbean Sea by high mountains and impenetrable jungles. As Munro notes, "they were thus to a great extent cut off from communication with the outside world until the construction of railroads to San Jose, Costa Rica, in 1891, and to Guatemala City in 1908, and the opening of the Panama Canal brought their more important towns within easy reach of the United States and Europe." However, while isolation certainly contributed to their slower economic and political development, it is the fact that they remained secondary or even weak powers rather than their isolation that prevented them from acting against U.S. expansion.⁷³⁶ Overall, physical barriers cannot be said to have played any role in limiting balancing against the United States.

2. Misperception (IV1.2)

Misperception stands out as a much more potent explanation. While most potential balancers possessed the correct information about the rise of the United States, their chronic underestimation of its expansionist intentions was partly to blame for their lack of action. As Mearsheimer underscores, "American foreign policy throughout the [late] nineteenth century had one overarching goal: achieving hegemony in the Western Hemisphere." The U.S. goal was to develop sufficient power to "dominate the other independent states of North and South America and also prevent the European great powers from projecting their military might across the

⁷³⁶ Munro 195.

Atlantic Ocean."⁷³⁷ Yet, despite being fully aware of the United States' tremendous growth in power, most potential balancers dismissed its intentions as benign, rather than expansionist. As a result, they did not take the U.S. rise seriously, especially in its early stages when they could have easily stopped it. Until the late 1880s "the United States was dealt with as a second-rate power." Foreign countries would send only minor diplomats as their U.S. envoys, for example. Even during the 1890s, "they [i.e., foreign statesmen] were never oblivious to its existence. But neither did they conceive of it, any more than of the Netherlands or Sweden, as a nation that might become a player of power in its own right." Only by the early 20th century, once the United States was already very powerful, did most other great powers realize its real potential. "By the early 20th century, some European statesmen looked upon America as a very formidable power indeed." But still not all powers did and by that point the United States was already quasi-hegemonic over the continent.⁷³⁸

British misperception of the United States was particularly evident, and a serious factor since the United Kingdom was the single most powerful potential balancer. Britain's provocative actions during the Venezuela crisis of 1895-6 show that its government did not take U.S. intentions seriously. The Britons not only denied the American claim that the Monroe Doctrine applied to Venezuela and rejected the U.S. demand that their dispute with Venezuela be submitted to international arbitration, but in a blatant dismissal of the Monroe Doctrine, they simultaneously offered France the Caribbean island of Dominica in exchange for France's fishing rights in Newfoundland and some African territory. Although the French refused, "the whole episode displayed ... how little attention the British government paid to the United States before the 1895 crisis." May even calls British PM Salisbury and Colonial Secretary Joseph

⁷³⁷ Mearsheimer 236.

⁷³⁸ May 3, 5.

Chamberlain "insouciant" in the face of the growing power and presence of the United States. Although President Cleveland used strong language to warn the Britons to relent in their claims on the Venezuelan boundary dispute in 1895, Chamberlain and Salisbury dismissed it. As May explains, the British government was "under the false impression that [the U.S.] government did not mean what it said." The reason was that "Salisbury, Chamberlain, and the majority of Englishmen had simply not been thinking of the United States as a potential factor in the balance of power." Chamberlain's biographer remarks that the Britons reacted with "stupefaction" when Cleveland eventually sent them his quasi-ultimatum on Venezuela, and for some time even believed that the United States would "recede from its extreme position," a clear indicator of misperception. But by the time they realized their mistake, the United States had already gained too much power and influence over the continent for Britain to handle on its own. At that point, for example, the War Office recognized that if the United States were to attempt to take over Canada, the overstretched British would find it "difficult if not impossible" to defend.⁷³⁹

To some extent, the European great powers' underestimation of U.S. intentions, though dangerous, was not entirely unjustified. Even though the United States possessed considerable power at the time of the Venezuelan crisis, for example, Grenville and Young argue that "the United States was militarily completely unprepared for war" and was therefore bluffing in its tough stance toward Britain, and "if the British cabinet had been aware [of] that ... their actions might have been different." For instance, the U.S. South Atlantic Squadron was at the time wintering far south in Montevideo, Uruguay, leaving "the Atlantic coastline and the Great Lakes exposed to attack" if Britain had suddenly sent an expeditionary force. Chamberlain even seemed to recognize the waning, though still considerable, mismatch between America's economic and military power, when he wrote to Salisbury in February 1896 that the Americans "are great

⁷³⁹ May 45-8, 50.

people for bluffing."⁷⁴⁰ Nevertheless, had the great powers not chronically underestimated the United States' expansionist plans during its early rise, they would have avoided the surprise of America's tough stance during crises and anticipated its numerous fait-accomplis throughout the region.

3. Deliberate Deception (IV1.3)

It does not appear that the United States tried to conceal its rise or trick the great powers to prevent them from balancing. However, the United States did not hesitate to use underhanded techniques as well as intimidation to force local secondary powers to behave in America's interest. Its sometimes coercive promotion of Pan-Americanism is a blatant example.

Pan-Americanism was a century-old movement reinvigorated by the United States in the 1880s with the apparent goal of advancing the common interests of the continent but in reality it can also be seen as a clever tool engineered by the United States to advance its own interests under the cover of a greater, common good. The United States called for the first Pan-American conference in 1881 under President Garfield, but it never materialized. President Cleveland again sent invitations to Central and South American governments and the first joint conference was held in November 1889 in Washington, D.C. However, the United States made certain that American Secretary of State James Blaine was elected chairman in order to keep control over the agenda, and displayed American power by taking the foreign delegations on a 6000-mile long luxury train trip that stopped at dozens of industrial sites, to establish the United States as the unequivocal leader of the movement and of the American continent. For American leaders, Zakaria contends, "Pan-Americanism was a means through which the United States would dominate its southern neighbors 'in the spirit of the Monroe Doctrine.'" When the foreign

⁷⁴⁰ Grenville and Young 171-4.

representatives became wary at American proposals for a common arbitration policy and a customs union, Congress passed a tariff bill in 1890 giving the president the power to stop the imports of certain goods from a foreign country if it did not conform to equal and reasonable trade norms, and President Harrison "used this power to force all Latin American countries except Colombia, Haiti, and Venezuela [who were subsequently blacklisted] to sign reciprocity treaties."⁷⁴¹

America's hijacking of the Pan-American movement and bullying of Latin American countries thus contributed to securing support and squashing resistance from local secondary powers. But such behavior was generally not widespread and can therefore not be considered a crucial factor in hampering balancing, particularly since it affected mainly the weaker, local powers. The great powers' tendency to misperceive American intentions, particularly England's, may have had a more direct impact on the failure of balancing. However, an even greater problem seems to have been at play in preventing balancing efforts.

b. Collective Inaction (IV2)

The European great powers could have achieved the best results against the United States by acting in concert. Because of the distance and consequent need for strong power projection to operate in the Western Hemisphere, they probably would have had a difficult time stopping even a slightly weaker United States individually, but together, given that they possessed footholds and influence in all different areas of the Hemisphere, they would likely have been very effective in providing an alternative to the United States and preventing its takeover of many of the local powers. Surprisingly, however, unlike in previous cases, there were virtually no cooperative

⁷⁴¹ Lloyd C. Gardner, Walter F. LaFeber, and Thomas J. McCormick, *Creation of the American Empire: U.S. Diplomatic History* (Chicago, IL: Rand McNally & Company, 1973), 233-4; Zakaria 138-9; Spykman 360.

efforts undertaken against the United States, whether by the great powers, the local secondary powers, or a combination thereof. Collective efforts were not laggard (IV3), but rather plainly nonexistent (IV2). While potential balancers, particularly the great powers, made a number of promises of support and cooperation, they never took an actual step toward common intervention. The few countries that attempted to stand up to the United States—England, Spain, Germany, and a few secondary powers—did so on their own.

A transcontinental cooperative effort, with European great powers backing up local resistance against the United States, would have had particularly great potential. "The real danger that the United States faced in the nineteenth century ... was the possibility of an anti-American pact between a European great power and a state in the Western Hemisphere," as happened when France installed Maximilian on the Mexican throne during the Civil War, because such an alliance could have "ultimately be[en] powerful enough to challenge U.S. hegemony in the Americas." However, the only such cooperative effort during the United States' rise was Germany's attempt to train a number of armies from independent Latin American states after its victory in the Franco-Prussian War in 1871 by sending German officers there to act as instructors. For example, Germany trained the Bolivian and Chilean armies and provided them with weapons. The cooperation was not aimed at the United States, however, and in reality mostly served to fuel the conflict between those states and their neighbors. Both became subsequently entangled in the War of the Pacific, pitting Chile against Bolivia and Peru from 1879 to 1884 (which Chile won).⁷⁴²

Attempts at setting up a union of secondary powers went further than usual in Central America, but again were not relevant for our purposes because they were not directed against the United States and did not consist of balancing. Additionally, such attempts all eventually failed.

⁷⁴² Mearsheimer 249.

After Central America obtained independence from Spain, it formed the Federal Republic of Central America in 1821, but that union soon fell apart in a bitter civil war and was replaced by independent states. There were recurrent efforts to bring back the union throughout the late 19th and early 20th centuries, but none succeeded because of each country's strong nationalist movements. The region was ripe with revolutions, and each country participated in its neighbor's instability by propping up its revolutionaries. The last attempt to reestablish a Central American federation in 1921 failed because various groups plotted coups not only in their own country but also against neighbors, straining their relations even further. Finally, the Pan-American Union could also be considered an instance of Latin American cooperation, but not only was it not directed against the United States, it was even led and manipulated by the United States, as mentioned above. Its general secretariat was located in Washington, D.C., and even if it had aimed at opposing or slowing down the United States, it would have been unable to do so since it was set up as a purely administrative agency and lacked any means of common action and enforcement. As Nicholas Spykman points out, "the Pan American system contains no guarantees of territorial security and political independence and makes no provision for coercion by the Pan American community." In fact, its achievement remained modest and pertained mostly to issues like transportation, communication, sanitation, education, and various cultural exchanges.⁷⁴³

Thus, the absence of common balancing efforts is particularly intriguing in this case. Why were America's potential balancers so overwhelmed by collective inaction? While the communication problems mentioned above—essentially misperception—prevented individual great powers from taking the United States seriously, it is unclear whether these issues also played an indirect role in preventing alliances (IV2.1). One can safely assume, though, that if

⁷⁴³ Spykman 214, 353-4; Munro 195-6, 207-10.

individual great powers hesitated to face off with the United States because they underestimated America's expansionist intentions, that same misperception only reinforced their reluctance to intervene collectively. In reality, the cause of the potential balancers' inaction was a combination of lack of trust (IV2.2), lack of sufficient interest (IV2.3), and buckpassing (IV2.4) on the part of the European powers, and rampant lack of trust and immediate gains focus (IV2.2) on the part of the local powers.

1. Lack of Trust (IV2.2)

Both the great and secondary powers were plagued with profound distrust that pitted them against each other and rendered collaboration unattainable. First, the great powers were increasingly preoccupied with disputes and competition amongst each other during the latter part of the 19th century, which gradually degenerated into open war in 1914. Some of the disputes were severe even before World War I, with a number escalating into military face-offs (the Franco-Prussian War of 1870-1, for example) and quasi-wars (such as the Anglo-Russian crisis of 1877-8 over Gallipoli and the Turkish Straits), generating distrust so profound that the European powers quartered themselves in a network of shifting but increasingly rigid alliances meant to protect themselves from each other. The great powers' involvement in a different, parallel regional balance of power system in Europe with its own aspiring hegemons, hardened their distrust and made it difficult to eschew local conflicts to cooperate against the United States. The suspicions and rivalries arising from the European great power competition naturally played in the favor of the United States. Spykman points out that "preoccupation of the European nations with the balance of power at home gave us [i.e., the United States] the opportunity to grow to our present position of power" while the great powers focused on outpacing each other, especially as European belligerence grew at the turn of the century. Those internal great power

rivalries directly explain why the United States' most serious competitors never seriously challenged the Monroe Doctrine, even though they could have, "with the war potential of most of the Eurasian continent at [their] disposal."⁷⁴⁴

The European great powers' mutual suspicion toward each other was perhaps most evident in their quests to secure colonies in Africa and Asia. They viewed the race for colonies as a zero-sum game, where one great power's gain was another's loss. This short-term gains reasoning rendered their relations particularly antagonistic. The European great powers were involved in too many colonial flare-ups with each other to list here, particularly in the 1880s and 1890s at the very time U.S. growth was soaring. To name a few, the French, Germans, Russians, Japanese and Britons quarreled over establishing their respective protectorates in coastal China, as mentioned earlier; the French occupied Tunisia to circumvent the Italians; the Britons invaded Egypt to squeeze out the French; Germany and Britain clashed over East Africa in the mid-1880s, resulting in a German protectorate over Tanganyika and British protectorate over Kenya; Britain and Russia collided over the control of Afghanistan; Britain took over Baluchistan, Burma and Tibet to counter the ambitions of France in Indochina and of Russia in Afghanistan; the Britons squabbled over some spoils in the Pacific, sharing New Guinea with the Germans and the Dutch and the New Hebrides with the French; and the fiercest competition probably occurred between France and Britain over the control of central Africa, with each country closing in from a different direction on the Sudanese outpost of Fashoda in 1898 and barely avoiding full-out war.⁷⁴⁵

While European divisions were already present in the 1880s and 1890s, distrust only grew more bitter at the turn of the century, and "the growing rivalries of Europe were to aid the

⁷⁴⁴ Spykman 71-3; 89.

⁷⁴⁵ Serge Berstein and Pierre Milza, "L'Impérialisme Colonial," in Serge Berstein and Pierre Milza, eds., *Histoire du XIX^e Siècle* (Paris: Hatier, 1996), 500-504.

United States in pursuing [its] policy of dominance in the Caribbean" in the early 1900s.⁷⁴⁶ French-British relations remained particularly strained after Fashoda. For a while, the French considered getting over their fundamental disagreements with Germany stemming from their loss in the Franco-Prussian War to ally with the Germans against Britain.⁷⁴⁷ But even when France and Britain eventually signed an Entente in 1904 because of the imminent danger arising from Germany, which was engaging in an arms race with the European continent, it was far from a natural cooperation. In fact, "before the French and the British signed the Entente..., the two nations indulged in glorious reciprocal hatred and disdain clearly evident in the acid tone of the articles which the press and the periodicals of the two countries published about each other." The Entente was not a fundamental change of heart, just "the inevitable result of having to work together against a common enemy." Spykman concludes that "sympathy does not determine policy; policy tends to determine sympathy." As a number of conflicts emerged between the two despite having signed the Entente, it is clear that the relationship was not one of trust.⁷⁴⁸

But France and England were not the only great powers wary of each other at the turn of the century, rendering a common action against the United States all the more unachievable. In the Pacific, the Japanese were worried about American expansion in their own East Asian zone particularly after the annexation of the Philippines and Hawaii in 1898 and its interest in China. However, Japan's ongoing competition over China with Germany, Russia, and England, all of whom had seized control of their own Chinese ports in 1897-8, made a collaborative effort against the United States impossible. In Europe, even though the Germans had carefully studied the question of sending an expeditionary force across the ocean to attack the United States between 1899 and 1903 and had put together a plan based on an initial landing in Puerto Rico,

⁷⁴⁶ Grenville and Young 301.

⁷⁴⁷ May 233.

⁷⁴⁸ Spykman 256.

they knew that the idea was flawed. To succeed against the United States required sending "so much of Germany's naval strength" to the Western Hemisphere that it would "leave the homeland unprotected," and given the danger of another crisis in Europe the risk was too high.⁷⁴⁹ Yet the Germans did not hesitate in provoking the ire of their fellow potential balancers. When Britain faced an embarrassing defeat by the Boers in South Africa and was forced to retreat in 1896, the Kaiser sent a message of congratulations to the Boer president acknowledging the independence of the Boer state, which the British government did not recognize. This prompted great hostility in Britain toward Germany, to such an extent that "German shopkeepers and tourists became ... fearful for their lives."⁷⁵⁰

Britain also remained at odds with Russia. In 1902 the British government concluded a formal alliance with Japan, for instance, whose purpose was to counteract the growth of Russia in Asia. One element further reinforcing the European great powers' distrust of each other was their polar opposite forms of government. Absolute monarchs like the German Kaiser, for example, considered the French Republican government illegitimate and was also hesitant to trust Tsar Nicholas II, who did not share his skepticism and sometimes cooperated with the French. For the Kaiser, kings governed by divine right and were the only legitimate rulers; a king could thus not cooperate with non-monarchic governments, which inevitably sought the downfall of monarchy and could not be trusted. On the eve of World War I, "the rivalry of the European powers [essentially] provided the North American mainland with a considerable degree of protection."⁷⁵¹

⁷⁴⁹ David Healy, *U.S. Expansionism: The Imperialist Urge in the 1890s* (Madison, Wisc.: The University of Wisconsin Press, 1970), 113.

⁷⁵⁰ May 49-50.

⁷⁵¹ Grenville and Young 304-8, 319; May 196, 237.

While the great powers were engaged in conflicts of their own that placed considerable hurdles on acting together to stop the U.S. rise across the Atlantic, the secondary powers of the American continent were similarly hampered by their own disputes and short-term gains quests that similarly froze their abilities to coalesce against the larger threat of the United States. South America in the late 19th century was rife with power struggles and competition over territory. South America had three recurrent areas of conflict, in addition to countless border conflicts: the west coast, with the Chile-Peru rivalry and the problem of Bolivia's access to the sea; the Western Amazon and its century-old quarrel between Ecuador and Peru; and the Southern plains disputed by Argentina and Brazil.

The Chile-Peru issue was particularly deep-seated and created a "heritage of bitterness" between the two countries. In 1879 Chile conquered Peru's nitrate-rich province of Tarapaca and Bolivia's coastal province of Atacama, its only sea access, in the War of the Pacific. But in 1883 Peru not only lost the Tarapaca province entirely, but also had to relent to a 10-year Chilean occupation of the Tacna and Arica districts, whose future was to be decided by a popular referendum after the occupation period. The referendum never took place and "the issue remained unsettled between the two nations," creating constant friction. The Peru-Ecuador issue revolved around a dispute zone located in the upper Amazon Basin, a little-populated, rubber-producing jungle zone of Amazon River tributaries where Brazil, Venezuela, Columbia, Ecuador, Peru, and Bolivia meet. A number of conflicts arose between Brazil, who tried to expand into the area, and the other states, and between Peru, who also sought to expand eastward into the region and had the best access via the Andes, and Columbia and particularly Ecuador, with all three claiming ownership of the region. It remained a very instable region throughout the period as the borders between Peru and Columbia, and between Peru and Ecuador, were "by no

means ... settled boundary." Finally, the third area of conflict revolved around the drainage basin of the La Plata River, over which the two most powerful states of Latin America, Brazil and Argentina, had been fighting. In colonial times, before Argentina's independence in 1816, the Viceroyalty of La Plata also covered parts of Bolivia, Paraguay, and Uruguay. Argentina had ambitions to expand into all three, and did so "in the form of economic penetration," while Brazil tried to oppose growing Argentine influence in the border regions. Argentina and Brazil eventually fought two wars with Uruguay and all three participated in a bloody war with Paraguay between 1865 and 1870 that decimated half of Paraguay's male population. An insurrection in Brazil in 1893 resurrected the issue once more.⁷⁵²

Further north in Central America, the story was much the same. Since the breakup of the Central American federation, the five states of the area had been frequently in quarrel over their borders. In 1894 disputes between Honduras and Nicaragua became more serious and turned into "open hostilities," while Nicaragua simultaneously tried to take over the autonomous Mosquito Indians. At the same time the Venezuela-British Guiana border dispute was flaring up. The intervention of the United States, both in the Venezuela issue and in the Mosquito region, did not quell the mistrust, and a series of wars eventually tore up Central America after the turn of the century, essentially rendering any kind of cooperation against U.S. intrusion unfeasible. In 1906 Guatemala and El Salvador went to war after the Salvadoran Minister of War supported a revolution against the Guatemalan president. Just months after the two countries ended their conflict, another arose between Honduras and Nicaragua because of the Nicaraguan president's support of Honduran revolutionaries, and "a general war involving all four countries seem[ed] inevitable" in 1907.⁷⁵³ It was barely avoided, but Nicaragua did not stop its meddling and openly

⁷⁵² Spykman 344-8.

⁷⁵³ Munro 199-202.

propped revolutionary movements in El Salvador. Simultaneously, the governments of El Salvador and Guatemala helped spawn a revolution in Honduras in 1908. Costa Rica and El Salvador also denounced Nicaragua for entering the Bryan-Chamorro Treaty in 1916, promising the United States an option to construct a second isthmian canal through Nicaragua and to establish a naval base in the Gulf of Fonseca, not because of the U.S. threat, but rather because they both claimed a right over the promised territory. The rivalries between the Central American countries ran so deep that they took precedence over virtually any other priority, and the United States intervened easily to secure its own interest: in 1910 it sent marines to Honduras after Nicaragua, with the likely help of Guatemala, invaded parts of Honduras, and then arranged for American bankers to take charge of Honduras's and Nicaragua's spiraling foreign debts.⁷⁵⁴

While the cause of mistrust among the great powers lay in their squabbles over colonies and the European balance of power, and perhaps reinforced by their incompatible forms of government, the roots of the secondary powers' mistrust of each other were somewhat different—their internal disarray and limited prospects led them to settle on extracting easy, immediate gains at the expense of their neighbors. For those Latin American countries that managed to achieve independence from Spain in the 19th early century, problems were not over and domestic confusion reigned, rendering alliances highly unlikely. "Independence from Spain did not end dictatorship and absolutism," and even though many republics copied American political structures, "geographic and social realities soon reasserted themselves over paper constitutions, and ... there emerged a political pattern that has been characteristic of Latin America ever since: government by the Caudillo," the reckless military dictator. 19th century Latin America was in many respects still a feudal, colonial system with relatively primitive economic and administrative structures concentrating wealth in the hands of a few landowners,

⁷⁵⁴ Grenville and Young 116-7.

engendering tremendous socio-economic disparities, constant class conflict and domestic chaos. Although the region was under growing influence of the West, most of the industrialization and development of modern capitalist economies did not occur until the First World War. The lack of social cohesion generated constant "stresses and strains." Those centripetal forces were further reinforced by the ethnic diversity within each republic, which increased the difficulty of generating unity and compatible political goals both within and across borders. In other words, Latin America was so rife with its own, multi-layered divisions and insecurities that it was in no position to provide a strong, united barrier to incursions from the United States.⁷⁵⁵

2. Lack of Sufficient Interest (IV2.3)

While lack of trust and endemic in-fighting was the primary cause of the secondary powers' collective inaction, another factor also played a key role in preventing the great powers' collaboration: they did not care enough. Lack of sufficient interest is somewhat to be expected given that all great powers in this case were external to the Western Hemisphere, and were based in a region that had its own balance of power struggles in the late 19th and early 20th centuries with the rise of Germany. Tending to the European balance naturally became a priority for the European great powers, and took preeminence over dealing with the rising United States. European states considered the possibility of a European hegemon next door a significantly more imminent threat than the existence of an American hegemon across the Atlantic, even though according to balance of power theory a regional American hegemon would soon become dangerous as a potential global hegemon and must thus be balanced as seriously as a potential European hegemon. The European disinterest in the Cuban crisis in 1898 was evident, and coincided with a decisive acceleration in the European arms race. In June 1897, German Admiral

⁷⁵⁵ Spykman 89, 218, 222-7.

von Tirpitz argued in a groundbreaking memorandum that Germany must build formidable battleship-based navy "aimed against Great Britain," and in late March 1898, just as the Cuban crisis was escalating, the Reichstag passed the First Fleet Act setting Tirpitz's plan into motion.⁷⁵⁶ Having observed U.S. power unfold in Venezuela and being aware of the growing strength of the U.S. position, Spain requested the support of the other European great powers in its Cuban standoff in March and April 1898, but it did not obtain it. President McKinley at first hesitated to escalate further because of the possibility that France, Russia, and Germany might side with Spain in case of war—hinting that a common balancing approach might have still worked at that point in forcing the United States to back down. But when each European power stalled, showing little interest in supporting Spain, the United States did not waver in driving Spain to war.⁷⁵⁷

In addition to the presence of threats right in their backyard, the European great powers were also hesitant to intervene against the United States because the distance made it particularly costly and risky. Organizing an operation into the Western Hemisphere required far greater resources than intervening close to Europe, and once capabilities were assigned there, even more time and resources would be needed not only to resupply and reinforce them, but also to repatriate them if necessary. Capabilities dedicated to an American mission would thus not be usable elsewhere in the near future, and the European great powers were therefore unwilling to commit the large amounts of resources necessary to succeed against the United States at such a distance from their home base. Germany's unwillingness to set in motion its plan to send an expeditionary force to the United States via Puerto Rico is a case in point—in the early 1900s

⁷⁵⁶ Robert J. Art, "The Influence of Foreign Policy on Seapower: New Weapons and Weltpolitik in Wilhelminian Germany," in Robert J. Art and Kenneth N. Waltz, eds., *The Use of Force: International Politics and Foreign Policy, Second Ed.* (Lanham, MD: University Press of America, Inc., 1983), 177-178.

⁷⁵⁷ May 119; Zakaria 153, 176.

Germany could not sacrifice enough troops and ships away from the European theatre to mount a successful operation on the American continent. The crisis between the United States and Chile in 1891 is another pertinent illustration. Not only was Chile across the Atlantic, but it was one of the furthest points from Europe, on the Pacific side of the continent. Chile eventually had to back down from its escalation with the United States because it "had received messages from Europe that ... the great powers of Europe would not intercede on Chile's behalf in the event of war" even though they were likely sympathetic to its cause.⁷⁵⁸

Finally, a third factor in Europe's lack of interest in stopping the United States was the great powers' preoccupation with other regions, particularly Asia and Africa, where they were also competing for influence. For the great powers, the American continents were one possible reservoir of influence over their peers, but there were several others. Most great powers, especially Britain, France, and Germany, were expanding in all corners of the world in the late 19th century, and the Western Hemisphere was thus not the only priority outside Europe. Sometimes they managed to form coalitions in other regions, though those peripheral conflicts distracted them from developments on the American continent. In the mid-1890s in Asia, for example, British interests in the Far East were threatened by combined German and Russian expansion. In early 1895, France, Russia, and Germany began cooperating diplomatically to curtail another East Asian competitor's expansion, Japan. At the same time, the British, French, Germans, Italians, and Belgians were battling for the division of Africa, which monopolized vast amounts of resources and may have seemed a more urgent priority than the rise of the United States. Indeed, colonial powers faced recurrent difficulties in multiple theatres from the 1880s all the way through the 1910s. Germany faced resistance in southwestern Africa and in the Horn region, while Britain was confronted with revolts in its colonies in Burma, Siam, India,

⁷⁵⁸ Healy 113; Zakaria 141.

Afghanistan, Turkey, Nigeria, and West and South Africa, and France was battling insurgents in Indochina, North Africa, the Gold Coast, and central and equatorial Africa.⁷⁵⁹

While the large countries of Europe that sought influence in America were distracted by their race for colonies, the smaller European states often had a total disconnect with the American continent. The Austro-Hungarians, for example, had virtually no interest in the Spanish-American war and therefore felt no urge to intervene. As May points out, unlike the great powers on the continent, who silently supported Spain, the Austrians were entirely "indifferent" to Spain's cause. Although it was sympathetic to Germany's call for support to Spain by all the monarchies of Europe in the name of regal solidarity, Austria was busy on its eastern border with Hungarian separatism. Because it had very little economic interaction with the United States, and possessed no colonies across the Atlantic, it felt no interest in sharing in a potential common action. The Italian government similarly had no stake on the American continent and thus no urge to intervene. It sought to prevent the Spanish from becoming too close to the French, however, with whom Italy was at odds in Africa, so it decided to follow the other great powers' lead.⁷⁶⁰

3. Buckpassing (IV2.4)

In the end the European great powers, overwhelmed by their own mistrust and internecine conflicts and lacking interest in intervening, passed the buck to each other. None wanted to take the risk of bearing the price of action against the United States and possibly becoming the sucker if the others faltered. Buckpassing was clearest during the events leading up

⁷⁵⁹ See Ronald Robinson and John Gallagher, *Africa and the Victorians: The Climax of Imperialism in the Dark Continent* (New York: St. Martins Press, 1961); Grenville and Young 119; May 46.

⁷⁶⁰ U.S. Dept. of State, *Papers Relating to the Foreign Relations of the United States – Transmitted to Congress Dec. 2, 1902* (Washington, D.C.: Government Printing Office, 1903); McCormick to Hay, U.S. Delegation in Vienna, March 7, 1902), 27; May 201-3, 209-10.

to the Spanish-American War, because it was the most serious incident between a great power and the United States and Spain requested help from each of the great powers. Yet despite promises and calls of sympathy, they eventually bowed out one after the other, and their reticent behavior throughout the crisis demonstrates their expectation that another power would take charge.

During the escalation leading to the Spanish-American War, Spain repeatedly appealed to the other European great powers to form a united front against the United States, which it reasoned would cause the United States to yield. By July 1896, the Spanish government had already issued the Tetuan Memorandum, asking for help against the United States. In January 1897, German newspapers claimed that France and Russia promised to intervene, backed by strong anti-American sentiments. In addition, the Kaiser reportedly "called for European unity not only in support of Spain but also against the Monroe doctrine." In a speech in front of the Reichstag, a prominent conservative leader called for "united resistance" against U.S. economic expansion. Germany also officially encouraged emigration to Latin America. In 1897, Spanish high authorities stepped up their requests and spoke with the British, Austrian, French, and Russian ambassadors. They also played the religious card, seeking the Vatican's support to build a Catholic network against anti-clerical America. The results seemed initially promising. The European ambassadors announced that their countries were ready to help Spain, at the condition that it be through common European action. But it was already obvious in the conditional offer that no great power was individually willing to take charge but instead preferred to pass the buck to others to take the lead on a common action. When in December 1897 Russia abruptly rescinded its commitment, the Spanish efforts seemed derailed until the French managed to convince Russia to participate in a joint action again in late March 1898. Then, however, the

British government began having second thoughts, and as a result even French support became doubtful since the French "insisted on England's participation," a clear indication that France sought to pass to England the responsibility for making the decision. Finally, after the Spanish Queen Regent sent a personal message to Queen Victoria, the French government announced that Britain had relented and accepted again to participate. As May notes, "no sooner had the sun broken through, however, than clouds closed in again." The British ambassador to Spain declared that the government in London in fact considered "intervention at this stage premature." The Spanish ambassador in Berlin also reported that Germany would probably renounce its support too, although the Kaiser had been the first to express support for the Spanish monarch. Even the Pope suggested that Spain give up and sign an armistice on April 2.⁷⁶¹

Each great power found excuses to suggest someone else take responsibility. Although the German emperor impulsively supported Spain, members of his cabinet, including foreign minister von Bülow, did everything they could to stall his eagerness, conscious of the economic loss a fallout with the United States would represent since Germany was the United States' second trade partner, just behind Britain. Therefore, the German cabinet insisted that France and England must participate before they commit themselves. The French were also delaying commitment, torn domestically by the Dreyfus Affair, and engaged in typical buckpassing. A French Foreign Office memorandum noted in the fall of 1897 that "it is important that we not become committed prematurely," and suggested a wait-and-see approach. The French foreign minister decided to "temporize," telling the Austrian ambassador that "France would willingly agree if all others did." Russia should have been most worried about the rise of the United States and particularly about its expansion into Hawaii and the Philippines because of Russia's own

⁷⁶¹ Memoranda by Hanotaux, *Espagne: Cuba, III and IV* (Fr.), by Leon y Castillo and Gullon, *Legajo 2417* (Sp.), qt. in May 119, 164-5, 169-72, 176.

goals in the Far East. However, the Russians remained "untroubled," because they were counting on great power distrust and petty short term gain focus to generate antagonism between the United States and Britain, and between both Anglo-Saxon powers and Germany, which were also interested in the Far East.⁷⁶² In other words, Russia expected to pass the buck to Britain and Germany for taking care of the United States. Because Russia was also behind in terms of military capability, being for example surpassed by German and French progress in artillery, it also avoided to take a lead in a common action front with Spain.

The final test of collective European action to stop the United States in the Cuban crisis occurred just two weeks before the Spanish-American War started. Although in early April 1898 the European great powers had been fidgeting about acting together, they managed to pull one last show of solidarity. On April 6, the European ambassadors visited the White House to collectively express their support for peace. They had not been able to agree on any other move. Yet even there, their uneasiness at taking a stance was obvious to the American officials. Before attending the White House meeting, the British ambassador first asked whether his presence would be acceptable to the president. "The behavior of the British had given McKinley whatever final proof he may have been seeking that no European intervention on behalf of Spain was possible."⁷⁶³ In the end, on the eve of the Spanish-American War, Britain found itself being the last link in the buck passing chain—all the European powers had passed the buck to Britain. The French, Germans, Austrians, Italians and Russians had agreed to participate against the United States if Britain did. However, the European great powers had not realized that a few years prior at the end of the Venezuelan crisis, as the next section shows, Britain had essentially made the decision to bandwagon with the United States rather than balance against it. Germany, France,

⁷⁶² Memorandum of October 8, 1897, and of March 12 and 15, 1898, *Espagne: Cuba IV* (Fr.), quoted in May 205, 207; May 198, 234-5.

⁷⁶³ Grenville and Young 260-1.

and Russia had passed the buck to Britain, in large part because they needed the help of the British Navy to put pressure on the United States, and they were relying on Britain to make the final decision, yet they would not get it.

When it became clear that Britain was not going to side against the United States in early April 1898, the European great powers quickly withdrew their previous assurances to Spain, although most were still sympathetic to Spain's cause, and Spain remained on its own to face U.S. demands. The Germans suddenly retracted their offer of participation, and when the French found out, they also rescinded, bringing an end to the short-lived potential European unity.⁷⁶⁴ And the great powers continued to pass the buck amongst each other even as the war between Spain and the United States had commenced. After the German Kaiser learned through his consul in Hong Kong that the U.S. Navy was about to attack the Philippines and relayed the information to Spain, the Spanish Queen Regent personally appealed to him and "suggested that he could best prove his friendship by sending German warships on a friendly visit to Manila immediately." But the Kaiser remained cautious, and his foreign minister von Bülow even suggested that "Spain should ask the French for support" instead.⁷⁶⁵ In the end lack of trust and constant in-fighting among both the great powers and the secondary powers, and insufficient interest and buck-passing among the European great powers led to their complete collective inaction. May compares the potential balancers' "hesitancy, torpor, and disunity" of the European great powers in front of the U.S. rise in 1890s to "that of the Greek city states when confronting Rome in the second century B.C.," describing it as "an instance of momentous inaction."⁷⁶⁶ Yet

⁷⁶⁴ May 212, 217-9.

⁷⁶⁵ Memorandum by Bernhard von Bülow, German Foreign Ministry Archives, G.F.M. 13/141, qt. in Grenville and Young 282-3.

⁷⁶⁶ May 181,185.

collective inaction was not the sole cause of balancing failure on the potential balancers' side; widespread bandwagoning also served to facilitate the rising hegemon's success.

c. Bandwagoning (IV4)

The United States' rise was greatly aided by some of the potential balancers' decision to bandwagon instead of balancing. Not only did most of the secondary powers on the continent side with the United States, many on their own and some with a bit of pressure, but some of the great powers, in particular the United States' strongest potential rival, did too. Bandwagoners who sought collaboration with the rising hegemon did so exclusively out of profit (IV4.2). In no case was fear the primary vector of bandwagoning (IV4.1). Even in the few cases where secondary powers were bullied into siding with the rising hegemon, there was generally a strong profit incentive for them to give in rather than remain on the balancing side (IV4.3).

One may wonder whether the absence of fear-driven bandwagoning stemmed from the United States' smooth, informal expansionary practices. Because the United States rarely ever forcibly invaded its targets—except perhaps for the Philippines and Cuba temporarily—secondary powers may have wrongly believed that the consequences of bandwagoning might not be severe. As this section shows, the United States' use of the Monroe Doctrine and its claim to safeguard the independence of the Western Hemisphere against European great powers, promote Latin American development, and later prevent internal disorder portrayed the rising hegemon as a benevolent protector, gaining the trust of the secondary powers while obscuring the price of bandwagoning. The best example is that of Pan-Americanism, which the rising hegemon used to coax secondary powers to join it, offering commercial and other rewards. Spykman calls the U.S. drive for Pan-Americanism "a super-colossal trade promotion scheme in dignified attire." The

United States portrayed it as the "concept that the New World was basically distinct from the Old, that the states on this side of the Atlantic were similar in spirit and ideology and resembled each other more closely than they resembled the nations of Europe" and should therefore collaborate closely. But in reality this particular Pan American thesis was a myth to attract the Latin American states into bandwagoning by making them believe they would benefit from a closer relationship with the United States. "This thesis of a Pan American identity, cultural affinity, and similarity of political outlook was a noble idea, but completely invalid" given how distinct societies in United States and secondary powers were in reality.⁷⁶⁷

The United States' informal expansion and quest for the "friendship" of its targets is reminiscent of Rome's *clientela* strategy. Rome sought to become friends, rather than conqueror, of a number of its targets during its early rise, at a time where it was not yet powerful enough to take them over but still sought to have them in its orbit. The United States, similarly starting its rise with limited military capabilities, may have wanted to capitalize on its economic assets while minimizing the risk of full-blown military interventions and thus used primarily informal, seemingly benevolent expansion techniques rooted in Pan-Americanism and the Monroe Doctrine. As Gallagher and Robinson add, "perhaps the most common political technique of [informal] expansion was the treaty of free trade and friendship made with ... a weaker state." Such informal techniques gave secondary powers the impression that they could gain tremendously by siding with the United States—they retained their formal independence and enjoyed economic exchange and developmental support.⁷⁶⁸

Numerous secondary powers on the continent would have likely bandwagoned even without special American efforts to convince them of the advantages of doing so. The

⁷⁶⁷ Spykman 235, 246.

⁷⁶⁸ Gallagher and Robinson 11.

geographical setup and distance of the other great powers likely played a significant role in triggering the unusually large occurrence of secondary-power-bandwagoning in this case. Weaker powers often look for a protector to help them ensure their security, and because they do not have as many options as great powers have, they often join the most accessible great power. "The placement of states affects their behavior."⁷⁶⁹ In this case, geography mattered because all great powers needed far greater power projection than the United States. The United States had the advantage of location, and hence its support for a small state was more valuable and reliable than support from another great power, which would have to travel further to assist it and might hesitate to intervene because of the distance. As Mearsheimer notes, "the United Kingdom had to project power across the Atlantic Ocean into the Western Hemisphere, whereas the United States was physically located there." The spread of steam navigation in the late 19th century allowed the United States to reach even South America faster than its European rivals, particularly as its footholds in Central America and the Caribbean multiplied.⁷⁷⁰

As is always the case with bandwagoning, however, support was not free. Just like Rome's friends, Latin American bandwagoners had to cope with varying degrees of American intrusion into their domestic and foreign policy decisions in exchange for the collaboration. The Monroe Doctrine, which began as the U.S. defense of continent's integrity against European conquest and colonization and upholding of Latin American independence and self-rule, evolved during the U.S. rise to hegemony into a principle allowing American intervention and encroachment into the government of Latin American states, almost the polar opposite of the Doctrine's original meaning. The heyday of America's rise, particularly the McKinley and Roosevelt presidencies, witnessed "the evolution of a doctrine which was intended for the

⁷⁶⁹ Kenneth Waltz, *Theory of International Politics* (Boston, MA: McGraw-Hill, 1979), 127.

⁷⁷⁰ Mearsheimer 247.

protection of Latin American states by the United States into one justifying and even sanctifying American intervention in and control of the affairs of the independent republics of this continent." But since "the heart of Monroe's message [of 1823] lay in the recognition of the independence of the new states of the New World, it is curious indeed that it should be twisted with time to justify a course which involved ... interference with that independence."⁷⁷¹ The Spanish–American War set the precedent for U.S. meddling with its targets' sovereignty. After the Spanish-American War, the United States planned to withdraw from Cuba but only after "a period of American tutelage" to ensure that "they could be trusted to make wise use of a representative government." The Platt amendment ensured that the United States could exercise such "tutelage" again later on.⁷⁷² Thus, with the 1901 Platt Amendment "the American government proceeded to claim for itself a special right of interference in the affairs of the country which it had set forth to free," and it acted on it repeatedly in Cuba and beyond.⁷⁷³

In hindsight, then, the Monroe Doctrine can be really seen as a modern equivalent to Rome's clientelae diplomacy. Just like Rome, the United States was able to develop a system allowing it broad intervention powers into its weak neighbors' foreign and domestic policies. The 1904 Roosevelt Corollary declaration, later formally ingrained in a treaty, completed the "transformation" of the Monroe Doctrine and reinforced the new U.S. policing powers on the continent. The United States was subsequently able to control the financial, budgetary, electoral, commercial, and other policies of a number of its neighbors. Even though many secondary states in question had requested American support and voluntarily bandwagoned, the price was steep and reactions varied. Some governments embraced American intrusions and called for more to help themselves stay in power, while others simply discretely coped with them, and others

⁷⁷¹ Perkins 396-7, 451.

⁷⁷² Grenville and Young 263.

⁷⁷³ Perkins 398.

openly disapproved—like Cuba did of the Platt Amendment, for example—but due to their weak state and absence of alternate international support, they did not have a choice. By the time of the Roosevelt Corollary, there was "a growing distaste for American interference with the republics of the Caribbean [and Latin America], ... but in 1905 and 1906 one can hardly speak of any widespread or flaming indignation at the Roosevelt Corollary," especially since the United States was careful to frame its interventions as "friendly," which was a style reminiscent of Rome even in rhetoric, and took steps "to make it more palatable," emphasizing America's benevolent intent.⁷⁷⁴

Regardless of the risks, the proximity of the United States and its smooth informal expansion techniques made examples of Latin American states willingly seeking to bandwagon for advantages—domestic political benefits, economic gains, or simply advantages over their neighbors—particularly plentiful. In some Latin American and Caribbean countries, Perkins explains, "it was at times the fashion for a hard-pressed government to insinuate to the United States that sinister foreign intrigues were afoot, in the hope that these insinuations would stimulate the American government to appropriate measures of protection." These efforts were more or less subtle. Haitian president Salomon, for example, who was confronted with an insurrection in the 1880s, told the U.S. delegate in Haiti that the rebels were being supported by Britain in exchange for a promise to reward Britain with the area of Môle St. Nicolas as a naval base should the revolution succeed. Salomon then promised the base to the United States against gunboats and financing. Although American authorities did not fall for it, the scheme shows how far local Latin American leaders were willing to go in order to secure U.S. support for their own political gains. Seeing that his ruse failed, Salomon later upped the ante by taking steps toward requesting a French protectorate in exchange for some territory, in an effort to play into the

⁷⁷⁴ Ibid., 403, 452-3, 455-6.

Monroe Doctrine to obtain U.S. material support for his regime. The profit motivate became even clearer a few years later, as Grenville and Young stress that "once the ... Haitian President no longer needed American aid, he refused to live up to his earlier promise" of granting the base at Môle St. Nicolas. A similar incident occurred with San Domingo a few years later during the Harrison administration (1889-1893). The Dominican president "practically invited [the United States] to ... seize a naval base there" to help his own political purpose.⁷⁷⁵

Besides domestic political gains, another expected benefit from bandwagoning was to score points at the expense of a neighbor. Panama bandwagoned with the United States in 1903 to obtain independence from Columbia, for example. While the United States prevented Columbian troops from intervening in Panama's revolution and thus effectively guaranteed Panama's independence, the newly formed state paid a hefty price to the United States: rights to build and operate the isthmian canal. The canal deal included U.S. control of a ten-mile radius, which raised numerous issues of sovereignty for Panama, and led the United States to repeatedly become involved in Panama's domestic politics, sometimes at the request of Panama's authorities but more frequently of its own volition. Hostility and lack of trust were so great between some states on the American continent that some voluntarily appealed to the United States to solve their disputes. For example, Guatemala and Honduras sought to have their boundary dispute arbitrated by the President of the United States, and in the mid-1920s the United States set up a conference led by an American judge to resolve the issue. The U.S. government was directly involved in negotiating the settlement of the boundary. Some Central American governments, such as Nicaragua in 1910-1911, also directly requested American intervention when they were unable to tackle their internal revolutions or their financial problems. Nicaragua went even further in 1914 when it tried to obtain a Platt-like provision in the Bryan-Chamorro Treaty giving

⁷⁷⁵ Grenville and Young 94-5; Perkins 34-5.

the United States the right to intervene in the country to secure the country's independence and protect life and property there.⁷⁷⁶

Profit for yet other small Latin American countries took the form of leverage against another great power. Venezuela's behavior during its crisis with Britain provides the perfect example. While confronting Britain over its boundary with British Guiana, Venezuela "was doing all it could to entangle the United States." The Venezuelan government did not hesitate to lease land with valuable mineral resources to American businessmen, and it employed American diplomat William Scruggs to lobby its cause in front of Congress, the President, and the American public. The United States eventually intervened on behalf of Venezuela and forced Britain to back down, but Venezuela paid a hefty price for bandwagoning. Although the United States forced Britain to renounce war and submit to arbitration, it excluded from arbitration all properties owned for fifty or more years, which greatly favored Britain, and at no point consulted with Venezuela. In 1899, when the arbitration committee eventually handed most of the territory at stake to Britain, the United States supported the decision and did not intercede for Venezuela.⁷⁷⁷

But secondary powers were not the only ones to bandwagon with the United States; some great powers sided with the rising hegemon, too. The first to do so, Britain, was perhaps the most surprising, because it was the most powerful potential balancer against the United States. Britain expected two benefits from bandwagoning with the United States. First, it hoped that by recognizing American supremacy in the Western Hemisphere and resolving ongoing areas of friction with the United States it would be able to free its own resources and redirect them on its overstretched imperial commitments around the world and its more pressing conflicts with the

⁷⁷⁶ Munro 68-9; 223-4, 232, 235.

⁷⁷⁷ Zakaria 149, 152.

other European great powers. Secondly, it anticipated that the weight of a transatlantic partnership would reinforce its power position vis-à-vis its European rivals.

Even before the Venezuela crisis of 1895-6, after which Britain decided to formally side with the United States, there were hints that it might join the rising hegemon's side. In fact, pressed by the fierce competition of other European great powers like France and Germany, as well as of the United States, Britain repeatedly contemplated cooperating with the United States despite their many differences in the early-to-mid 1890s, to form an Anglo-Saxon barrier against the other great powers. For example, Colonial Secretary Chamberlain sought to act jointly with the United States in Armenia, suggesting a joint U.S.-British naval exercise to force the Ottoman Sultan to stop the ongoing genocide of Armenians and to engage in a number of reforms sought by the British government. Chamberlain expected to gain from the cooperation, as a letter to Prime Minister Salisbury shows: "If the Americans were with us we need not fear any interference from France or Russia—they dare not provoke a combination of two Anglo-Saxon nations." In addition, Chamberlain believed the cooperation would alleviate the United States' increasingly hostile position on Venezuela, writing that "in the stir created by such an alliance the Venezuela difficulty would be lost sight." Although PM Salisbury ultimately rejected the idea, Chamberlain continued to publicize the possibility of an alliance and put it on the table again a year later when Russia renewed its expansionist pursuits in China. This time the British government officially asked for a U.S.-British rapprochement in a secret message to Washington, but the U.S. response was "evasive," showing that while other great powers might have been tempted by bandwagoning, the United States did not need or seek British help in establishing hegemonic control over the continent.⁷⁷⁸

⁷⁷⁸ Chamberlain to Salisbury, Dec. 24, 1895, qt. in May 53-4.

Britain's decision to side with the United States became fully apparent when it backed down during the Venezuela crisis and opted to yield to the United States' quasi-ultimatum, after realizing that it could not afford to keep pushing back against the United States and its other rivals in other regions all at once. Starting in the mid-1880s, British naval supremacy had begun to wane all over the world as Britain was increasingly confronted with the competition of European and other local rivals. "As a consequence, because it was overstretched, Britain began to shift its naval power from abroad to home waters and to strike deals." Britain's first deal was to agree to grant the United States primacy on the American continent and to become an unofficial ally of the United States. Overcoming their previous hostility that had culminated in the Venezuelan crisis, the British government began envisioning a "partnership" between the United States and Britain and initiated talks to resolve all disputes with the United States, including the uncertain Alaska-Canada border and the projected American-run trans-American canal. Britain began negotiating the Hay-Pauncefote Treaties, which was eventually signed in 1900 and 1901, in which the Britons agreed to leave the responsibility of building and operating a trans-isthmian canal solely to the United States, an issue on which they had previously been particularly inflexible. Britain yielded again in 1903 and supported the United States in its land dispute with Canada over a gold-rich stretch of Alaskan coast.⁷⁷⁹

Britain's refusal to support Spain against the United States in 1898 further deepened its rapprochement with the United States. Because it was last on the buckpassing chain, Britain's decision to bandwagon instead of balance ultimately became responsible for the failure of collective action in the wake of the Spanish-American War. Since Britain still had significant territorial possessions and a large military and naval presence in the Western Hemisphere, its

⁷⁷⁹ Art, *Grand Strategy*, 182-3. Britain later settled with other great power rivals (1902 Anglo-Japanese Treaty, 1904 Dual Entente, and 1907 Triple Entente), signaling further retrenchment.

unwillingness to participate in any balancing actions against the United States essentially doomed the efforts of other balancers seeking to act collectively against the United States. Without the support of the powerful British Navy, there was virtually no chance at projecting sufficient power in the Western Hemisphere to rein in the United States. In 1898, although there was "astonishment and alarm" in England after the seizure of the Philippines, the British perceived a direct U.S. claim against one of their own imperial possessions in the area as improbable and therefore continued "a policy of staying out of America's way," which was, at this point, the best Britain could do to appease the rising hegemon.⁷⁸⁰

By the late 1890s, Britain had entirely changed its stance toward the United States. It "had recognized that America was the most significant rising power ... and that British interests in the Western hemisphere were exposed." Instead of "contest[ing] this nascent power" it chose to "accommodate it."⁷⁸¹ After defeats in the Boer Wars and growing difficulties with other colonies, which highlighted their weakness and overstretch, Britain began to "withdr[a]w substantial military force from the Americas and limit ... its political interests there." At the turn of the century Britain was thus unlikely to revert to balancing the United States, given "all the areas of the world where its interests were imperiled." Indeed, the British government had decided to "concentrate Britain's resources where its most vital interests were at stake, that is, in the home waters and the Mediterranean."⁷⁸² As Mearsheimer concludes, "during the early years of the twentieth century ... the United Kingdom retreated across the Atlantic Ocean and left the United States to run the Western Hemisphere."⁷⁸³ While the Germans continued to be hostile to the United States, the British were "content with the status quo in the Western Hemisphere" after

⁷⁸⁰ May 225.

⁷⁸¹ Zakaria 166, 168.

⁷⁸² Grenville and Young 307-8.

⁷⁸³ Mearsheimer 246.

1903, which allowed it to focus its resources on its colonies in other regions and on the European balance of power. With the Hay-Pauncefote Treaty, the British Cabinet formally decided to "abandon all pretense of attempting to match the growth of American power in the Western Hemisphere."⁷⁸⁴

For a short while following the American victory over Spain, the German government also reflected on a rapprochement with the United States. "Shortly after the American naval victory at Santiago [i.e., Cuba], officials in Berlin began to think of an American alliance as a possible means of strengthening Germany in Europe."⁷⁸⁵ They also had their own profit in mind. Their main goal was to preempt a U.S.-British-Japanese offensive alliance, the worst possible nightmare for Germany. Foreign minister Richthofen calculated that "in order for us to upset this hope [i.e., Britain's alleged desire of an alliance with the United States and Japan], it is not England but America whom we should approach." The Kaiser agreed with him, noting in a memorandum that "we must hasten to reach an understanding with the Yankees." But he failed to make an overture, because just then the Manila battle occurred and Spain seemed poised to lose the Philippines, prompting Germany to look for its own piece of the pie in the Spanish colonies of Mindanao, the Carolines, and Samoan archipelago, which would naturally lead to rivalry with the United States. Thus, the idea of a rapprochement dissipated.⁷⁸⁶

Finally, after it had become clear at the turn of the century that Britain was firmly planted on the United States' side and intra-European rivalries had been accelerating, the European great powers, little by little, let the United States take greater responsibility in Latin America,

⁷⁸⁴ Grenville and Young 307.

⁷⁸⁵ May 225-8.

⁷⁸⁶ German Foreign Ministry, *Die Grosse Politik der Europäischen Kabinette, 1871-1914: Sammlung der Diplomatischen Akten des Auswärtigen Amtes - hrsg. von Johannes Lepsius, Albrecht Mendelssohn Bartholdy, Friedrich Thimme* (Berlin: Deutsche Verlagsgesellschaft für Politik, 1922), vol. XIV, 3530, 3808, 3823, 3839, vol. XV, 4151, 4160-5, 4232.

engaging in an inactive or passive type of bandwagoning of their own. In a telling evolution, the European great powers gradually agreed to let the United States become their intermediary for all debt recovery on the American continent. At the end of the 19th century it was still "accepted practice for states to use military force to collect debt owned to their nationals by other states," and there were a number of precedents where European powers sent troops to forcibly collect debts from Latin American countries—for example France in Mexico in 1838, and Britain, France, and Spain in Mexico in 1862. This explains why the United States was wary of foreign intervention in several cases where weak, unstable Latin American countries owed substantial debt to Europeans, such as Cuba, Nicaragua, Haiti, San Domingo, and Venezuela. In 1902, for instance, German and British warships established a blockade of Venezuelan ports, sinking Venezuelan boats and firing onto coastal forts, in an effort to obtain arbitration in the repayment of Venezuelans debts owed to German and British businesses. In 1907, to prevent further European debt recovery interventions, the United States obtained a treaty at the Hague Peace Conference that bound Latin American countries to a compulsory U.S. arbitration of debt issues and committed the Europeans to abstain from using force to recover their debts. In essence the treaty legitimized the American "dollar diplomacy" and secured the de facto approval of the other great powers for the United States' financial stabilization incursions into Latin America and recovery of debt on their behalf. When the United States intervened in Nicaragua in 1912 and in Haiti and San Domingo 1916 for such purposes, for instance, there was virtually no backlash in Europe. On the contrary, the Europeans were relieved that the United States bore the costs of intervention and recouped their funds.⁷⁸⁷

⁷⁸⁷ Martha Finnemore, *The Purpose of Intervention* (Ithaca, NY: Cornell University Press, 2003), 27-9, 33.

In the end, bandwagoning, both active and passive, was widespread and directly contributed to easing the United States' achievement of hegemony on the continent. As May concludes,

Not one of the [great] powers recurred seriously to the thought of uniting against an American peril. The British and German governments considered seeking [or sought] an American alliance. The French, Russian, and Japanese, whose interests were most seriously affected by America's move into the Philippines [and Hawaii], remained silent.⁷⁸⁸

The secondary powers also gave in to the potential gains they could derive from collaborating with the United States, and because of the distance of other potential protectors, did not hesitate to bandwagon en masse. Thus, collective inaction and bandwagoning largely explain the failure of the United States' rivals and neighbors to stop its rise. But their failure alone does not explain the United States' hegemonic success.

2. The Rising Hegemon's Side: Unmatched Military and Economic Growth

The United States' meteoric military (IV5) and economic growth, as well as its geographic advantage (IV6), may have played a crucial role also in enabling it to become the region's hegemon.

a. Military Achievements (IV5)

The United States' prowess in the military sphere lay not so much in dramatic innovations giving it the edge over its competitors, but rather in the sudden and spectacular buildup it

⁷⁸⁸ May 237.

undertook starting in the 1880s that quickly led it from being a military underdog to closing the gap with its fellow great powers. Coupled with its geographic advantage, requiring less power projection than its great power rivals to reach its target areas in the region, the United States' military buildup quickly transformed it into the leading great power of the Western Hemisphere. The U.S. military buildup was remarkable in that the United States started from far behind, not even considered a great power by many at the onset of its rise, and within a mere thirty year time span turned into the leading great power of the region. Following the Civil War, during which the United States had shown its ability to raise a large army and navy, the United States demobilized its conscript army, keeping only a small regular force, and let its navy fall into decrepitude. Successive U.S. governments made no efforts to keep up with the other great powers' naval constructions and technological innovations, and by the 1870s and early 1880s the United States would have been incapable of defending itself against a British attack on its harbors or naval commerce. In 1874 the authorized force consisted of 25,000 enlisted soldiers and 2,161 officers, though "the actual force was even smaller." It was mostly used to conquer and pacify the Western frontier, against a technologically inferior rival.⁷⁸⁹ The naval force was quasi inexistent. Purely defensive, it consisted of a few coastal defense ships, commercial raiding vessels and mines. The weakness of the American Navy in the early 1880s was "depressing." Reports from the Navy Secretary concluded it was in a "derelict" state. Even more tellingly, future Secretary of the Navy and then U.S. representative John Long described the Navy in 1885 as "an alphabet of floating wash-tubs."⁷⁹⁰

By the early 1880s, the United States had accumulated a large surplus thanks to its booming economy, and throughout the decade began spending more on improving its military.

⁷⁸⁹ Zakaria 122-3.

⁷⁹⁰ Grenville and Young 10; May 7.

President Chester Arthur (1881-1885) was the first to promote naval growth, after realizing that there was a large discrepancy between America's economic power and its international influence and reputation, as well as a vast gap between U.S. capabilities and the capabilities of all the other great powers. One of the first steps the U.S. government took was to recognize that both the army and navy were poorly organized, inadequately trained, and inefficient in their use of resources, and that led to drastic educational reforms. The army established a number of graduate schools for professional training such as Fort Leavenworth, and in 1884 the Naval War College opened. The army and navy also created intelligence divisions, and senior level staff positions in the administration, initiating a thorough reorganization and rationalization of the armed forces that would give the executive the tools to carry out a more emancipated foreign policy.⁷⁹¹

The U.S. naval buildup formally began in the mid-1880s and continued through 1922, when the Washington Naval Treaty freezing warship tonnage was signed. First, starting in 1883, the United States undertook a thorough modernization of its fleet, under the impulse of Secretaries of the Navy William Hunt and then William Chandler, to reverse the "technological and material" neglect it had fallen into since the Civil War.⁷⁹² Hunt and Chandler convinced Congress to fund a vast increase in existing naval capabilities—mostly defensive, like coastal guns, mines and ships. But "the strategic concepts of the Navy Department ... were [still] as outdated as the warships." They focused on narrow and "static" coastal defense and commerce-raiding capability to deter attacks, and lacked the most basic power projection elements.⁷⁹³ The "real turning point" for the rise of the U.S. Navy thus occurred in 1890, when the United States engaged in a strategic shift from a fleet primarily aimed at coastal defense and commercial raiding to "the construction of an ocean-going, blue-water, battleship navy" based on larger,

⁷⁹¹ Zakaria 77-8; 122-6.

⁷⁹² Art, *Grand Strategy*, 183.

⁷⁹³ Grenville and Young 11.

offensive heavy-armed boats like armored cruisers, clippers, and monitors.⁷⁹⁴ "By 1890, American grand strategy had been transformed."⁷⁹⁵ That year, the congressional Naval Policy Board and Secretary of the Navy Benjamin Tracy issued reports calling for the development of a battleship fleet, leading Congress to vote and allocate funds for the building of the first battleships, and the publication of Alfred Thayer Mahan's *Influence of Sea Power on History*, praising the primacy of trade and naval capabilities as sources of power, gave the naval buildup and switch to the offensive an intellectual justification.⁷⁹⁶

The new strategy's impact on American capabilities was dramatic. The United States constructed fifty-two battleships between 1891 and 1919, including some of "the most heavily armored and gunned vessels in the world."⁷⁹⁷ In 1890 the United States was 6th in warship tonnage after Britain, France, Russia, Italy, and Germany, and it could not have withstood a conflict with another European great power. But by the eve of World War I, the United States had climbed to 3rd in tonnage, still way behind Britain but closing in on Germany, both of which were engaged in a naval arms race of their own in Europe. By the end of World War I, the United States possessed more battleships than the fleets of Japan, France, and Italy combined. At the Washington Treaty, the United States obtained parity with Britain in allowed tonnage, a remarkable feat considering that just thirty years earlier, its tonnage was over six-and-a-half-times less than Britain's.⁷⁹⁸ The new capabilities immediately boosted the United States' influence over the continent and beyond. Even though the U.S. economy had been booming throughout the 1870s and 1880s, "statistics on resources and production remained irrelevant until

⁷⁹⁴ Art, *Grand Strategy*, 183.

⁷⁹⁵ Zakaria 80, 146.

⁷⁹⁶ May 8-9.

⁷⁹⁷ *Ibid.*, 7.

⁷⁹⁸ Art, *Grand Strategy*, 183. Since the United States was an aspiring *regional* hegemon and the American continent was outside of Britain's core region, U.S. naval parity with Britain in global terms meant massive U.S. naval supremacy on the American continent.

the United States possessed some evident capacity to assert that it had interests beyond its shores and that it was prepared to protect them."⁷⁹⁹ The construction of a "battleship fleet with a large coal-carrying capacity ... ma[d]e possible the assertion of American power beyond the continental confines."⁸⁰⁰

Because the other great powers were external to the Western Hemisphere, the United States started to acquire growing influence over the region soon after it began its military growth, even before it had formally outpaced its great power rivals militarily. President Cleveland wrote in 1993, for example, that the U.S. Navy had been instrumental in advancing American interests in several areas: Nicaragua, Guatemala, Costa Rica, Honduras, Argentina, Brazil, and Hawaii. By the time of the Venezuelan crisis of 1896, President Cleveland was ready and willing to risk war with Britain. He asserted that he was "fully alive to the responsibility incurred and keenly realiz[ing] all the consequences that may follow." The reason was that "for Washington, the controversy between London and Caracas was a welcome pretext for asserting America's claim to geopolitical primacy in the Western hemisphere" and showed that the United States was starting to have the military capabilities to back up its claims. The United States essentially provoked a "showdown" with Britain to demonstrate that Britain was no longer a match for it militarily in the region. Cleveland "expected" the Britons to back down. Salisbury refrained from going to war in 1896 because the United States possessed increasingly strong military capabilities and could prevail in the long term due to its vast economic resources, especially as Britain could not afford to spare large amounts of resources to combat the United States when it

⁷⁹⁹ May 7.

⁸⁰⁰ Grenville and Young 10, 34.

was also involved in other theatres and did not want to risk possibly losing Canada if things went wrong with the United States.⁸⁰¹

Illustration 8.2: "Columbia's Easter Bonnet"



Source: Samuel D. Ehrhart, "Columbia's Easter Bonnet," illustr. cover of *Puck*, a political satire magazine, April 6, 1901 (Library of Congress Prints and Photographs Division, Keppler and Schwartzmann, copyright).

⁸⁰¹ Christopher Layne, "Kant or Cant: The Myth of the Democratic Peace," *International Security* 19.2 (August 1994), 22-3, 25; Zakaria 146.

The easy victory in the Spanish-American War in 1898 made the other great powers plainly aware of the tremendous U.S. military growth that was underway, even though Spain was a declining great power. "Most European magazines and newspapers depicted the United States as a bully picking a quarrel with a nation she badly overmatched." The war was quick, famously nicknamed "splendid little war" by U.S. Secretary of State John Hay. It was certainly "little" for the United States. Even on the naval side, where the United States had been particularly weak until just a few years before, the United States dominated Spain. Six American warships defeated a numerically superior Spanish naval force in just seven hours off of Manila. Even "the best informed writers had not credited the American navy with such enterprise and efficiency." More importantly, the United States was able to mobilize in record time despite its diverse population, something the other great powers did not expect: "old Confederates [took] command alongside former Union officers, Democrats and Republicans enlist[ed] with matching zeal, [and] northern immigrants and southern Negroes line[d] up before recruiting booths."⁸⁰² Yet despite the emerging superiority the military buildup did not cease there. President McKinley, who took office in 1897, further pursued the American naval surge. By the end of his presidency in 1901, he had "ensured that the United States had access to and a large measure of control over sea lanes, port facilities, strategic locations, and even the domestic order of foreign countries in the Caribbean and Latin America."⁸⁰³

By 1901, when President Roosevelt took office, "the Navy was large, continuously expanding, and steadily acquiring new bases abroad." A former Assistant Secretary of the Navy who had been instrumental during the Spanish-American War, Roosevelt "preached to Congress and the nation that the defense of America's interests required a large fighting fleet." He not only

⁸⁰² May 220.

⁸⁰³ Zakaria 155, 164-5.

further continued to grow the size and scope of the U.S. Navy but also sent sixteen battleships and their escorts on a trip around the world between 1907 and 1909 demonstrating to the world the United States' growing military power and its blue-water capability. At the same time Roosevelt and his Secretary of War Elihu Root reformed the organization of the professional army to maximize its efficiency. They adopted the praised Prussian-inspired general staff system, improved promotion methods, instigated rotating officer positions, and refined the army education system with a new Army War College and separate curricula for special branches.⁸⁰⁴

In the end, even though at the beginning of its rise the United States lay far behind its rival great powers in terms of military capabilities, the massive and sustained military—and particularly naval—buildup it initiated in the 1880s enabled it to surpass its competitors militarily in the Western Hemisphere by the first decade of the 20th century and soon become one of the world's most powerful military actors.

b. Non-Military Achievements and Characteristics (IV6)

But military preponderance alone, especially since it came relatively late and did not involve any major innovation, cannot explain the United States' vast success in extending its control throughout the American continent. Besides military capabilities, the United States developed and took advantage of a number of other assets that gave it a decisive edge over its rivals: geographic proximity was a crucial factor (IV6.1), and the United States' overwhelming economic power (IV6.2) and masterful political and diplomatic influence (IV6.3) were key elements because the rising hegemon relied primarily on informal expansion—i.e., economic and political/diplomatic—rather than military expansion.

⁸⁰⁴ Grenville and Young 309.

As already demonstrated by the great powers' lack of interest in balancing and the secondary powers' unusually high tendency to bandwagon, the unique geographical isolation of the United States in the Western Hemisphere, with all rival great powers based in a different region and separated from the United States by two oceans except for a few outposts, gave the aspiring hegemon a considerable advantage (IV6.1). But that was not always the case, as "before the construction of the Panama Canal, Europe was not only nearer than the United States to the east coast of South America, but also nearer to the west coast." In addition, during most of the 19th century, Britain, France, and Spain possessed naval bases in the Caribbean which were closer to the South America than the United States. Those "European states were, therefore, in a better position to exert naval pressure on the republics to the south than the government in Washington." This explains why the United States did not apply the Monroe Doctrine beyond Brazil during the early phase of its rise: it simply could not. The United States merely "protested" when Spain took over the Chicha Islands in 1864 during its war with Peru, but it could do nothing more. And it did not even object when Britain annexed the Falklands in 1833 or when the French and British made incursions in Argentina in the 1840s and 1850s. The expanding use of steamships for commercial navigation and the growing military capabilities of the United States in the late 19th century, which increased the United States' projection power, dramatically changed that, enabling the United States to reach the confines of the continent faster than all rival great powers. By the late 19th century, "the relative position of Europe and the United States ... [in South America] changed in favor of the latter," leading President Roosevelt to extend the doctrine to apply not only to European powers but also to Asian powers and to both oceans.⁸⁰⁵

⁸⁰⁵ Spykman 87-88.

Once the power projection of the United States surpassed that of the other great powers, the United States became well aware of its geopolitical advantage, guaranteeing security, prosperity, and the greatest chance at hegemony. It was a "continent-sized power possessing abundant natural resources, surrounded by militarily weak neighbors, and separated by two vast oceans from the bulk of the world's miseries and from easy reach by the forces of most other powers of consequence."⁸⁰⁶ U.S. leaders knew their geographic situation helped compensate for their late military growth. In 1889, a policy board of naval officers charged by Secretary of Navy Benjamin Tracy to conduct a strategic review recognized that the United States was blessed with a favorable geography. They wrote that "for the United States it may be confidently asserted that the chances of war with any nation comparable in wealth and power are much less than the chances among the nations of Europe." At the onset of the Venezuelan crisis, Secretary of State Richard Olney similarly remarked that "today the United States is practically sovereign on this continent ... because ... its infinite resources combined with its isolated position render it master of the situation and practically invulnerable as against any or all other powers."⁸⁰⁷

In addition to its highly favorable geographic position, the United States' booming economy also played a significant role in its hegemonic success because it became a central vehicle of U.S. influence throughout the Hemisphere (IV6.2). During the first phase of its rise in the aftermath of the Civil War, the United States developed a strong, modern, capitalistic economy "with a highly developed industrial system supplementing a vast extractive economy devoted to agriculture and mining," and became an economic powerhouse. The massive economic growth was partly induced by a jump in U.S. population. Pushed by a large middle class, the United States went through a massive physical growth through the latter part of the 19th

⁸⁰⁶ Art, *Grand Strategy*, 172, 178.

⁸⁰⁷ Zakaria 129, 149.

century.⁸⁰⁸ Not only did the expansion of the frontier mean that it covered more territory than any European state save Russia, but it also underwent a dramatic population increase: while in the 1850s the United States had a smaller population than France or England, by the 1880s "it led every industrialized nation but Russia." In addition, the United States benefited from a very fluid social structure and thus did not experience the class conflicts that undermined European economies in the 19th century. As American industrial production rose sharply throughout the 1870s and 1880s, "American businessmen were pushing aggressively into foreign markets," competing with the British, French, Dutch, Russians, and Germans in the oil, cooper, and steel markets. Boosted by "economic and political freedom under a 'laissez-faire' system combined with almost inexhaustible resources and a large market" both domestically and throughout the continent, the United States reached economic hegemony long before it achieved political hegemony over the continent.⁸⁰⁹

U.S. economic growth was unmatched, both in agriculture and industry. Virtually every sector of the U.S. economy thrived after the Civil War, despite several depressions in the 1870s and 1890s. Agricultural production soared: wheat production grew by 256% between 1865 and 1898, corn by 222%, and sugar by 460%. By the 1880s the United States had become the first producer of wheat in the world. Industrial output was even more impressive: coal production climbed by 800% between 1865 and 1898, and steel rails by 523%. Oil production jumped from 3 million barrels in 1865 to 55 million barrels in 1898. The boom attracted a massive wave of immigrants, on top of an already soaring natural population growth, leading the population to double between 1865 and 1900, which in turn further fueled economic production. But in comparison to other industrializing countries, America's economic rise was even more

⁸⁰⁸ Spykman 213-5.

⁸⁰⁹ May 9.

staggering: while the U.S. annual growth rate reached 5% for most of the second half of the 19th century, the growth rate of its closest economic rival, Britain, averaged only about 1.6%. In fact, by the mid-1880s the United States had surpassed Britain in shares of manufacturing and steel production and by the early 1890s it had also become the first producer of coal and iron in the world. In 1892 the United States was "overtaking or surpassing Europe by nearly every measurable standard," and the disparity between the U.S. economy and all other industrialized economies would only continue to increase. In addition there was more capital in U.S. banks than in any other country, ready to fuel even further growth.⁸¹⁰ Eventually, "by 1900 ... [the United States] had the most powerful economy in the world." While in 1860 Britain accumulated 59% of the world's relative wealth and the United States still lagged behind at 13%, by 1890 the United States had taken the lead with 35% share of the world's wealth against 32% for Britain and 16% for Germany. In 1907, for example, the U.S. steel output had soared to 23.4 million tons, 3.6 times that of Britain and about twice that of Germany. By 1910, the gap had become momentous, with the United States claiming 48% of the world's wealth against a mere 15% British share and 20% German share.⁸¹¹

Foreign trade still seemed relatively low—8% of the U.S. GNP in 1913 as opposed to Britain's 26%—but these figures are deceiving because of the massive size of the U.S. GNP. Moreover, the United States' vast internal market enabled it to grow considerably without relying heavily on foreign exchange. In reality, the United States had been steadily gaining shares of the Latin American market from its European competitors since the 1880s, when the United States realized that European great powers still dominated trade with Latin America and decided to step up their commercial dealings with their neighbors. John Bacon, U.S. representative in Paraguay

⁸¹⁰ Ibid., 6.

⁸¹¹ Mearsheimer 220, 235. Note that Mearsheimer calculates relative wealth as an aggregate value of relative iron and steel production and energy consumption, see p. 71. See also Zakaria 45-6.

and Uruguay, noted to Secretary of State Bayard in 1888 that Europeans vigorously pursued trade opportunities in the area, and that the United States should do the same. Out of all foreign ships that docked at Montevideo in 1883, for example, only one was American while 407 were British, 231 were French, 209 were German, and 90 were Italian. In several Latin American countries, the Germans outpaced the Americans in trade levels, even though they were newcomers in many areas. And while the British had considerably invested in their merchant marine, its U.S. equivalent had been neglected. Beginning in 1889 with the Harrison administration, the United States became more assertive and succeeded in securing increasingly larger shares of the Latin American market. The first step consisted of calling in the first Conference of American states in 1889, which was attended by every American republic save San Domingo and resulted in dramatically increased trade relations. The second was to secure American control of the trans-isthmian canal at Panama, essentially establishing U.S. control over a large part of the continent's commercial relations.⁸¹²

The United States' two-phased rise with an initial exclusive focus on economic growth thus paid off. Not only did the United States become an economic giant, but in addition, thanks to the vast economic assets it developed, it was able to get involved with Latin American economies early in its rise without yet appearing militarily intimidating. The United States similarly exploited the vast political advantage it possessed over the other great powers. Its stable political system and informal, almost restrained, political expansion differed starkly from its rival great powers' and made it more appealing to its targets throughout the region (IV6.3).

The United States' own political system, first, stood out from the other great powers' by the stability and equal opportunity it offered, in sharp contrast to the more archaic structures of Europe, which consisted either of rigid, un-democratic monarchies or highly unstable

⁸¹² Grenville and Young 79-81, 91-2; Zakaria 131.

parliamentary systems. As Albert Blaustein, a constitutional scholar and consultant who helped draft a number of constitutions in emerging countries, notes, the American political system became an attractive role model for the entire region, to such an extent that upon obtaining independence from Spain throughout the 19th century, a majority of new Latin American republics adopted American-style systems replicating the U.S. constitution and institutional structure—Mexico, Venezuela, Argentina, Uruguay, and Chile, for example. The United States even contributed to writing some constitutions in the region in the early 1900s, like that of Cuba, the Philippines, and Haiti.⁸¹³ At a time when the European great powers, the United States' main competitors for influence on the American continent, were struggling with their own domestic political structures, oscillating between failed democratic attempts, violent revolutions, and remnants of authoritarian rule, the United States "operate[d] as a representative democracy on the basis of a free vote expressed in a secret ballot." The two-party presidential system "contribute[d] significantly to the preservation of liberal government and the prevention of political extremism" that plagued Europe, with both parties firmly anchored in the large centric middle class and providing an eminently stable social environment, prompted large emulation throughout Latin America.⁸¹⁴

In addition, because it did not engage in formal conquest but instead used informal expansion tools, the United States appeared more restrained than European colonial powers and treated its targets significantly better, reinforcing its attractiveness in the region in contrast to its rival great powers. The United States' new, more subtle type of hegemonic quest significantly improved its chances of generating political and diplomatic cooperation from its targets, as

⁸¹³ Albert P. Blaustein, "The U.S. Constitution: America's Most Important Export," *Issues of Democracy* 9.1 (March 2004), 9; Robert J. Knowlton, "The Early Influence of the United States Constitution in the Western Hemisphere: The Cases of the Mexican Constitution of 1824 and Bolivar's Ideas," *Discussion Paper No. 78*, University of Wisconsin-Milwaukee Center for Latin American and Caribbean Studies publication (March 1988).

⁸¹⁴ Spykman 215-6.

opposed to the frequent insurgencies the colonial powers were confronted to simultaneously in their possessions all over the world. The novelty of the United States' rise, seeking control not only through military pressure and occupation but also through more insidious political and economic influence, created a much more sustainable hegemony than previous empires had obtained, because it relied less on war, created less resentment, and encouraged growth and stability by "promoting democracy [and] interdependence" rather than suppressing local initiative.⁸¹⁵ While the United States still relied to some extent on formal annexations (Puerto Rico, Hawaii, the Philippines), the primacy of informal expansion undoubtedly contributed to the longevity of the American domination over the continent. The political capital the United States was able to draw from its informal strategy and preservation of local independence was quite evident when comparing it with Spain's actions in Cuba. The Spaniards practiced heavy-handed colonial rule, not hesitating to use violence to establish Spain's domination. As Spykman explains:

Spain systematically denied her colonies opportunity for self-rule. She granted a very limited degree of municipal self-government through participation in town councils, but important affairs were all regulated from the home country by an absolute monarchy. Government posts were reserved for officials sent out from Europe, and no one born in America had an opportunity to display his administrative talent. Spanish rule offered no opportunity to learn the difficult art of self-government.⁸¹⁶

In contrast, the United States' informal influence was much more appealing to the populations of Latin America. In fact, the U.S. public often sided with insurgents seeking to loosen the Spanish rule, as was the case with Cuba.

⁸¹⁵ Bruce Russett and John R. Oneal, *Triangulating Peace: Democracy, Interdependence, and International Organizations* (New York: W.W. Norton & Company, 2001), 300-302.

⁸¹⁶ Spykman 226.

Conclusion

Thus, the United States followed a unique, innovative path to hegemony. Despite facing a number of powerful but extra-systemic great powers, it succeeded without engaging in the traditional, conquest-based expansion, instead gaining control with a subtler, informal strategy based on economic and political exchange with its targets and occasional use of coercive power to safeguard their internal stability and guarantee their integrity against other great powers' intrusion attempts. The United States realized its hegemonic bid by taking advantage of a number of unique characteristics and achievements. On the one hand, its highly favorable geographic position as the sole great power from the region, isolated from its rivals by two oceans, helped it market itself as the ideal protector of the region's secondary powers. On the other hand, its two-phased growth, focused first on becoming an economic powerhouse and the trade partner of the region's secondary powers in the aftermath of the Civil War, and then from the 1880s-onward on rapidly developing massive military and power-projection capabilities, gave its spectacular rise a benign feel that both made it attractive to its targets and led the other great powers to underestimate its intentions.

The rising hegemon was tremendously helped by the other states' inability to mount a collective balancing effort. The potential balancers of the region, all secondary powers, were constantly distracted by their own conflicts and bandwagoned en masse with the United States. The European great powers suffered even greater dissent and internecine conflicts, worsening at the turn of the century and as World War I drew near. Their attention to the growing U.S. influence over the Western Hemisphere was side-tracked by those intra-European conflicts, as well as by their quest and competition for colonies in other faraway regions. They each passed the buck to each other instead of balancing the United States, and a few even chose to

accommodate the United States or considered doing so. In the end, the United States' unorthodox choices paid off, and it successfully became the first and only regional hegemon in modern history.

Conclusion

This chapter answers three main questions. First, what factors are most important in understanding the few historical instances where balance of power failed? The first section provides a comparative analysis of the factors of balance of power failure across the four cases and identifies the three principal variables responsible for the success of rising hegemons. It also examines a number of short research probes where the balance of power was restored to validate the case studies' findings and confirm that the factors identified as the main causes of balance of power failure are also key in explaining a variation in the outcome. The section then offers a few suggestions as to why hegemony may have been so infrequent historically. Why were the factors identified by this thesis as the prime causes of hegemony salient in a few cases but not in the numerous other cases where rising hegemons failed?

Second, do the causes of balance of power failure operate differently in the modern world than they did in ancient times? This section asks why results in the U.S. case appear somewhat unique and inquires whether the causal path to balance of power failure may have evolved overtime. It then explores why there has only been one instance of regional hegemony in the modern world.

Third, what do the findings of this thesis suggest about the world's current configuration of power? The final section opens the debate by discussing the United States' unusual status as the world's sole superpower and by questioning its position with regard to balance of power failure, and then concludes by offering predictions concerning the future of American unipolarity.

1. Findings

a. What the Cases Tell Us

The qualitative analysis of each case enables an assessment of the causal values of the twenty-one factors of balance of power failure identified in the theoretical framework of Chapter 2. Table 1 below summarizes the causal impact of each variable in triggering the outcome of balance of power failure in each of the four cases. Each factor is given a score based on the frequency with which it is observed in a case and the prominence with which it is featured in the historical narrative of the case. For instance, how often were Mongol victories over an adversary due to their innovative military tactics? Or, how many potential balancers engaged in profit-driven bandwagoning during the United States' rise? Factors that are not easily subject to a frequency assessment, mainly the rising hegemon's non-military skills and characteristics, are judged based on their significance in historical perspective. How ground-breaking was Rome's decentralized administration or citizenship policy, for example, and how much did they contrast with the practices of its contemporary rivals and grant Rome a more efficient tool of control?

Factors that were absent altogether from a case were given the score of 0. Marginal factors, that were observed only in a few instances in a case and that, if absent, would not have affected the outcome, were scored 1. Secondary factors, scored 2, were factors that were observed repeatedly in a case but were not always present or prominent. They occurred in some instances but not others. Primary factors, scored 3, were factors that stood out as omnipresent. For example, if only a few secondary powers bandwagoned (or failed to join a balancing coalition due to lack of trust, or misperceived the rising hegemon's intentions, etc.), the score would be 1. If a number of secondary powers and some great powers bandwagoned, the score would be 2. If a majority of secondary powers and a majority of great powers bandwagoned, the

score would be 3. No factor was sufficient to singlehandedly engender the outcome, not even primary factors. In each case there was more than one primary factor at play.

Table 9.1: Variable Scores in Four Cases of Balance of Power Failure

Variable \ Case	Mongols	Rome	Qin	U.S.A.	Total/ Variable Strength
1. COMMUNICATION PROBLEMS					
1.1 Physical Difficulty	0	0	0	0	0
1.2 Misperception	1	1	1	2	5
1.3 Deception	3	2	2	1	8
2. COLLECTIVE INACTION					
2.1 (Indir.) Communication Problems	2	1	1	0	4
2.2 Lack of Trust	3	3	3	3	12
2.3 Lack of Sufficient Interest	0	0	0	3	3
2.4 Buckpassing	0	0	0	3	3
3. LAGGARD BALANCING					
3.1 (Indir.) Communication Problems	1	1	1	0	3
3.2 Lack of Trust	2	3	3	0	8
3.3 Lack of Sufficient Interest	0	0	0	0	0
3.4 Buckpassing	0	0	0	0	0
4. BANDWAGONING					
4.1 Fear-Based	0	1	1	0	2
4.2 Profit-Driven	1	1	2	2	6
4.3 Combination	0	2	0	0	2
5. MILITARY ACHIEVEMENTS					
5.1 Weaponry/Technology	2	2	1	2	7
5.2 Strategy/Tactics	3	2	1	2	8
5.3 Organization	3	3	3	1	10
6. NON-MIL. SKILLS & CHARACT.					
6.1 Geopolitical Advantage	0	1	3	3	7
6.2 Economic Advantage	2	2	2	2	8
6.3 Political/Admin. Advantage	3	3	2	2	10
6.4 Social Advantage	2	2	0	1	5

Variable scores and levels of causal importance:

- 0 = Absent
- 1 = Marginal
- 2 = Secondary
- 3 = Primary

The scores allow for comparison both between variables and between cases. By comparing how each variable fares across the four cases, we can clearly establish the causal path leading to hegemonic success, or balance of power failure. Communication problems (IV1), which can occur as a result of both the potential balancers' or the rising hegemon's actions, have a direct impact on internal balancing and individual balancing decisions, but not in the way one might have thought. A central finding of this study is that contrary to conventional wisdom, physical communication impediments (IV1.1) like primitive technology or poor diplomatic networks do not prevent the potential balancers from acquiring critical information about the rising hegemon and its growing power. Even when communication tools were archaic, information about the aspiring hegemon always found its way to even the most remote potential balancers in time to organize a balancing effort. In ancient cases where information did not circulate as freely and rapidly as in modern times, potential balancers still knew about the growing power of the rising hegemon and could thus have balanced. While modern communication technology might allow for earlier appraisal of the rising hegemon and hence a faster response, the slower pace of travel and deployment in the ancient world made up for the difference by giving ancient balancers additional reaction time.

The potential balancers' misperception of the rising hegemon's intentions (IV1.2), while being present in each case, similarly failed to play a larger causal role. Misperception generally remained isolated to a few actors and therefore did not determine the failure of balance of power. Interestingly, the only case where misperception was more widespread and may have had a larger causal impact was the United States. Once more, conventional wisdom proves wrong. One would assume that in the modern world, thanks to the prevalence of diplomatic exchange channels and continuous communication at both the private and public levels due to better

technology, intentions would be better publicized and misperception would thus be minimized in comparison with the ancient world, but that was clearly not the case. The mechanisms of misperception remain eminently psychological and repeatedly evade material evidence.

Deceptive manipulation by the rising hegemon (IV1.3) is the communication variable that had the most impact on balance of power outcomes. While it was a secondary factor in two cases and a primary factor in one case, however, it was only marginal in the fourth case, making deception an important yet overall not critical causal factor. Interestingly, deception was at least a secondary factor in all ancient cases but was barely present in the United States case, and used only against minor targets. One might suspect that the United States' informal expansion strategy, based on influence rather than force and on coaxing targets to bandwagon, required it to uphold a benevolent reputation, which blatant deception attempts might tarnish, particularly given the accrued transparency of modern international interaction. The Mongols, by contrast, who relied heavily on manipulation and for whom psychological warfare played a critical causal role, were entirely unaffected by subtleties of reputation issues and in fact thrived with a reputation of cruelty and barbarism since they used that reputation to deceive enemies into surrendering without a fight. They had no incentive to hold back on tricking their targets.

Overall, communication problems (IV1) thus appear to be a second-rank causal variable in the outcome. While they did contribute to creating balance of power failure in all cases, they were more potent in some cases than others and individual communication factors had a different impact in each case. Material communication aspects—the state of technology and networks—were surprisingly irrelevant, while cognitive and psychological aspects, whether or not prompted by the rising hegemon, were more essential in disrupting the potential balancers' response. This study therefore substantiates other works in the international relations literature that highlights

the importance of perceptual assessments of a situation and of other actors, which matter beyond the individual reaction because they "are common to diverse kinds of people" and can therefore "account for patterns of interaction," as Robert Jervis observes.⁸¹⁷

The potential balancers, because they are the opponents and competitors that stand in the way of the aspiring hegemon, naturally share a large part of the responsibility for the success or failure of a hegemonic bid. Their actions—whether they balance or bandwagon, how and when they balance, and how much effort they put into pushing back the aspiring hegemon, can either help, undermine, or rout the rising hegemon, and thus determine whether balance of power succeeds or is overcome by the rising hegemon. The results from the four cases confirm that the potential balancers' mistakes—collective inaction (IV2), laggard balancing (IV3), and bandwagoning (IV4)—play a large causal role in producing balance of power failure, but they also show that some mistakes are more consequential than others.

External balancing is often necessary to stop a rising hegemon, for a multitude of reasons. Individual states can be unwilling to bear the cost of stopping a stronger power by themselves, for instance, or they may lack the resources and capabilities to sustain their own balancing effort, particularly in systems with numerous secondary powers and few or distant great powers. Moreover, as the rising hegemon grows in power, the likelihood of successful internal balancing declines. When the rising hegemon reaches near or full unipolarity, external balancing becomes the only option for balancers to stop its rise. Although collaborative balancing can thus be essential to preventing hegemony, collective inaction (IV2), or the total absence of a joint balancing effort, was prevalent in all four cases and therefore stands out as a major cause of balance of power failure. As one can expect, indirect communication problems (IV2.1), which

⁸¹⁷ Robert Jervis, *Perception and Misperception in International Politics* (Princeton, NJ: Princeton University Press, 1976), 3, 10-15.

are communication problems pertaining specifically to alliance arrangement and collaboration between potential balancers, were present since individual states experienced direct communication problems hampering their own balancing decisions (IV1). But indirect miscommunication, just like direct communication problems, remained relatively marginal. It was understandably more common in the Mongol case, where individual states also encountered more communication issues.

Collective inaction due to lack of trust (IV2.2) turns out to have a much more substantial causal impact. In fact, it is singlehandedly the most potent, as well as the most consistent, source of balance of power failure in the study, scoring at the primary level in all four cases. It is the only variable to reach the maximum mark in each case. We can therefore conclude that lack of trust-induced collective inaction is the single most important cause of balance of power failure. In all four cases there were repeated incidents in which potential balancers that could have formed a countervailing alliance failed to do so because they were too preoccupied with rivalries amongst each other, short-term gains, and mutual suspicion to agree to risk cooperating against the larger threat of the rising hegemon. The fundamentally anarchic nature of international relations and the resulting unreliability of international commitments is a force that remained constant throughout the cases and throughout time, and that weighed heavily in the potential balancers' decisions. Anarchy breeds mistrust, which in turn, if great enough, produces balance of power failure. In each case an overwhelming majority of potential balancers refused to cooperate against the rising hegemon altogether, despite the massive risks that inaction entailed for their security and survival, because they could not trust their potential partners to adhere to their obligations and stop taking advantage of them. While other potential balancer mistakes may perhaps be corrected, lack of trust-induced collective inaction is particularly unforgiving. It

inexorably yields balance of power failure because it diverts actors and resources away from the rising hegemon and pits them against each other instead, leaving few actors, if any, to take on a balancing task that would require extensive participation to succeed.

Lack of sufficient interest- (IV2.3) and buckpassing-induced collective inaction (IV2.4), on the other hand, had an overall low causal impact on balance of power failure, as both were absent in all cases save the United States. Yet, in the U.S. case, they scored at the primary level and thus played a critical role, and one may wonder what may explain such irregular scores. While distance and the existence of other priorities and areas of interest may account for many European great powers' lack of interest in intervening against the United States and their eventual decisions to pass each other the buck, this conclusion cannot be generalized. Rome and Qin grew in relatively compact regions and thus cannot be compared, but many of the Mongols' potential balancers were based thousands of miles away from the rising hegemon—in Japan, the Southeast Asian islands, or Western Europe—and often had other interests to pursue in their own areas. Yet they showed no evidence of a lack of interest in the Mongol rise or any intention to pass the buck. Distance and busyness alone do not explain the differentiated result, but it raises the issue of the United States as a distinctive case.

In large part because of the unique salience of trust problems, collective inaction (IV2) appears overall as the central cause of balance of power failure. Collective inaction immobilizes potential balancers, particularly when the most powerful ones abort balancing efforts entirely, and thus rids the rising hegemon of its competition, leaving the path wide open for an unopposed rise.

Even when potential balancers manage to overcome their collective inaction problems and actually set forth a collective balancing attempt, they can still fail if their countervailing

effort is laggard (IV3). In other words, they balance, but too late, too little, or too weakly to stop or even slow the rising hegemon. A rather interesting finding is that unlike collective inaction, laggard balancing was a second-rank causal factor, sharing roughly the same overall causal impact as communication problems. One might expect that given the high threat posed by a rising hegemon, some cooperative balancing would be more likely to occur than no balancing at all, but the reverse is true. Laggard balancing was absent altogether in the case of the United States, meaning that the United States' potential balancers did not engage in any cooperative balancing. In the other cases laggard balancing was weak or absent except for trust-induced laggardness. Cooperative balancing was therefore so difficult to achieve in the four cases that it either failed entirely, or succumbed quickly, mainly due to trust-related problems.

Communication problems (IV3.1) hindered alliance efforts, as one can expect given the existence of individual communication issues among the potential balancers. If states' individual balancing efforts were affected by their misperception and the rising hegemon's deceptive tricks (IV1), those same issues were bound to affect alliance management. Again, however, communication problems played only a marginally disruptive role in balancing coalitions. Some clever divide-and-conquer tactics used by rising hegemons were fruitful in deceiving allies and pitting them against each other. Yet even in the ancient cases, where one would think communication issues of all sorts would severely handicap an alliance, particularly in military maneuvers, communication was not the main cause of laggard balancing.

It was trust issues, just like with collective inaction, that were in large part responsible for sabotaging existing alliance efforts (IV3.2). External balancing efforts are particularly prone to trust problems because of the inherent difficulties of sustaining cooperation in anarchy, which the rising hegemon, acting on its own, does not face. Alliance members are concerned about

cheating and being double-crossed by their partners. Allies constantly worry about division of labor within the alliance: they want to have a say in decisions and hold a prominent position, but at the same time they do not want to bear the brunt of the cost, which would put them at a disadvantage vis-à-vis their partners. Trust issues are particularly salient when the alliance includes military action, which is frequently the case when balancing a rising hegemon, because the core of states' sovereignty and their whole security is at stake if their partners balk or switch sides. States that cooperate against a rising hegemon frequently have a history of rivalry and seeking immediate gains to outdo each other, which only reinforces their mutual suspicion. Lack of trust causes allies to sit back and wait for a partner to take the next step, to act with caution and not commit their full resources or effort, and it leads to extra tensions and inefficiencies within the alliance, rendering the balancing endeavor laggard. It is not surprising, therefore, that lack of trust is a decisive type of laggard balancing, and scores at the primary level in three of the cases.

While lack of trust thus led balancing coalitions to break down or be inefficient, there was no evidence that a lack of sufficient interest (IV3.3) and buckpassing (IV3.4) undermined balancing alliances. Along with physical communication problems (IV1.1), those two factors were the only ones to be entirely absent in all four cases. If collective balancing, even laggard, is generally less likely than collective inaction, there were naturally fewer instances where a lack of sufficient interest or buckpassing could have been at play in generating laggard balancing, which may be one explanation for their conspicuous absence.

Overall laggard balancing (IV3) thus was not a primary but rather a secondary causal factor of balance of power failure. Its impact remains overshadowed by that of collective inaction, which shows that in the four cases of successful hegemony, external balancing was

particularly difficult to achieve. When cooperative efforts were successfully launched, however, they all fell apart for one reason: the prevalence of trust issues within the alliances.

Bandwagoning (IV4) is the final course of action potential balancers can choose to take that is conducive to balance of power failure. Instead of siding with the other weaker states in trying to stop the rising hegemon, they join the rising hegemon's side, subtracting their capabilities from the balancers' and adding it to the rising hegemon's growing power. The four cases show that bandwagoning was also a second-rank causal factor, just like laggard balancing and communication problems. Bandwagoning does not score higher than laggard balancing overall; however, just like communication problems, it is a variable that appears much more consistently than laggard balancing, which gives it somewhat more causal weight than laggard balancing despite remaining a secondary variable. Some form of bandwagoning was present in each case, just like communication problems; but unlike communication problems, all types of bandwagoning were present in one case or another, making bandwagoning a fairly more consistent factor than even communication problems. Bandwagoning can thus be considered the most potent of the secondary factors discussed so far.

Fear-based bandwagoning might appear to be the most likely type of bandwagoning (IV4.1). Confronted with a powerful opponent on the rise, many potential balancers, particularly weaker states whose security hinges upon a great power protector, may be scared into accommodating the rising hegemon and hoping for its leniency instead of balancing it and provoking its ire. Once more, however, this study disproves the conventional wisdom. Bandwagoning out of fear was only independently present in two cases, and even in those cases its causal impact remained marginal at best. In reality, few bandwagoners chose to join the threatening side out of fear. This result is particularly unexpected with regard to the ancient

cases, where rising hegemons widely used violence and intimidation to lure bandwagoners. It shows that most potential balancers were not oblivious to the danger posed by the rising hegemon and were aware that voluntary collaboration alone might not buy its clemency.

But while they were not willing to take the risk of cooperating with the rising hegemon in the sole hope of securing the aspiring hegemon's benevolence, potential balancers were more than ready to put their survival in peril to secure profits (IV4.2). Profit-driven bandwagoning was present in each case and was often widespread. This only reinforces the previously emphasized finding that trust issues are rampant and in anarchy states tend to favor even costly short-term gains at the expense of long term security. Indeed, the main benefit expected from bandwagoning was generally the rising hegemon's support to secure territorial or other material gains at the expense of a neighbor or immediate rival. However, despite being widespread, profit-based bandwagoning did not have an overwhelming causal impact in any of the cases because it never drew a majority of the potential balancers. In fact, except in a handful of cases, profit-driven bandwagoning concerned primarily secondary powers whose balancing power was relatively limited to begin with and who thus provided only marginal added power to the rising hegemon. Secondary powers in general remain more likely to bandwagon for profit because they possess limited capabilities to take advantage of their neighbors on their own and thus welcome any help they can get to advance against them. This finding confirms Stephen Walt's argument that "weak[er] states are more likely to bandwagon" than great powers, but highlights that they may choose their protector depending on its ability to help them secure profits rather than simply out of fear.⁸¹⁸

Combined fear-and profit-based bandwagoning (IV4.3) remained absent except in one case and thus has only marginal overall causal power, which one can expect given the low

⁸¹⁸ Walt, "Alliance Formation," 16-17.

incidence and causal impact of fear-driven bandwagoning mentioned above. The only case where it was present, and significantly so, was Rome, because the rising hegemon enticed bandwagoners with its clientela strategy linking the fear and profit incentives of bandwagoning, which attracted a number of secondary states and weakened great powers who were interested in obtaining Rome's support in taking advantage of their neighbors while at the same time being particularly concerned with remaining on friendly terms with Rome because of the fragmented layout of the region. The success of combined fear- and profit-driven bandwagoning can thus be explained by the peculiar characteristics of the Roman case.

Bandwagoning (IV4) remains in the end a potent, yet not principal, cause of balance of power failure, primarily driven by profit incentives. Among the variables associated with the potential balancers' actions, it shares a second-rank causal strength with laggard balancing (IV3), while collective inaction (IV2) is the most powerful, and thus the primary, cause of balance of power failure. Communication problems (IV1), a variable pertaining to the actions of both the potential balancers and the rising hegemon, stood out as a secondary causal factor. As the aspiring hegemons' involvement in potential balancers' communication troubles suggests, the potential balancers' actions alone are insufficient to judge the success and failure of balance of power. While the balance of power literature typically stresses only the role of balancers in upholding the balance of power, that picture is incomplete because the potential balancers control only one side of the balance of power equation. An investigation of balance of power failures can thus not claim to be thorough without assessing the performance of the aspiring hegemon, whose actions determine the other side of the balance of power equation. If the potential balancers manage to balance better than in the four cases examined in this thesis and nevertheless fail to stop the rising hegemon, it means that they likely still did not do enough or

waited too long until the rising hegemon was too powerful to rein in (IV2, 3, or 4), but it may also mean that the rising hegemon proved particularly skillful in garnering and concentrating power fast and efficiently enough to outdo them. Conversely, even if the potential balancers acted in ways that tilted the equation in favor of the rising hegemon by bandwagoning and failing to balance, as in each of the four cases in this study, a poor performance by the rising hegemon might still prevent it from surpassing the other powers and taking control of the system. It may just remain another great power and fail to achieve hegemony not because it was actively stopped but because it did not develop enough power and capabilities, military or other, to take over the system. Thus, the aspiring hegemon's own abilities can also have a crucial causal impact on the success or failure of its hegemonic bid (IV5 and IV6).

The four cases highlight that the rising hegemon's military achievements constitute a particularly potent cause of balance of power failure (IV5). Not only did the rising hegemon possess unique military skills in each of the cases, but each rising hegemon displayed all three types of military advantage, making this causal category the only one without a single 'absent' score and thus the most recurrent in the study. Of course, one might expect that a successful rising hegemon might have some kind of superior military skills. It is noteworthy, however, that each rising hegemon in the study scored on every type of military achievement, showing that successful rising hegemon's aim for military superiority across the board. It is also important to point out the respective roles of innovation and adaptation in the rising hegemon's achievement of military superiority. Contrary to what one might expect, the rising hegemon's tended to be quick learners and emulators rather than innovators. While a number of rising hegemon's in the four cases made novel military discoveries that enabled them to surpass their rivals, they mostly resorted to observing, copying, and improving upon the inventions of others. In fact, successful

rising hegemon generally adopted the most efficient military techniques and technologies available at the time to others also, but they were the first to rationalize them or systematize their use, which gave them an edge over their competitors.

Weaponry and technology (IV5.1) is surprisingly the lowest-scoring military factor. It has a definite causal impact on balance of power failure since it obtains mid-level scores in all but one case; nevertheless, the other military factors outscore it. This is a significant conclusion. Even in the ancient world, where military conquest was the primary tool of expansion for rising hegemon, weapons and technology, while important, did not constitute the most powerful tools of success in the rising hegemon's military arsenal. This may reflect the rising hegemon's tendency to emulate the technological advances of their neighbors, which suggests that the rising hegemon may have suffered from a technological disadvantage at least in the early stages of their rise. In fact, some rising hegemon even started building whole sections of their military, in particular their navies, from scratch after realizing that such capabilities would be necessary to challenge a rival that held naval supremacy, such as Carthage or 19th century Britain. In all cases the aspiring hegemon embarked on its rise as a technological underdog, but extensive investment and other military aspects made up for its material deficit.

Strategy and tactics (IV5.2) have a slightly higher causal impact than weaponry and technology, but this is due singlehandedly to the Mongols' uniquely innovative strategies and tactics since scores in the other cases are exactly the same as for weaponry and technology. This result highlights that particularly inventive strategies and tactics can make up for and help overcome a deficit in technology, as happened with the Mongols. Superior strategies are strategies that assess and prioritize targets in a way that diversifies and minimizes the rising

hegemon's military risk, while better tactics are those that give it an operational advantage over individual targets.

Military organization (IV5.3), lastly, was by far the most crucial military factor at play. It outscored both weaponry and technology, and strategy and tactics, and reached the critical level in all cases except the United States. This is an important and somewhat unexpected finding. While one might have assumed that material or even operational choices would most influence the military achievement of a rising hegemon, it turns out that it is instead the less spectacular, behind-the-scenes organizational choices that matter most. Each rising hegemon made comprehensive organizational changes to its military, introducing rationalized, streamlined structures coupled with democratized access and reinforced discipline that enhanced the effectiveness of its entire military machine from the command hierarchy to logistics and gave it a decisive edge over rivals, though the United States' organizational changes were the least dramatic.

Military achievements (IV5) thus belong to the primary causal factors of balance of power failure along with collective inaction, highlighting the importance of the rising hegemon's own choices in achieving a victorious course. But military action alone is insufficient to succeed as a hegemon. Other, non-military skills and attributes (IV6) also play crucial roles in enabling a great power to become a hegemon, such as its geographic position, its ability to generate economic growth, its state-building skills and treatment of newly acquired members. While the military can obtain initial control of a vast system, for instance, it is not meant to exercise that control indefinitely and unless a rising hegemon possesses a bottomless reservoir of military force it cannot keep all its targets under tight military supervision at the same time. To move on to the next target, a rising hegemon needs to create a rule that is accepted and internalized in the

newly acquired territories. It must also develop a particularly powerful economic system to be able to fuel its military efforts, and its position in the system can also greatly facilitate its success. This study determined that the rising hegemon's non-military skills and characteristics were in fact of equal, if not more important, causal impact than its military skills, and thus constituted a third primary cause of balance of power failure.

Geopolitical advantages (IV6.1), first, including the proximity to targets, and security and richness of the home base, had a mixed impact on balance of power failure. In two of the cases geographic location in particular was a decisive causal factor—without their unique position, Qin and the United States would likely not have become hegemon. But geography is a pre-existing rather than acquired characteristic, and its absence or limited presence in the two other cases did not prevent the rising hegemon from succeeding. The Mongols' and Romans' lack of geopolitical advantage did not particularly hinder their rise. On the contrary, it forced them to develop efficient long-range transportation techniques to accommodate their distance from the targets—naval power and roads for the Romans and light-weight but self-sufficient steppe riding and post stations for the Mongols—which became their best tools of supremacy. It is interesting to note, moreover, that all rising hegemon except perhaps Qin developed considerable naval capabilities during their rise. It begs the question, must one be or become a naval power to succeed as a rising hegemon? Qin, an originally landlocked state, did not develop considerable naval power and still succeeded because it sought domination over a compact, continental area with limited sea access. But the other three cases indicate that any rising hegemon aiming for a larger region cannot forsake the development of massive naval projection capabilities. Both Rome and the United States were naval powers, and while the Mongol Empire was primarily a continental power, it overcame its reluctance to sea power and acquired naval

skills. Another remarkable finding is that all rising hegemons except Rome originated in the periphery, rather than in the center, of the system, which perhaps gave them an advantage by shielding their rise and removing them from the more competitive international environment of the system's core. Both issues—the impact of naval primacy and of peripherality on the success and failure of balance of power—could be topics for further research.

Economic advantage (IV6.2) had a central causal impact in each of the four cases, making it a strong and consistent factor. The rising hegemon always became an economic powerhouse, which helped it not only fuel its rise and buttress its military capabilities, but also become attractive to its targets and later conquer or control territories. Indeed, in all cases the rising hegemon played a proactive role throughout the system by providing infrastructure and economic growth beyond its immediate boundaries, via extensive trade partnerships. It is safe to say that without their economic leverage, both domestically and externally, those rising hegemons would have found it difficult to control their respective systems.

Political and administrative advantage (IV6.3) stands out as the most important non-military skill and characteristic, scoring at the primary level in two of the cases and at the secondary level in the two others. This result is not surprising, because a significant institutional apparatus is necessary to be able to control a territory the size of an entire system. As mentioned above, while military supremacy allows a rising hegemon to secure territory in the first place, it is its state-building efforts that enable it to retain the territory and sustain its rule. Without a superior political and administrative plan to oversee the extensive territory, then, the rising hegemon will not succeed in establishing its hegemonic rule throughout the system. But all such institutional plans are not equal, and their effectiveness at not only controlling the territory and

population but also generating the allegiance of the governed has a direct effect on the longevity of the hegemonic rule, as the Qin rulers found out.

A related, though less crucial skill, is the rising hegemon's ability to garner a social advantage (IV6.4) throughout the system, by offering unique social goods not previously available or not available among its rivals, giving it a reputational edge over the other powers of the system and winning over the adherence and loyalty of populations under its control. This factor was absent in the Qin case, which also helps explain the advent of a rebellion and rapid downfall of a hegemon that lacked the social bases of control. The rising social advantage was a secondary factor in two cases and a marginal factor in the remaining case, giving social skills moderate causal power across cases.

Non-military skills and characteristics (IV6) thus have an overall crucial causal impact on balance of power failure and hegemonic success. The rising hegemons in the study developed significant to considerable advantages geographically, economically, politically, and socially, and generally combined at least three types of non-military advantage. We can conclude that there are three primary overall causes of balance of power failure: the collective inaction of the potential balancers, the rising hegemon's superior military achievements, and its non-military skills and characteristics. The three single most important factors are lack of trust-induced collective inaction, the rising hegemon's improvements in military organization, and its political and administrative advantage. Three secondary causes may also contribute to generating a balance of power failure, though less directly: communication problems, particularly the rising hegemon's deliberate deception, and the potential balancers' bandwagoning and mistrust-related laggard balancing.

Table 9.2: Individual Factors in Declining Order of Causal Impact

SCORE	FACTOR	RESPONSIBILITY
12	Lack of Trust-related Collective Inaction	Potential balancers
10	Military Organization	Rising hegemon
	Political/Administrative Advantage	Rising hegemon
8	Deception	Rising hegemon
	Lack of Trust-related Laggard Balancing	Potential balancers
	Economic Advantage	Rising hegemon
	Military Strategy/Tactics	Rising hegemon
7	Military Weaponry/Technology	Rising hegemon
	Geopolitical Advantage	Rising hegemon
6	Profit-driven Bandwagoning	Potential balancers
5	Misperception	Potential balancers
	Social Advantage	Rising hegemon
4	Communication-based Collective Inaction	Both
3	Lack of Sufficient Interest-based Collective Inaction	Potential balancers
	Buckpassing-based Collective Inaction	Potential balancers
	Communication-based Laggard Balancing	Both
2	Fear-based Bandwagoning	Potential balancers
	Combined Bandwagoning	Potential balancers
0	Physical Communication Problems	Potential balancers
	Lack of Sufficient Interest-based Laggard Balancing	Potential balancers
	Buckpassing-based Laggard Balancing	Potential balancers

One of the notable contributions of this study is to show that the mishaps of the potential balancers are not entirely to blame for the occurrence of balance of power failure, but that the proactive role of the rising hegemon in generating and exploiting both military and non-military advantages is also critical. Although there are slightly more factors pertaining to the potential balancers than to the rising hegemon, it is noteworthy that the rising hegemon-related factors consistently score higher than potential balancers-related factors (see table 2). The significant exception to this pattern is lack of trust-related collective inaction, the single highest scoring factor, and to some extent also lack of trust-related laggard balancing. This shows that much is in the hands of the rising hegemon, but there remains one sizeable uncertainty for any rising hegemon—will the potential balancers be able to overcome their deep-seated trust issues?

b. Research Probes

When balance of power is restored and rising hegemons fail, the variables should operate in reverse. If the three primary causes of balance of power failure are the balancers' collective inaction and the rising hegemon's unique military achievements and non-military skills and characteristics, we should observe that these three variables are also key in generating successful balances, albeit with opposite scores. In other words, when balances are reestablished, it should be primarily because balancers are able to surpass their mistrust and collective action problems and establish effective balancing coalitions, and because the rising hegemon fails to develop unique military and non-military advantages, or if it does, because the balancers are able to catch up or counteract its superior efficiency. In order to ascertain that it is the case and validate our findings, a few control cases of successful balance of power need to be examined. Such research probes serve to establish that the factors that are responsible for balance of power failures are also critical in explaining why balances succeed, and confirm that the three variables that stood out in the cases of hegemony are the correct variables.

The clearest successes of balance of power provide the best research probes because they leave the least room for possible non-related interfering variables and offer the most unambiguous variation of the dependent variable. The following set of mini-case studies will therefore provide a suitable control checklist of balance of power success: Spain under Philip II, France under Louis XIV, and France under Napoleon. Each of these aspiring hegemons was stopped before achieving hegemonic status. These cases represent the classic era of European balance of power politics, when scholars agree that balance of power between great powers worked particularly well. From the 16th to the 20th century various European kings and states incessantly pursued hegemonic bids, but all were checked.

1. Spain under Philip II

One of the most noticeable European aspiring hegemon was Spain under Philip II, the Habsburg ruler of the second half of the 16th century who led Spain to the peak of its power, with control over Portugal, the Netherlands and Low Countries, Jerusalem, parts of Italy with Milan, Naples, and Sicily, parts of eastern France, the vast Spanish overseas colonial possessions in Latin America and the Caribbean, and even briefly England and Ireland while he was married to Mary I during the four first years of his rule. If one considers that his uncle, Ferdinand I, and later his cousin, Maximilian II, both Holy Roman Emperors, ruled over Austria and Germany, as well as parts of Italy and eastern Europe, then it can be said that Philip II and the Habsburgs encircled France and virtually controlled the European continent. But Philip's ambitions were stopped by a multitude of forces.

Communication problems, first, hampered not the balancers, but rather the rising hegemon (IV1). It was the balancers that successfully used deception against the rising hegemon (IV1.3), by stirring elements within the Spanish empire to rebel and then supporting their insurrections, thus artificially inflating their power and disrupting Philip's control over his territory. For instance, Ottoman emperor Selim II threw his weight behind the Moor minority in southern Spain and encouraged its open rebellion in 1568-71, which turned out to be particularly damaging for Philip's rule as it increased the fragmentation of his regime and further undermined an already frail economy. Selim helped send Turkish and Berber fighters to support the Moorish insurgents, and spread rumors that he was planning a diversionary attack on Malta, which was quickly picked up by Spanish spies. The French Huguenots similarly decided to help the Moors in an effort to undermine Philip's rule, and in the mid-1570s plotted an attack on their side, even

obtaining a promise of naval support from the Ottoman emperor. Although the attack never took place, its possibility threw Spanish leaders into disarray. In addition to outside disruption attempts, Philip's regime also suffered from domestic destabilization attempts, with never-ending palace intrigues, including murder plots involving the highest-ranking officials.⁸¹⁹

There does not seem to have been any laggard balancing (IV3), or any buckpassing (IV2.4) or lack of sufficient interest (IV2.3) among Spain's potential balancers. Instead, most of Europe went to war against Philip and many kings did so eagerly. Potential balancers not only did not hesitate to individually resist Philip—Britain fought Spain intermittently between 1585 and 1604 and France battled Philip from 1590 to 1598—but they displayed unusual unity and a remarkable ability to act in concert against the Habsburg king, transcending their numerous divergences and the pervasiveness of immediate gain temptations that resulted from the unstable boundaries 16th century Europe (IV2.2). The Dutch revolt, which led to the Eighty Years War and ultimately defeated the Spanish ambitions over Europe, is a remarkable example of balancing cooperation. When some Dutch provinces rebelled against Philip and sought to secede in 1568, they were able to gather support from the most unlikely of coalitions. England and Ireland, some German protestant kingdoms, Catholic France, the French Huguenots, and the Ottomans all sent military expeditions and participated in the Calvinist and Catholic Dutch efforts to boot out Philip, which were eventually successful and resulted in secession and eventual Dutch independence.⁸²⁰

Bandwagoning, while it occurred, remained relatively minor (IV4). Philip managed to organize the Holy League in 1560 to battle the Ottomans, but it was mostly composed of smaller secondary powers—the Papal states, the Habsburg Italian kingdoms of Naples and Sicily, as well

⁸¹⁹ J.H. Elliott, *Imperial Spain 1469-1716* (London: Penguin, 2nd ed., 2002), 255-60.

⁸²⁰ Geoffrey Parker, *The Grand Strategy of Philip II* (New Haven, CT: Yale University Press, 2000), 2-3, 6, Chs. 2, 4-6.

as Venice, Genoa, Tuscany, Savoy, Parma, Urbino and Malta, all placed under the lead of Philip II's half-brother—and was relatively short-lived. The League suffered a massive defeat at the hands of the Ottomans at Djerba in 1560, but was able to regroup and beat the Ottoman forces at Lepanto in 1571. It disbanded soon afterwards, however, and the Ottomans retook much of the territory that the League had deprived them of. In its war with France (1590-1598), similarly, Spain was joined by the Catholic League, a union of papal and French Catholic forces under the Duke of Guise, but that League enjoyed the support of only parts of France and lost strength in 1593 when French King Henry IV converted to Catholicism. The Catholic League was defeated by Henry IV two years later, leaving Spain the sole belligerent against France and forcing it to eventually withdraw in 1598. Thus, the few bandwagoners that teamed with Spain did little to help the Spanish cause and in fact contributed virtually no power to the aspiring hegemon's side. Because they were minor actors, their defection from the balancers' side did not fundamentally affect the chances of the balancers either.⁸²¹

As for Spain's achievements, they remained too modest to ultimately supersede the rest of Europe. Spain's military was powerful, but not superior (IV5). Spain had considerable maritime strength, in particular, which had helped it build its empire across the Atlantic, but it was not able to outpace its competitors in Europe and did not achieve naval supremacy there. The Ottomans remained powerful in the Eastern Mediterranean, while Britain showed significant strength in the Western Mediterranean, successfully repelling repeated assaults from the Spanish Armadas between 1588 and 1604. In addition, although a coalition of Habsburg forces, with Philip's Holy Roman Emperor uncle and cousin, would have multiplied Spain's power over the European landmass, the relations between Spain and the Holy Roman Empire were not always cordial and

⁸²¹ See Robert J. Knecht, *The French Religious Wars 1562-1598* (Wellingborough, UK: Osprey Publishings, 2002), particularly 57-60, 75-79; Henry Kamen, *Empire: How Spain Became a World Power 1492-1763* (New York: Harper Collins, 2003), 171-6.

did not permit such hopes for Philip. The Spanish military suffered greatly from Spain's financial troubles, which deprived it of crucial resources, and it was also particularly difficult to control for Philip because a large part of Spain's military power was in the hands of local aristocrats due to the political arrangement of the empire. In the end, while Philip's military might have been sufficient to expand at the expense of the indigenous peoples of Latin American and the Caribbean, it proved critically inadequate to surpass rival European great powers.⁸²²

While superior non-military assets and achievements may perhaps have partially made up for Spain's military deficiencies, Philip failed to develop those too, essentially dooming his expansionist effort (IV6). Geography did not offer him any particular advantage (IV6.1). On the contrary, his European territory was fragmented and separated by other powers, including France, and that distance and disconnection made it easier for the Dutch to rebel. And while the overseas colonies provided critical resources to a relatively poorly endowed Spanish homeland, it further contributed to segmenting Philip's possessions. The Spanish economy did not fare particularly well under Philip, either (IV6.2). The home base of Spain had limited economic prospects and was scantily populated, thus yielding insufficient revenue for Philip. In addition, because the land was largely controlled by local kings and aristocrats, Philip encountered difficulties appropriating the tax revenue. The richer regions of his kingdom, the Netherlands and the colonies, where most of the funding for military campaigns originated, were both far away and difficult to control. Moreover, Philip inherited large debts from his father and predecessor, Charles V, and incurred a significant annual deficit with costly domestic policies. Coupled with massive inflation, those problems caused considerable hardship not only on the Spanish population but also on Spain's industry and trade, which were therefore unable to support the military effort. Spanish aristocrats often favored high-end imported goods to Spanish products,

⁸²² Kamen 167-170, 173-5, 189.

and Philip's government, having trouble finding domestic sources of revenue, became increasingly dependent on resources from the colonies and on foreign debt, which soon became a significant part of the revenue. Philip's financial troubles became so dire that he was forced to declare a moratorium on Spanish debt in four occasions between 1557 and 1596.⁸²³

On top of its economic shortfalls, 16th century Spain was a real governing quandary, and Philip proved incapable of exerting sovereign control over his possessions (IV6.3). The main issue was a lack of central authority. Spain was not one monarchy but a federation of various kingdoms and fiefdoms in the hands of aristocrats, all with their own pre-existing, and often incompatible, legal authority and military forces. As a result, Philip was frequently overruled by local lords and regional assemblies, which did not hesitate to disregard his authority. For instance, the Kingdom of Aragon, located in Philip's heartland of Spain, frequently undermined Philip and even openly rebelled in 1591-2, causing a civil war. Spain was thus extremely difficult to rule. The king attempted to centralize power by appointing local representatives to enforce his will in each region, but that backfired because the representatives competed for authority with local entities and generated even further resentment and disobedience against Philip. The Dutch revolt started in large part as a response to Philip's attempt at centralizing power, and the governor he sent to the Netherlands for that purpose was a particularly acute source of discontent.⁸²⁴

Finally, Philip reinforced the social fragmentation of the country and further encouraged dissent and revolt by repressing religious minorities, which were numerous in Spain (IV6.4). The most blatant example concerns the Moors of Muslim origin, though discrimination extended to other minorities like Jews and Protestants. The repression of the Moors became problematic for

⁸²³ Elliott Ch. 8.

⁸²⁴ Geoffrey Parker, *The Dutch Revolt* (London: Penguin, Revised Edition, 1990), 40-45; Elliott 275-84; Kamen 180.

Philip because they had a large presence in Spain, particularly in the south. Most were forcefully converted under threat of expulsion in the early- to mid-16th century, and Philip continued the trend by ordering in 1567 that they abandon their Muslim names, garbs, and dialects, and send their children to Catholic education institutions. After Philip defeated a Moor uprising in 1571, the Moors resettled throughout Spain but continued to be consistently persecuted by the Inquisition.⁸²⁵ Thus, in the case of Philip II the factors of balance of power failure took on exact reverse values, substantiating the findings of this study. Philip's hegemonic bid failed for the same three primary factors that enabled other rising hegemony to succeed. First, the balancers were able to overcome their mistrust and collective action problems despite vast differences and pre-existing conflicts, and collectively balanced against Philip. Second, Philip did not develop superior military skills to overcome the balancers, and third, he was unable to generate non-military advantages to reinforce his position. Instead, he suffered from a severe lack of authority over his own land and aristocrats, and was also hindered by economic shortcomings and social fragmentation. Bandwagoning and communication issues played more minor roles.

2. France under Louis XIV

The rise and fall of absolutist France under Louis XIV in the following century further reinforces this verdict. Louis XIV began governing France in 1661 and, between the mid-1660s and 1715, expanded the country's boundaries in all directions to reach what he considered the 'natural frontiers' of France, repeatedly stretching into the Spanish Netherlands, northern Italy, the Rhine region, and parts of Spain. After a number of European-wide conflicts—two short wars, the War of Devolution (1667-8) and the War of the Reunions (1683-4), and three major conflagrations, the Dutch War (1672-8), the Nine Years' War (1688-97) and the War of the

⁸²⁵ See Richard Fletcher, *Moorish Spain* (Berkeley: University of California Press, 2003); Elliott, Ch. 6.

Spanish Succession (1701-1713)—French ambitions were checked and Louis XIV was forced to retreat to his original boundaries in the 1713 Treaty of Utrecht.⁸²⁶ What were the most important factors in explaining why a balance of power was successfully restored on the European continent after Louis XIV's rise?

Communication issues were present in the form of deception attempts by the rising hegemon (IV1.3), but played only a minor role. All the European powers seemed well aware of Louis XIV's expansionist aims and thirst for glory (IV1.1), and there is no indication that they misperceived French intentions (IV1.2). French deception was mainly focused on getting some secondary powers to act in France's favor and on strategically spreading information to confuse opponents in support of France's war efforts. France under Louis XIV had "more ambassadors and agents scattered around Europe than any other country." The French king also sent "special envoys" and "strange characters" abroad such as astrologers, clergymen, and scholars, and did not hesitate to have them spur "diplomatic intrigues" or even "bribe influential men such as English members of Parliament or Swedish senators" to sway them to his side. In fact, "the French Foreign Office was large, containing experts at ciphers, propagandists and archivists." However, the number of potential balancers France was able to influence with its diplomatic maneuvering remained limited.⁸²⁷

There were no instances of laggard balancing (IV3) during France's rise. Instead, the balancers' reaction was prompt and full-fledged every time Louis attempted to expand. Moreover, the potential balancers were able to overcome their difficulties to cooperate and avoid collective inaction (IV2). There is no evidence that any of the potential balancers were uninterested in holding France back (IV2.3), or that some sought to pass the buck and free ride

⁸²⁶ John Albert Lynn, *The Wars of Louis XIV 1667-1714* (Harlow, UK: Addison Wesley Longman, 1999), 6.

⁸²⁷ Maurice Ashley, *The Age of Absolutism 1648-1775* (Springfield, MA: Merriam Company Publishers, 1974), 77.

on others' balancing efforts (IV2.4). On the contrary, all of the great powers and most secondary powers appeared more than willing to participate in stopping France's rise. French ruses occasionally managed to destabilize a balancing coalition (IV2.1). For example, Louis used complex diplomatic tricks to damage the Netherlands' international support network during the Dutch War. But because those cases remained rare and the destabilization was temporary, French deception did not impact the overall success of the balancers.⁸²⁸ Most importantly, even though the balancers had contrasting backgrounds and often a history of conflict, they nevertheless repeatedly managed to supersede their distrust and the appeal of immediate gain to form sweeping alliances against France (IV2.2).

When Louis XIV attacked the German principalities of Cologne and Rhineland-Palatinate in 1688, initiating the Nine Year's War, he anticipated an easy victory because he did not think the two secondary powers would be able to resist. However, he did not count on the German princes' ability to raise a coalition. The Grand Alliance, as it was called, brought together all of the European great powers and some secondary powers in opposition to France, even though several had been hostile or had been reluctant to participate in joint continental actions in the recent past. The Holy Roman Empire, England, Scotland, Spain, the Dutch Republic, Piedmont-Savoy, and Sweden all participated. By 1697 the allies had surrounded France and blocked its advance in the Spanish Netherlands, the Rhineland, and Savoy, forcing Louis to return most of the territory he had conquered. Initially the allies did not possess a strong, united army, and encountered a number of disagreements. For example, the Dutch were suspicious of the Britons, who had sided against them just ten years earlier in the Dutch Wars.⁸²⁹ However, "the aggressive

⁸²⁸ Ibid.

⁸²⁹ Arthur Hassall, "The Foreign Policy of Louis XIV (1661-97)," in Lord Acton, A.W. Ward, G.W. Prothero, and Stanley Leathes, eds., *The Cambridge Modern History Vol. 5: The Age of Louis XIV* (New York: The MacMillan Company, 1908), 57-8.

policy of the... [French] king, backed by his formidable forces, had caused a fright" and led the Allies to quickly overcome their early difficulties.⁸³⁰

A few years later, as the European powers proved unable to reach an agreement on the partition of the declining Spanish Empire following the death of its ruler without an heir, the British, wary of Louis XIV's expansionist plans, convened another Grand Alliance with the Dutch and Austrians, which was sealed by treaty in 1701. Hostilities with France resumed soon afterwards in the Spanish Netherlands and in northern Italy, beginning the War of the Spanish Succession. The alliance was again shaky at first. The Swedes were simultaneously at war against the Danes, Poles, and Russians, so that the Grand Alliance could not count on any help from northern Europe. However, the allies quickly made up for that absence. Most of the German principalities joined the alliance, and "an impressive accumulation of strategic resources was steadily built up," fueled by the Netherlands and Britain, who pitched in financially to support the secondary states that had joined. "The wealth of the English kingdom and the Dutch Republic, deriving from commerce, industry and banking ... made them the paymasters of the Grand Alliance which was able finally to bring the Bourbons to heel." Despite their diverse background, the allies found ways to coordinate their actions, and joint campaigns became a success. For instance, in 1702 the Dutch and Britons won a critical sea battle at Vigo Bay off the coast of Spain together, and Britain and the kingdom of Savoy inflicted a massive defeat on France at Blenheim in southern Germany in 1704, a turning point in the war. The allies' ability to surmount their distrust and collective action problems and balance in common was therefore a primary cause of France's downfall.⁸³¹

⁸³⁰ Ashley 81, 83, 86.

⁸³¹ Quotes are from Ashley 91-3. See also John B. Wolfe, *Louis XIV* (New York: W.W. Norton & Company, 1968), 509-554, and C.T. Atkinson, "The War of the Spanish Succession," in Lord Acton, A.W. Ward, G.W. Prothero, and

A few potential balancers bandwagoned with France (IV4), but they were mostly secondary or weakened powers and none supported France for the entire period of its rise. During the Nine Years' War France was on its own, for instance, as Louis XIV was unable to secure a single ally. During the War of the Spanish Succession, Spain joined the French side, undoubtedly because the will of last Spanish king Charles II had formally put a descendent of Louis XIV on the Spanish throne. But Spain was not in a position to supply significant amounts of troops or military equipment, and its contributions to the French war effort were minimal. France was able to persuade the Portuguese to join its side too, which allowed Louis to use a number of Portuguese naval bases in the Mediterranean and Atlantic, from which the French launched raids to hurt the allied commerce. Portuguese bandwagoning was short-lived, however. After the 1702 allied naval victory at Vigo Bay, the Portuguese king renounced his treaty with France and joined the Grand Alliance instead. France initially secured the support of the Duke of Savoy, but by 1703 he too defected and joined the Grand Alliance. The German electorate of Bavaria also bandwagoned with France, motivated by the potential territorial profit it could extract from the Habsburg, but it was so badly defeated in 1704 at Blenheim that it was effectively knocked out of the war. The French finally lent their support to Hungarian nationalists, in an effort to open a second front against the Holy Roman Empire and distract the Emperor from the war with France, but the Hungarians did not actively contribute to Louis's own effort. In the end, bandwagoning occurred, but since it was mostly temporary and attracted mainly weaker actors, it did not dramatically help or hinder either side and therefore played only a secondary role.⁸³²

Stanley Leathes, eds., *The Cambridge Modern History Vol. 5: The Age of Louis XIV* (New York: The MacMillan Company, 1908), 401-436.

⁸³² Atkinson 402-404, 407, 411-12; Ashley 82, 92-3; Lynn 13.

While the balancers overcame their difficulties to cooperate and established a lasting, well-coordinated countervailing alliance, France was in contrast an underachieving rising hegemon. Louis XIV failed to develop the skills necessary to use and extract France's resources more efficiently than its rivals and thus failed to outdistance them. On the military side (IV5), although France had the largest and best trained army in Europe at the time of the Nine Years' War, it was unable to retain that advantage. Britain and the Netherlands, even though they were maritime powers, used their trade surpluses to finance large armies so that by the time of the War of the Spanish Succession they and Austria had almost matched France's land power.⁸³³ France's special skill was in its fortification system, designed by military engineer Vauban to secure the French borders (IV5.2). However, the allies eventually found ways to conquer Vauban's fortresses, thus denying France another advantage. In an era dominated by attrition and siege warfare, Vauban also came up with a number of critical innovations in offensive siege tactics, such as the use of ricochet fire, but most remained underused due to the French artillery's skepticism.⁸³⁴ In addition, France suffered greatly from its sub-par naval capabilities (IV5.1). Although Jean-Baptiste Colbert, Louis's Comptroller of Finances and Minister of the Navy, expanded French naval power in the 1670s, the program was abandoned ten years later due to its cost and the French navy was never able to rival the navies of Britain and the Netherlands.⁸³⁵ In the end, the French were severely "handicapped by their inferiority at sea, particularly in the Mediterranean," during the War of the Spanish Succession.⁸³⁶

France's non-military skills and characteristics were also insufficient to hoist it ahead of the balancers. The country was geographically endowed, but so were its rivals (IV6.1). France

⁸³³ Lynn 266-75, Ashley 82.

⁸³⁴ Jamel Ostwald, *Vauban Under Siege: Engineering Efficiency and Martial Vigor in the War of the Spanish Succession* (Leiden, NL: Brill, 2007), 31-4, 64-67.

⁸³⁵ Lynn 14.

⁸³⁶ Ashley 77, 93.

benefited from a rich and fertile soil and therefore possessed a strong agricultural basis with which to sustain its war efforts, but the Netherlands and Britain generated even more wealth thanks to their greater maritime exposure, which allowed them to engage in extensive trade. France's position was central in the region and thus allowed it easy access to the north, south, and east, yet it also left the country surrounded by rivals, with the Habsburgs in three directions (Austria, the Netherlands, and for some time Spain) and Britain in the fourth.⁸³⁷

France's economic achievements were similarly average (IV6.2). Unlike some of its neighbors, the French economy was a legacy from the Middle Ages, based almost exclusively on agriculture and therefore dependent on unreliable harvests. In 1696, for example, an economic crisis fueled by a number of bad harvests precipitated France's defeat in the Nine Years' War.⁸³⁸ In addition, there were virtually no improvements in agricultural technology during Louis's reign. The French economy failed to modernize and was thus unable to come to par with the booming economies of its rivals, particularly Britain and the Netherlands, which prided themselves with large trade profits and technological innovation. In France by contrast, commerce and industry constituted only a small part of the state income. Comptroller of Finance Colbert attempted to improve the state's economy by reducing the national debt, increasing the land tax base, promoting exports, and subsidizing the domestic production. While his reforms helped provisionally build up the country's resources for war, mainly by increasing the burden on the peasantry and lower classes, they failed to regenerate the state's financial system and produce long-term economic growth because they remained too rigid and did not renounce the full tax exemption granted to the wealthiest, the nobility and the Church. To encourage commercial exchange, Colbert created trading companies, but merchants refused to join them because they

⁸³⁷ Lynn 14-15; Ashley 76.

⁸³⁸ Hassall 61.

were too tightly regulated by the government. Thus, although France was not poor during Louis's reign, it was not at the forefront of the European economies, which remained dominated by the Dutch and British.⁸³⁹

France's political and social system was even more archaic, hindered by profound disunity and remains of feudalism (IV6.3). The rigidity of the political system, an absolute monarchy with all authority and control of every administrative office exclusively concentrated in the hands of Louis XIV, fueled a number of popular revolts, which in turn further ossified Louis's grip on power. France's fifteen million peasants bore the brunt of the tax and production burdens and suffered widespread abuse, contributing to the popular grievances, which erupted notably in the midst of the Dutch War. Louis XIV was worried not only about such uprisings, but also about the treachery of the nobility, as palace intrigues and power conspiracies were numerous. As a result Louis excluded all princes of the blood from governmental positions, as well as his immediate family, including his mother, who had gained valuable experience as a regent for nearly two decades, his only brother, and his only son. Instead of surrounding himself with the most qualified statesmen, Louis retained the same three loyal ministers for the main advisory positions, war, foreign policy and state finances. In addition, he displaced the entire court from the rowdy capital to his father's former hunting lodge at Versailles in 1682. The move was meant to protect him from incidents in Paris while at the same time allowing him to keep the higher nobility in check, but it also contributed to further isolating him.⁸⁴⁰

Finally, Louis XIV increased the already acute fragmentation of French society by reigniting the country's religious rifts (IV6.4), which had started to heal after his grandfather

⁸³⁹ A.J. Grant, "The Government of Louis XIV (1661-1715)," in Lord Acton, A.W. Ward, G.W. Prothero, and Stanley Leathes, eds., *The Cambridge Modern History Vol. 5: The Age of Louis XIV* (New York: The MacMillan Company, 1908), 8-14; Ashley 71, 75-6, 78-9.

⁸⁴⁰ Lynn 15-19; Ashley 66-7, 70-1, 80. See also Roger Mettam, *Power and Faction in Louis XIV's France* (Oxford: Blackwell Publishers, 1988).

signed the 1598 Edict of Nantes conferring French Huguenots religious and civil rights. Presenting himself as the Catholic 'savior of the West,' Louis revoked the Edict and outlawed Protestantism in 1685, and resumed the forced conversion and persecution of French Protestants. This not only led to a number of domestic uprisings that interfered with France's wars abroad, but also resulted in sending hundreds of thousands of Protestants, many of them merchants, artisans, and skilled laborers, into exile to the Netherlands and England where they reinforced the power base of France's opponents.⁸⁴¹ As Maurice Ashley stresses, Louis's "failure to create political unity, to destroy the galling remains of feudalism or to accept religious toleration paved the way ultimately for revolution."⁸⁴² In sum, the rise and demise of France under Louis XIV confirm that the three factors that appeared most prominent in enabling hegemony are the same that prevent it from succeeding when they are reversed. Balance of power was restored against absolutist France primarily because the balancers overcame their distrust and collective action problems and built sweeping balancing coalitions; France's military skills failed to surpass those of its rivals; and France's non-military achievements, particularly in the political realm, remained mediocre. Bandwagoning and communication issues were secondary factors.

3. Napoleonic France

A third shadow case, the rise of Napoleonic France less than a century later, further substantiates these findings. Napoleon Bonaparte rose to prominence as a general during the French Revolution, then became consul through a coup in 1799 and eventually crowned himself emperor in 1804. Napoleon led France through successive offensive campaigns and expanded the country well beyond the frontiers left by Louis XIV in 1715, which had remained virtually

⁸⁴¹ Warren C. Scoville, *The Persecution of Huguenots and French Economic Development 1680-1720* (Berkeley, CA: University of California Press, 1960), 98-99; Lynn 15.

⁸⁴² Ashley 80.

untouched until the French Revolution. Napoleon ventured as far as Egypt, Syria and Galilee. At its height between 1810 and 1812, the Napoleonic empire formally included all territory west of the Rhine, the Low Countries, northwestern Italy with Piedmont and Liguria, as well as the Papal states and Illyria. In addition, Napoleon increased his control informally by acquiring a number of puppet states, often by installing his relatives on their thrones. Spain and parts of Portugal, Naples and the remainder of Italy, Switzerland, the Confederation of the Rhine and all the German states, and Poland all became French proxies. By 1811-1812, after defeats forced Austria, Prussia, Denmark and Sweden to ally with France, continental Europe was almost entirely under Napoleon's direct influence.⁸⁴³ Nevertheless, the balance of power was reestablished just three years later.

The Napoleonic rise offers no evidence of physical barriers depriving the potential balancers of the correct information about France's growth (IV1.1), nor is there any indication that potential balancers may have misperceived Napoleon's imperialistic intentions (IV1.2). Communication issues were present in the form of deception, however (IV1.3). Napoleon extensively used espionage and ruses to destabilize his opponents both in battle and in diplomacy. The impact of his deceptive efforts remained limited, though, because the potential balancers were just as shrewd and managed to deceive Napoleon in return in numerous instances. For example, in 1812 Prussian military leaders concluded a secret alliance with Russia against France while at the same time still feigning to be on France's side on the battlefield. The same year Austrian foreign minister Metternich concocted intricate diplomatic tricks to retrieve the 30,000 troops Austria had been forced to supply to France, rearm beyond the ceiling imposed by Napoleon, and switch sides to join the balancing coalition, "all without bringing down upon her

⁸⁴³ Edward Vose Gulick, *Europe's Classical Balance of Power: A Case History of the Theory and Practice of One of the Great Concepts of European Statecraft* (Wesport, CT: Greenwood Press, 1955), 95-8.

[Austria] the wrath and arms of France until Austrian arms were adequate for defense." Metternich clearly managed to fool Napoleon, who still thought, even after Austria had secretly joined the alliance against France, that Austria would remain on France's side.⁸⁴⁴

Not only did Napoleon's potential balancers manage to coordinate their efforts and overwhelmingly joined balancing coalitions (IV2), but they also did not waver in doing so and balancing was never laggard (IV3). There is no evidence that any of the potential balancers sought to pass the buck to others (IV2.4) or lacked sufficient interest in participating in the common balancing effort (IV2.3). European monarchs first tried to jointly stop France during its early rise in the Revolutionary years, and wound up calling together seven successive coalitions. The First Coalition (1782-97) was formed by Austria and Prussia to respond to France's encroachments on Austria. In the Second Coalition (1798-1802), Russia joined Austria against France. Participation grew during the Third Coalition (1803-05), where Britain and Portugal joined Austria and Russia, but led to an ambivalent result. Britain secured a major naval victory at Trafalgar but a few months later France destroyed the combined Russian and Austrian forces at Austerlitz. The Fourth Coalition (1806-7) nonetheless resumed the fight, with England, Prussia, Russia, Saxony and Sweden, eventually reinforced by Spain and Portugal. In 1809 a Fifth Coalition was briefly formed as Austria was able to break away from its forced alliance with France resulting from Austerlitz and join Britain on the resisters' side. However, it was short-lived and France scored another victory. The balancing effort grew in size with the Sixth and most important Coalition yet (1812-13), which was comprised of Prussia, Austria, Sweden, Russia, England, Spain and Portugal. The allies succeeded in surrounding Napoleon and forced him to retreat, even taking Paris in 1814. The final, Seventh Coalition was formed after a brief resurgence of French power following Napoleon's return from his forced exile at Elba in 1815.

⁸⁴⁴ Ibid., 108-110, 112-122.

This time it included England, the now-liberated Netherlands, Prussia, Russia and Austria, and brought about Napoleon's final defeat at Waterloo.⁸⁴⁵

What was unique about the alliances against Napoleon was not just their sheer number, but also the balancers' remarkable ability to overcome their underlying distrust of each other and transcend their numerous pre-existing conflicts to face France (IV2.2). For example, the Nordic countries were wrought with long-standing hostility, which led to flare-ups between Sweden and Prussia and between Sweden and Russia. The latter eventually went to war over Finland in 1808-09, but that did not prevent them from uniting their efforts against Napoleon barely two years later. Similarly, while France was retreating from its failed Russian campaign in 1812, the Russians and Prussians formed an alliance despite having outstanding border disputes themselves. They hastily concluded an agreement on the borders in order to be able to join forces against Napoleon. Austria and Britain also suffered from "important differences in their diplomatic positions," particularly as they both competed for influence in central and eastern Europe, and Austria was staunchly anti-Russian, yet all managed to put their hostility on hold and sign alliance treaties in 1813 to form the Sixth and victorious Coalition. Even after the coalitions were formed, the allies occasionally faced internal dissent that at times "almost visited disaster upon the coalition[s]." In 1814, for instance, distrust led the great powers to clash on the coalition's plans and priorities. "Alexander, ever skeptical of Austrian loyalty to the coalition, and Metternich, suspicious of Russian motives and fearful of Russian preponderance in European affairs, were continually at loggerheads," and their disagreements almost led to the breakdown of the alliance. However, British Foreign Secretary Castlereagh was able to defuse the crisis and consolidate the alliance. Thus, a primary cause of Napoleon's downfall was the

⁸⁴⁵ Charles Esdaile, *Napoleon's Wars: An International History 1803-1815* (New York, NY: Viking, 1997), Chs. 4-10.

balancers' success in overcoming their numerous trust issues and coordinating their opposition to France.⁸⁴⁶

Bandwagoning played a more muted role (IV4). Although Napoleon succeeded in establishing a number of puppet states and military allies, their support should not be overstated. Most provided only minimal or temporary contributions to Napoleon's power, either because they were weaker powers (Denmark-Norway, or Switzerland, for example, which suffered from a civil war) or because they were coerced into provisionally joining Napoleon's side as a result of defeat and thus were not true bandwagoners.⁸⁴⁷ A number of states wound up on the French side after suffering severe defeats—Austria after Austerlitz in 1805, Prussia after its defeats at Jena and Auerstadt in 1806, Russia after its defeat at Friedland in 1807, and Austria after its defeat at Wagram in 1809—only to switch sides again whenever their strength returned or an opportunity arose. Such states cannot be considered bandwagoners because they did not voluntarily join France but rather ended up on its side because they were strapped with harsh peace terms and had little few other choices. The example of Prussia's participation in the French war effort is telling. When Napoleon extracted Prussian troops to serve in his 1812 Russian campaign, "these troops were a reluctant portion of ... his army, reluctant because the real enemy of Prussia in that era was Napoleon himself and not Alexander, against whose soldiers the Prussians were compelled to fight." The Prussian forces, which were under French command, waited for the right moment and then secretly defected to the Russians in December 1812. Napoleon succeeded in securing a Franco-Ottoman alliance in 1806, and a Franco-Persian alliance in 1807, but the alliances never bore fruit and both the Ottomans and the Persians soon regularized their relations

⁸⁴⁶ Gulick 109-11, 149-50.

⁸⁴⁷ William H. McNeill, *The Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000* (Chicago, IL: University of Chicago Press, 1982), 199; Anton Guiland, "France and Her Tributaries 1801-3," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907), 95-7.

with at least one of France's balancers. In the end, therefore, although some bandwagoning with France occurred, its overall impact on either side remained limited.⁸⁴⁸

The crucial factor that accounted for the balancers' success, then, was their ability to surmount their trust issues and thus coordinate their balancing endeavors. Nevertheless, such efforts might have proved insufficient if France outpaced the balancers with unique innovations and efficiency. That was not the case, however. While Napoleonic France displayed unusual military skills, it failed to generate a great enough edge to reach hegemony (IV5). Although the French army, driven by the levee en masse, was "the most imposing to be seen on the Continent" throughout the Napoleonic Wars, its strength rested on numbers alone, because overall military technology remained "static" throughout the Napoleonic period (IV5.1). After the Russian campaign, which decimated the French forces, Napoleon found it increasingly difficult to sustain France's high troop levels. In addition, naval power was still a French weakness, and Napoleon was aware that the French Navy could not defeat the British Navy in a frontal assault. In fact, at Trafalgar in 1805 Britain defeated France on the seas and gained maritime supremacy over the continent.⁸⁴⁹

More so than technology and weaponry, the conduct of warfare was Napoleon's great strength. The French emperor spawned major innovations in operational mobility. His brilliant tactics, which enabled his rise to fame during the Revolutionary Wars of 1797-9, included among other things the recurrent use of troop concealment, envelopment and pincer movements, the focus on the hinge sections of allied fronts, and the deployment of artillery as infantry support. But Napoleon's tactical genius did not always yield results, and Napoleon lost a few critical

⁸⁴⁸ Gulick 101, 106-8.

⁸⁴⁹ H.W. Wilson, "The Command of the Sea 1803-15," in Lord Acton, Adolphus W. Ward, G.W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907), 206-233; Christon I. Archer et. al., *World History of Warfare* (Lincoln, NE: University of Nebraska Press, 2002), 383; Gulick 97.

battles, such as Aspern-Essling (1809), Leipzig (1813), and Waterloo (1815), as well as the campaign in Russia (1812). Napoleon was undermined by his quest for victory, which led him to recklessly overreach by invading Russia, a fatal strategic mistake that nearly annihilated his army and planted the seeds of his downfall. During the Russian campaign Napoleon also met his tactical match in the person of Clausewitz, a Prussian advising the Russian forces. Clausewitz helped orchestrate a deliberate Russian retreat that lured the overstretched French troops far into Russian territory without any control over their surroundings or their line of retreat.⁸⁵⁰

Napoleon made a few improvements in army organization (IV5.3), such as developing the staff system or standardizing the merit ranking inaugurated by the Revolution, but none were dramatic enough to affect the outcome.⁸⁵¹ He also took some steps back that critically hampered France's chances. For example, although the Revolutionary government had ameliorated and systematized the French draft system, which had been arbitrary and irregular, the system was changed a few years later under Napoleon's rule to allow the wealthiest to pay their way out, which placed an uneven burden on the lower, productive classes. In fact, by 1812, Napoleon faced difficulties in raising enough manpower for his army without disrupting civil and economic life.⁸⁵² In addition, some aspects of the French military organization were inferior to that of its rivals. In particular, France suffered from a disadvantage in logistics. The French army relied mostly on costly and slow overland transport to supply troops, while balancers such as Britain or Russia supplied their forces by boat. As a result, the French often resorted to plundering, which only increased the antagonism of the locals. In contrast, the Britons "paid

⁸⁵⁰ Gulick 101, 126; Carl Von Clausewitz, *The Campaign of 1812 in Russia* (Cambridge, MA: Da Capo Press, 1995), 15-17, 175-200.

⁸⁵¹ Archer 397.

⁸⁵² Georges Pariset, "France under the Empire 1804-14," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907), 114-5.

negotiated prices to the local inhabitants" for the goods they could not supply by sea. During the 1808-12 conflict in Spain and Portugal, for example, "the French starved while British troops were more ... adequately nourished." Supply was particularly problematic during the Russian campaign, even though Napoleon had taken extra precautions. While the French army was supplied by carts transported over land, the Russians had control of the rivers and canal system to supply their soldiers faster and more efficiently.⁸⁵³

If France's military improvements were insufficient to overtake the allies, its non-military skills did not buttress its position either. France's geographic assets had not dramatically changed since Louis XIV (IV6.1). And although Napoleon tried to use economic policy as a vehicle of hegemony, he was ultimately unsuccessful (IV6.2). By the time Napoleon assumed power in 1799, the French state was near bankrupt and faced serious domestic instability, which was one reason why Napoleon consented to the Louisiana Purchase in 1802 although it amounted to a loss of imperial land. Napoleon was able to assuage the economic crisis, notably by restoring order and therefore being again able to effectively collect taxes, but that proved insufficient to generate the long-term growth necessary to contend with the booming British economy.⁸⁵⁴ Even though French manufacturing production had jumped by 40% between the Revolution and 1811, it still "lagged far enough behind the British to make it difficult" for French products "to compete ... with British products." Just as was the case a century earlier, the early 18th century British economy, based extensively on free trade and markets, provided a vast reserve of wealth and resources to prop up British power, while the French economy, which combined "compulsion and reliance upon a more or less free market for mobilizing resources for state purposes," lagged

⁸⁵³ McNeill 199-200, 204.

⁸⁵⁴ Georges Pariset, "The Consulate 1799-1804," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907), 25-6; Pariset, "France under the Empire," 124-5.

behind. The contrast between the two systems became especially marked during the Napoleonic years. Britain "exported goods as well as men ... and thereby contrived to establish a market-supported system of power that proved more durable than anything the French achieved."⁸⁵⁵

The large economic advantage of Britain led Napoleon to devise the Continental System, a continent-wide boycott of Britain started in 1806. Napoleon enlisted the participation of his puppet states and forced allies. The plan failed for several reasons. First, Britain came up with innovative smuggling practices using its naval power. It developed new channels of trade in the Mediterranean and in the Baltic, which the French were unable to stop and which thus only reinforced Britain's commercial and economic power. Second, a large number of signatories violated the blockade, because it clashed "with the interests of a considerable proportion of the European population," for whom cheap British goods were vital.⁸⁵⁶ Finally, French manufacturers were in no position to supplant Britain and fill the European demand for British goods, especially since Britain was providing a number of unique primary goods such as tea, sugar and cotton, for which there were no continental substitutes. Russia, for example, had been one of the "greatest violators" since the end of 1810 and became the main gate through which British products entered Europe, precipitating Napoleon's attack on Russia in 1812. Another violator was Sweden, which really helped Britain because the Swedish defection opened up the northern European sea route.⁸⁵⁷

Napoleon's political and administrative reforms also proved insufficient to grant France an edge over its rivals (IV6.3). Napoleon instituted a series of administrative reforms, many of which have remained in place and now constitute the backbone of modern France. He

⁸⁵⁵ McNeill 186-7, 199, 202.

⁸⁵⁶ Ibid., 202.

⁸⁵⁷ Eli F. Heckscher, *The Continental System: An Economic Interpretation* (Oxford: Clarendon Press, 1922), Part II Chp. 5, 175-82, 230-245; Gulick 98-100; Pariset, "France under the Empire," 122-3.

established the country's first civil, criminal, commercial, and tax codes, the centralized administrative structure by departments, the French central bank, and a public university system.⁸⁵⁸ Although the Revolution had attracted the attention of populations in a number of France's rival countries and France's attempts to export it in the late 17th century had greatly worried Europe's monarchs, Napoleon's administrative reforms failed to generate the same uproar mainly because they were coupled with the return to a rigid, reactionary political rule. Similarly, French society remained fragmented despite some ground-breaking social reforms, such as the emancipation of the Jewish ghettos and the restoration of the Catholic Church's civil status in the wake of the destabilizing anticlerical rules of the Revolution (IV6.4). A rapid population growth, coupled with the pressures of constant warfare and the resurgence of socio-economic inequalities not resolved by the Revolution, generated constant "social frictions" and domestic turbulences under Napoleon's reign.⁸⁵⁹

The Napoleonic case thus further confirms that the factors of balance of power failure operate in reverse to produce balance of power success. The balancers' remarkable ability to overcome deep-seated hostility and collective action problems, as well as France's failure to develop dramatically superior military skills, particularly its organizational problems, and its lack of economic and political precedence, stand out as the primary factors allowing for the restoration of a balance of power in 1715. Other factors, such as bandwagoning, deception, and the rising hegemon's geopolitical and social attributes, played secondary or marginal roles, while

⁸⁵⁸ H.A.L. Fisher, "The Codes," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907), 148-79; Pariset, "The Consulate," 27-8, 33; Pariset, "France under the Empire," 127.

⁸⁵⁹ McNeill 213-215; Pariset, "France under the Empire," 126, 141-2; Bruce 321-3; L.G. Wickham-Legg, "The Concordats," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907), 180-9.

physical communication problems, misperception, laggard balancing, insufficient interest and buckpassing were absent altogether.

c. Research Agenda: the Rarity of Balance of Power Failure

The three instances of successful balance of power examined as research probes substantiate the findings from the case studies and provide greater confidence that the three factors identified among twenty-one as the most important causes of balance of power failure are indeed the correct factors. The identification of the three main factors of hegemony—lack of trust-related collective inaction, the rising hegemon's military achievements, particularly in organization, and its state-building skills—begs one further question. What makes those three factors exceptionally acute in the few cases where hegemony has triumphed, and why were they not as salient in the numerous historical cases where the rising hegemon failed and balance of power was restored? In other words, what makes instances of hegemony so infrequent historically? While this thesis's primary contribution is to find out why regional hegemons sometimes emerge and to locate the unit- and system-level factors most important in explaining that outcome, it cannot ignore the related and equally important question of what makes those factors act powerfully at certain times and not at others. The reason for the historical rarity of hegemony is the second larger question this thesis asks. However, because identifying the causes of hegemony required vast empirical research and thus occupied the bulk of the thesis's effort, we can only offer a few speculative thoughts as to the causes of hegemonic infrequency, and leave in-depth treatment of the issue for a further project, perhaps for the transformation of this thesis into a book manuscript.

First, lack of trust-related collective inaction is the single most important causal factor explaining the occasional success of hegemon, but historically, although distrust and suspicion have always been endemic in international relations, they have most often not been prominent enough to prevent most external balancing from succeeding. Distrust-induced collective inaction is a systemic effect observed at the unit level, since it derives from structural anarchy, which intrinsically breeds distrust. Structural, neorealist theory recognizes that distrust is endemic in international relations and makes cooperation particularly difficult but also maintains that the security challenge posed by a rising hegemon is so grave that other states are able to overcome their suspicions and balance jointly when facing a would-be hegemon. In most cases they do, since balance of power is the most common historic outcome. So why was distrust so profound in the four cases of hegemony that it prevented the cooperation allowed by neorealism?

Second, the two other primary causes of hegemonic success are the rising hegemon's superior military skills, particularly in military organization, and its non-military skills, particularly in state-building, but again, in most historic cases the rising hegemon's unique traits did not procure it such extra efficiency that it was able to outpace the balancers and reach hegemony. Either its achievements were not superior enough to begin with, or the potential balancers were able to copy or counteract them. Why was the rising hegemon able to develop attributes so overwhelming in the four cases of hegemony that the balancers were unable to timely emulate them? Why was one state able to develop such a large comparative advantage in the first place? Neorealism argues that actors adopt the practices of the most successful. As Waltz explains, "contending states imitate the military innovations contrived by the country of greatest capability and ingenuity, ... and even [its] strategies." In addition, because of the "disadvantages that arise from a failure to conform to successful practices," emulation is "not

confined narrowly to the military realm" and should also encompass the rising hegemon's non-military achievements.⁸⁶⁰ So why did this not happen in the four cases?

This thesis offers a few preliminary answers to explain the exceptionality of hegemony, that need to be further pursued. It points clearly to why the successful rising hegemons were unusually proficient. The rising hegemons in the cases all had a variety of unique traits that enabled them to develop and use their resources more efficiently than their rivals: a unique nomadic lifestyle for the Mongols, a unique peripheral location for the Qin, a unique intersection of diverse backgrounds and practices for the Romans, a uniquely shielded position that allowed an exclusive focus on economic growth for the United States. In fact, there is an entire section of literature that focuses on U.S. exceptionalism and highlights the roots of the United States' unique economic, political, social, and military development. It might be less clear, however, why the potential balancers were unable to find ways to copy or counteract the rising hegemons' achievements as fast as neorealism predicts. The thesis highlights that misperception slowed down balancers' response, and even more importantly, that the rising hegemon's deceptive tactics frequently enabled it to mislead the balancers and gain critical time, perhaps enough to overpower the balancers before they could recover and copy its achievements. The literature on adaptation to innovation may help inform us as to the balancers' lack of emulation.

Why potential balancers were unable to overcome their distrust may be more puzzling. This thesis finds that the rising hegemon's deceptive tactics were a secondary or primary factor in each case of hegemony except for the United States, and frequently contributed to aggravating tensions and encouraging immediate gains among the potential balancers, reinforcing their mutual distrust. In comparison, deception was present but seemed more muted in most of the shadow cases. Another critical element that stands out from the four cases of hegemony is the

⁸⁶⁰ Kenneth N. Waltz, *Theory of International Politics* (Boston, MA: McGraw Hill, 1979), 127-8.

potential balancers' reckless leadership. Each of the four cases exhibited rampant reckless behavior among the potential balancers, at a level not present in the shadow-cases, that may have exacerbated trust issues beyond their normal scope. During Rome's rise, for example, all three great power rulers—Hannibal of Carthage, Philip of Macedon, and Antiochus of Syria—displayed unusual brutality and greed when dealing with each other and with the secondary powers that could have joined them. The Khwarezmians, Assassins, Eastern Europeans and others displayed similar traits toward each other and their neighbors as the Mongols approached. During the U.S. rise, Central American and Caribbean leaders incessantly attempted to stir up and support revolutions and coups in their neighbors' states. While reckless behavior by balancers was far from non-existent in the control cases, it appeared to have been less extensive and to have concerned fewer potential balancers, leaving room for balancing coalitions to form under the auspices of the better behaving, more dependable states. The question requires further investigation, but since distrust is structurally driven and thus always present in anarchic interaction, the reason it occasionally becomes so acute as to help produce hegemony presumably lies not with its structural determinant but with the unit-level attributes of the various actors.

2. The Impact of Modernity

After comparing results between variables, it is also enlightening to compare results across cases. The most notable observation that comes to mind is the striking, almost perfect, correlation between some of the variable scores in the Mongol, Roman, and Chinese cases: the important role of deception and distrust-related laggard balancing, but absence of disinterest- and buckpassing-based collective inaction. The U.S. case, in contrast, stands out with its own distinct

scores, notably the larger impact of misperception, the absence of any laggard balancing and communication-related collective inaction, and the primary role of lack of sufficient interest- and buckpassing-induced collective inaction. It thus appears that the U.S. case follows a slightly different pattern than the Mongol, Roman, and Qin cases, with some balancer-related variables performing differently in the three ancient cases than in the modern one. One may wonder whether the divergences on those variable scores reflects the existence of somewhat distinct hegemonic paths in modern and ancient times. However, the results of this study do not show a consistently ancient or modern path to hegemony. First, all four cases share lack of trust-induced collective inaction, as well as the rising hegemon's military achievements and state-building skills as their central variables. Second, the scores of the different types of bandwagoning, military achievement, and non-military skills and characteristics blur the distinction between the ancient cases and the U.S. case entirely. When it comes to bandwagoning, the United States is most similar to the Mongol Empire, with both exclusively displaying instances of profit-driven bandwagoning. With regard to military achievement, the United States scores closest to Rome, with identical material and operational results, but different organizational scores. The United States' non-military skills and characteristics, however, show an almost identical causal impact as Qin's.

As a result, it is clear that the distinct U.S. scores in the first three variables are more due to the United States' unique characteristics and environment as a rising hegemon rather than to a question of era. The European balancers' atypical lack of sufficient interest and decision to pass each other the buck in the U.S. case was a consequence of their extra-systemic roots, distance from the American continent, and concern for the balance of power in their own regional system and had little to do with the century in which it occurred. None of the other

aspiring hegemony shared a similar situation. The Romans and Chinese operated in a more compact system with no external great power intervention. The Mongols spread beyond their immediate region, but had only limited extra-regional great power competition and were not separated from their rivals by an ocean. Had any of the three faced a situation similar to that of the United States, one can safely assume that we would have observed at least some buckpassing and lack of sufficient interest. The absence of laggard balancing in the U.S. case derives directly from the same unique characteristics. Since the rival great powers were so concerned with their own regional balance that they all passed the buck and lacked interest in acting against the increasingly powerful United States, it is unlikely that they would balance jointly at all, even laggardly. In the end, scoring differences among cases stem primarily from the different environmental setup of each system, which led the potential balancers to be particularly apathetic in some instances or the rising hegemon to be particularly thriving in certain domains. This analysis is therefore applicable to all eras and offers predictive power through the ages.

The only variable where the time period might have played a role may be deception. While the rising hegemon's deceptive practices have a significant to decisive causal impact in the Mongol, Roman, and Qin successes, it is only marginal, and in fact almost absent in the U.S. case. The United States did not conceal its rise or use other deceptive means to trick rival great powers. It only occasionally used mild psychological propaganda with minor targets, like the promotion of continental cultural affinity via Pan-Americanism, for example, and bullied Latin American states into granting it commercial privileges in a few rare instances. One can speculate that the United States' choice to abstain from deceptive tricks and psychological warfare reflects the modern evolution toward informal expansion: the United States aimed not at intimidating but at convincing its targets to obtain access, and sought to avoid direct confrontation with European

great powers at least in the early part of its rise, making the preservation of a benevolent reputation crucial. While the choice of informal tools and thus limited deception is at least partly a result of the modern evolution of international interaction, however, it may also be due, to a great extent, to the unique location and situation of the United States. The United States may simply have shied away from deception to avoid provoking the extra-systemic great powers' direct intervention—by triggering pleas for help from intimidated local powers, for example. The United States probably knew that the great powers would likely not intervene except in the face of a major provocation given their distance and the cost of intervention. Refraining from tricking its targets thus helped ensure that the United States would preserve a free hand in the affairs of the hemisphere.

Even though there may not be a uniquely modern balance of power failure pattern, the United States' regional hegemony still remains a puzzle because it is the only case of hegemony in modern history. One must wonder why balance of power failure seems to have been even less frequent in modern than in ancient history. Does it mean that potential balancers are now more successful than in the past, or that aspiring hegemons are less resourceful? It is possible that the modern international environment has become increasingly constraining for rising hegemons, and several arguments have been made that may substantiate such a claim. The central liberal proposition, for example, maintains that the modern world makes it harder for potential hegemons: since the 19th century the world has witnessed growing economic interdependence, multilateralism, and democratic affinity that severely curtail an aspiring hegemon's expansion options or desires to become a hegemon. Great powers are now so intertwined by trade and commercial and monetary exchange, the argument holds, that attempting to break out of the pack and dominate others against their will would be costly and even potentially crippling for an

aspiring hegemon. Similarly, international institutions and multilateral practices restrain great power action by rendering unilateralism costly and creating a normative interest in the preservation of the existing order. Constructivist scholars add that great powers are now increasingly constrained in their actions by the development of international norms of behavior to which they are forced to adhere to uphold their reputation and international legitimacy. Finally, the spread of democracy throughout the world, particularly among great powers and thus potential hegemonic candidates, raises domestic hurdles on the growth of military power and the use of force, hindering hegemonic bids.⁸⁶¹ Yet, while the forces highlighted by the liberal and constructivist scholarship might indeed have transformed great power interaction in the modern world, it is unclear that they would deter a great power from attempting to become a hegemon or prevent it from succeeding.

In fact, such forces have not stopped a number of great powers besides the United States from attempting hegemonic bids since the 19th century. Even though Germany and its European neighbors, particularly Britain, were highly interdependent commercial partners sharing unprecedented levels of trade in the early 20th century, for example, Germany twice attempted to become the European hegemon in the first half of the century, and it was only the collective military intervention of its neighbors that defeated it. Participation in multilateral institutions and the existence of specific normative practices do not seem to restrain aspiring hegemonies either. On the contrary, some great powers are able to use those institutions and norms as a vehicle to enhance their power and control over a region while simultaneously obtaining

⁸⁶¹ See, for example, Bruce Russett and John Oneal, *Triangulating Peace: Democracy, Interdependence, and International Organizations* (W.W. Norton & Company, 2001); Robert O. Keohane, *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton, NJ: Princeton University Press, 1984); John G. Ruggie, "Multilateralism: The Anatomy of an Institution," *International Organization* 46.3 (Summer 1992); Robert Keohane and Lisa Martin, "The Promise of Institutional Theory," *International Security* 20.1 (Summer 1995); Martha Finnemore, *National Interests in International Society* (Ithaca, NY: Cornell University, 1996).

international blessing and thus upholding their reputations, as the United States did with NATO during the Cold War, for instance, or more subtly with conditional financing hinging on human rights practices or representative institutions in the developing world.⁸⁶² As for democratic norms and domestic structures, they have not prevented great powers from engaging in wars and military expansion, and there is insufficient evidence to conclude that democratic affinity has restrained democratic great powers' actions against fellow democracies. Thus, it cannot be concluded that the changes the international environment has undergone in the modern world have sufficiently constrained great powers to render hegemonic bids prohibitively costly or difficult.

Could it then be that the modern world has had a positive impact on the potential balancers' side of the equation and that modern developments have made balancing easier? Does the fact that several rising hegemony were stopped in the modern era mean that the modern world has removed or loosened the potential balancers' barriers to cooperation—their trust issues, buckpassing tendencies, and lack of sufficient interest—or reduced their incentive to bandwagon? This does not seem to be the case either. For example, even though the European great powers and the United States brought the rise of Hitler's Germany to a halt in the 1940s with a wide-ranging countervailing military alliance, there were ample examples of collective inaction, laggard balancing, and bandwagoning in the 1930s. When the Allies intervened, Germany had already become very powerful, and required the participation of all the major great powers to stop it, while earlier intervention could have stopped its rise with much fewer efforts. In the 1930s, France and Britain passed each other the buck and at times accommodated

⁸⁶² John Mearsheimer, "The False Promise of International Institutions," *International Security* 19.3 (Winter 1994-5), 7, 13-14; David E. Spiro, "Give Democratic Peace a Chance? The Insignificance of the Liberal Peace," *International Security* 19.2 (Autumn 1994), 51-53; Ido Oren, "The Subjectivity of the 'Democratic' Peace: Changing U.S. Perceptions of Imperial Germany," *International Security* 20.2 (Autumn 1995).

Germany, while the United States' efforts at stalling the rising hegemon were clearly laggard and the Soviet Union bandwagoned with Germany. In other words, the same balancing difficulties that were at play in the four cases of hegemonic success were also present in the 1930s. Although the outcome was different—the balancers eventually managed to overcome their difficulties against Germany and restore the balance of power—the modern characteristics of the international system do not seem to have alleviated the balancers' original difficulty to cooperate or their hesitation to balance until it was almost too late.

Thus, although some changes in the international environment and evolutions in international interaction have undoubtedly taken place, it does not appear that balancing has become easier or that aspiring hegemons are more constrained now than in pre-modern times. If the same factors are still at play, how can we then explain that the United States remains the only case of hegemony in the modern world, while ancient history has witnessed more successful hegemons, the largest being the Mongols, the Romans, and the Qin? When thinking in broad historical terms, it is in fact not surprising at all that the United States remains the sole modern hegemon. The international relations literature generally dates the onset of the modern world in 1815 with the Congress of Vienna and the reestablishment of the balance of power following the demise of Napoleon, and occasionally in 1713 after the Treaty of Utrecht and the restoration of the balance of power at the end of the War of the Spanish Succession. The modern world has only spanned two to three centuries, depending on which date one acknowledges. In comparison, the ancient and pre-modern worlds have lasted over two millennia, from the Greek city-states and Rome (but one could start much earlier) to 1713 or 1815. There were seventeen centuries alone between the early days of Rome's rise and the advent of the Mongol hegemony. The most obvious explanation is thus simply that not enough time has elapsed in the modern world to have

allowed for more successful hegemons. If a rising hegemon succeeds in its region every four or five centuries on average, as seems to have been the case in the pre-modern world, we may not expect the next hegemon for another couple of centuries, or perhaps longer since we are now in a quasi-hegemonic situation of unipolarity.

3. Balance of Power Failure and Unipolarity

The current distribution of power and exact nature of the United States' status in the global system has been subject to much debate and provides an interesting addition to the question of balance of power failure. While it is undisputed that the United States remains in control of the Western Hemisphere, its unique position resulting from the end of the Cold War and its unmatched capabilities begs the question of whether the United States might not control much more than the Western Hemisphere. The United States currently enjoys wide supremacy over all other great powers worldwide with regard to both military and economic indicators of power; it dominates other great powers technologically and benefits from the advantage of its location. In 2006, the United States accounted for over 65% of the great powers' military expenditures, while the other great powers, China, Japan, Germany, Russia, France, and Britain scored in the single digits, and it accounted for 46% of the world's total military expenditures. By comparison, in the late 19th century, Russia, Britain, and France each accounted for roughly 20% of the world's military expenditures, while the United States and Germany each had about 10%. The distribution of capabilities was thus much more balanced than it is today. In 2006 the United States also held over 46% of the great power GDP and over 27% of the world's GDP, with the other great powers lagging far behind with single digits. While the European Union's combined GDP surpassed that of the United States, its lack of unanimity in defense and security issues, and

even on some economic and monetary policies, restrains its potential compared to the United States. The United States was also responsible for 39% of the world's high tech production in 2003, far ahead of its nearest competitor, China, with 12%, although China's high tech exports have been growing steadily. "The main feature of the distribution of capabilities today is thus unprecedented American primacy," Stephen Brooks and William Wohlforth argue. While past great powers and hegemons have often enjoyed superior capabilities over their peers, "the depth, scale, and projected longevity of the U.S. lead in each critical dimension of power are noteworthy" and "what truly distinguishes the current distribution of capabilities is American dominance in all of them simultaneously."⁸⁶³

In addition to its sheer power, the United States also enjoys an unprecedented degree of influence over the world, enabled by its "command of the commons," as Barry Posen points out. Not only does the United States possess naval supremacy, as did a number of great powers before it such as Rome or Britain in the 19th century, providing it with virtually unrestricted maritime access around the globe, but it also controls the space and air. Since very few states have the means to bar the United States from using their high altitude airspace, the United States has wide leeway to patrol the globe's landmass in addition to its oceans, and gather information as it sees fit. "Command means that the United States gets vastly more military use out of the sea, space, and air than do others; that it can credibly threaten to deny their use to others and that others would lose a military contest for the commons if they attempted to deny them to the

⁸⁶³ Stephen G. Brooks and William C. Wohlforth, *World Out of Balance: International Relations and the Challenge of American Primacy* (Princeton, NJ: Princeton University Press, 2008), 29-35; Eurostat, "World Market Share of High Tech Exports, Leading High Tech Trading Countries" (2006), available at http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/High-tech_statistics (accessed Aug. 7, 2010).

United States," Posen remarks. Moreover, the United States' superior capabilities makes a successful challenge to the U.S. command in the short- to medium-term highly unlikely.⁸⁶⁴

Can it be concluded that the United States has become the world's first true global hegemon in the sense we have applied in this study, given its overwhelming power and influence throughout the globe? If not, what is the United States' current status and what future global power configuration can we expect? The press has widely called the unprecedented American primacy hegemony, and so have a number of international relations scholars.⁸⁶⁵ The use of the word hegemony as a generic description of overwhelming power and influence is misguided, however, and does not withstand the stricter definition used in this study. While the United States is the dominant actor in many regions led by secondary powers, such as the Middle East, it is unlikely that the United States would be able to contend against or exert full control over all of the system's great powers—China, Japan, Germany, Russia, France, and Britain—or a combination thereof, particularly in their respective regions, as a hegemon would be able to do. As Robert Ross notes, although "the United States is the world's only superpower" ... "the existence of a single superpower does not necessarily imply global hegemony." The United States enjoys naval supremacy over maritime Asia, for example, but it does not dominate mainland East Asia. Instead, the East Asian region is currently bipolar and "well into the 21st century neither China nor the United States will be able to challenge each other's dominance in their respective spheres of influence," Ross concludes.⁸⁶⁶

⁸⁶⁴ Barry R. Posen, "Command of the Commons: The Military Foundation of U.S. Hegemony," *International Security* 28.1 (Summer 2003), 8, 19.

⁸⁶⁵ See, for example, Brooks and Wohlforth, *World*, 23, 27-28, 35, 45; Posen, "Command," 5-6.

⁸⁶⁶ Robert S. Ross, "Bipolarity and Balancing in East Asia," in T.V. Paul, James J. Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford, CA: Stanford University Press, 2004), 267-8.

While the United States is thus currently still a regional, as opposed to a global hegemon, its dominant power qualifies it as the sole superpower, or unipole—a great power that lies on the threshold of hegemony but has not quite reached it.⁸⁶⁷ A unipolar system, unlike hegemony, is still a balance of power system because a “unipolar leader is not immune to balancing by a coalition of second-ranked states.”⁸⁶⁸ There remain a number of rival great powers in the global system—mainly China and Europe—that could balance against the United States and restrain its growth. However, because a unipole is so close to hegemony, its power is so great that balancing can only succeed if all the other powers in the system combine their efforts. Since the United States monopolizes almost half of the world's military capabilities and a substantial share the world economy, the individual states that possess the remaining 50% of capabilities would have to all throw in their weight in opposition to it to stand a chance of decisively curtailing its influence.⁸⁶⁹

One may wonder, then, why potential balancers have allowed the United States to acquire such overwhelming power in the first place and did not balance when it was easier and less costly to do so. Does such lack of action not constitute a quasi-balance of power failure in itself, since it has allowed the United States to reach power levels on the verge of hegemony? Unipolarity has been a rare historical occurrence precisely because it constitutes the last stage before hegemony; it has materialized mostly prior to a rising state's attainment of regional hegemony, for example in the Roman, Mongol, or Qin cases. In cases of balance of power success, however, which are preponderant historically, the balancers never let the rising hegemon go as far as unipolarity. In reality, the United States reached its unipolar status not

⁸⁶⁷ Stephen M. Walt, "Alliances in a Unipolar World," *World Politics* 61.1 (January 2009), 87.

⁸⁶⁸ Robert Pape, "Soft Balancing against the United States," *International Security* 30.1 (Summer 2005), 11-14.

⁸⁶⁹ For a useful discussion of how much power is needed to balance a unipole, see Barry Posen, "From Unipolarity to Multipolarity: Transition in Sight?" (forthcoming chapter, April 2010 draft), 7-8 and fn. 17.

because other great powers failed to balance but rather accidentally when the USSR, one of the two poles of the Cold War's bipolar power distribution, abruptly collapsed in the early 1990s, leaving the United States as the sole pole remaining. The current case is thus unique because the United States ended up a unipole fortuitously, rather than on its way to hegemony.⁸⁷⁰

Because of this unusual path, analysts in the 1990s and early 2000s expected the "unipolar moment" to be brief and multipolarity to return quickly. While most great powers had not needed to balance the United States throughout the bipolar Cold War because the USSR kept the United States in check, it was believed that after the USSR collapsed, they would quickly step in to put an end to unipolarity, and restore multipolarity.⁸⁷¹ Because the imminent danger of hegemony looms over the system, unipolarity is the "least stable international configuration," as Waltz points out, and should immediately lead to balancing and the restoration of multipolarity or quickly give way to hegemony.⁸⁷² The unipolar moment has lasted for the past twenty years, however, and the configuration of power appears to be roughly the same as at the end of the Cold War. The persistence of this unusual power pattern raises a number of questions. Why has there not be a significant balancing reaction against U.S. primacy? Even though the United States has not become a hegemon, does the persistence of unipolarity amount to a balance of power failure of sorts? What can we expect in the future—a return to multipolarity, as balance of power theory scholars argue; a U.S. move toward hegemony, as some say the United States' unilateral policy moves suggest; or a continuation and stabilization of unipolarity?

⁸⁷⁰ Stephen M. Walt, *Taming American Power: The Global Response to U.S. Primacy* (New York: W.W. Norton & Company, 2005), 29, 60; William C. Wohlforth, "The Stability of a Unipolar World," *International Security* 24.1 (Summer 1999), 5.

⁸⁷¹ See, for example, Charles Krauthammer, "The Unipolar Moment," *Foreign Affairs* 70.1 (Winter 1990-1991), 23-33; Christopher Layne, "The Unipolar Illusion: Why New Great Powers Will Arise," *International Security* 17.4 (Spring 1993), 5-51.

⁸⁷² Kenneth Waltz, "Evaluating Theories," in John Vasquez and Colin Elman, eds., *Realism and the Balance of Power. A New Debate* (Upper Saddle River: Prentice Hall, 2003), 54.

The first observation is that analysts pointing to lasting unipolarity may have shown signs of impatience. They expected balancing mechanisms to immediately jump into full gear as soon as power appeared unbalanced and have thus judged anything short of a massive anti-American reaction as "sluggish." But because balance of power remains a structural outcome, the reaction may take some time to occur. "A new balance is always emerging slowly; in historical perspective, it will come in the blink of an eye," Waltz reminds us.⁸⁷³ More importantly, although the presence of a unipole is "uncomfortable," it monopolizes such overwhelming power that balancing requires vast resources and bears exponential risks, so that balancing a unipole will not be undertaken casually or quickly.⁸⁷⁴ As G. John Ikenberry point outs, "even as unipolarity increases the incentives for counterbalancing, it also raises the costs." Not only do many states benefit from the unipole's leadership and provision of public goods, but carelessly provoking the unipole may yield destruction, as Saddam Hussein's Iraq discovered in 2003.⁸⁷⁵ In fact, rather than the massive, frontal retaliation expected in traditional balancing cases, potential balancers in a unipolar situation are likely to favor subtler approaches.

Because of the great asymmetry of power, states are more likely to balance against the United States when Americans are "unable or unwilling to respond," for example, when the costs of retaliation are greater than the benefits, by multiplying small scope operations, or when the United States is already "busy elsewhere," Walt explains. They look for "windows of opportunity" to challenge U.S. power. For example, Saddam Hussein's decision to invade Kuwait in 1990 was based on a calculation that the United States was "preoccupied by the collapse of the Soviet Union" and would not commit itself to defend the small state of Kuwait.

⁸⁷³ Kenneth Waltz, "Structural Realism after the Cold War," in G. John Ikenberry, *America Unrivaled: The Future of the Balance of Power* (Ithaca: Cornell University Press, 2002), 52-6; Walt, "Alliances," 91.

⁸⁷⁴ Posen, "From Unipolarity to Multipolarity," 12-13.

⁸⁷⁵ G. John Ikenberry, Michael Mastanduno, and William C. Wohlforth, "Unipolarity, State Behavior, and Systemic Consequences," *World Politics* 61.1 (January 2009), 18.

North Korea's decision to reopen its nuclear program in 2003 was similarly based on a calculation that the United States was preoccupied with Iraq and would not want to be engaged in a second theatre. The same logic may lie behind Iran's current quest for nuclear weapons. It is thus obvious that potential balancers in unipolarity take advantage of situations unfavorable to the unipole to strike modest blows against it.⁸⁷⁶

In reality, there have been a number of instances of balancing against the United States since the 1990s. Balance of power is clearly not "inoperative" as some have claimed. It merely appears so because countervailing efforts have been more muted than one expects in a traditional multipolar setting due to the "sheer size and comprehensiveness of the power gap favoring the United States," which makes it not only much harder and costlier to resist, but also raises the hurdles traditionally associated with balancing, such as trust issues. As Brooks and Wohlforth point out, for instance, the unipolar distribution of power increases the opportunity-cost of balancing and "raises still higher the coordination and collective action barriers to external balancing," which in a unipolar world is the most likely, and perhaps even only, form of balancing that has a chance to succeed. Nevertheless, the persistence of even modest balancing efforts, despite the large power gap, highlights that the rest of the world is not indifferent to U.S. power and that we should therefore not expect hegemony or durable unipolarity, but rather an eventual return to multipolarity.⁸⁷⁷ In addition, the presence of a number of potential rivals to U.S. power, both regional great powers like the European Union and China, and medium and emerging powers, such as India, Brazil, Japan, and perhaps Russia, indicates that there are candidates for restoring multipolarity in the future.⁸⁷⁸

⁸⁷⁶ Walt, *Taming*, 112-120.

⁸⁷⁷ Brooks and Wohlforth, *World*, 36-7.

⁸⁷⁸ Posen, "From Unipolarity to Multipolarity," 8-10.

The most serious instances of anti-American balancing so far have consisted of some key actors reinforcing their military posture and building up their capabilities, both internally and externally. The European Union has the potential to become a serious balancer because it possesses a vast economic reservoir that could fuel long-term military growth, with a combined GDP that surpasses that of the United States.⁸⁷⁹ In addition, although its coordinated security and defense arm, now called the Common Security and Defense Policy (CSDP), is still in its infancy, it could become an effective tool because it would be embedded in the pre-existing multilateral or even supranational framework of the Union and would thus not require building collaborative structures from scratch, and because the individual European members' military and technological assets are among the world's most advanced. Britain and France are the only states besides the United States to possess global power projection capabilities, for instance. The CSDP combines both "decision making and operational capabilities." While common European military and defense actions still require unanimity, the EU has issued a common security strategy in a format similar to the United States' and set up a foreign ministry and an agency that implements the strategy and coordinates military activity, capabilities, research, and investment among the members. The common efforts have succeeded in sending a number of joint deployments abroad, including large peacekeeping and security missions in the Balkans and Africa, with over 6,000 troops deployed. The CSDP had a budget of 285 million Euros for 2008, only a fraction of America's military spending, but was complementary to the individual European state defense budgets.⁸⁸⁰

⁸⁷⁹ 2009 figures from International Monetary Fund, "World Economic Outlook Database" (April 2010), available at <http://www.imf.org/external/pubs/ft/weo/2010/01/weodata/index.aspx> (accessed Aug. 7, 2010).

⁸⁸⁰ Giji Gya, "ESDP and EU Mission Update," *European Security Review* 36 (December 2007), 3-4; Barry R. Posen, "European Union Security and Defense Policy: Response to Unipolarity?" in *Security Studies* 15.2 (April-June 2006), 179-182.

While many pundits point out that the CSDP, like previous European collaborative defense attempts such as the European Defense Community of the 1950s, was not principally meant to balance the United States but rather to address regional security issues and perhaps respond to (unfounded) fears of U.S. withdrawal from the European theatre, CSDP cannot be divorced from balancing concerns.⁸⁸¹ Balance of power was at least one motive for the creation of the CSDP in 1999, at a time when Europe became increasingly conscious of the growing gap between its capabilities and those of the United States, blatantly exposed by the U.S.-NATO intervention in Kosovo, Europe's own backyard.⁸⁸² Moreover, Europe already has a highly efficient military wing at its disposal to take care of regional security issues, NATO, under American leadership, so one may wonder why the EU would seek a separate military instrument parallel to, and virtually overlapping with, NATO, if not to assert its ability to act independently of the United States. "European states ... are balancing U.S. power," Barry Posen explains, even if still modestly. The CSDP gives them the means to have their own voice in security affairs and stand up to the United States if they wish to do so.⁸⁸³

China is another actor that has shown signs of balancing against the United States, both internally, by reinforcing its military position, and externally, by cooperating militarily with states intent on resisting the United States. Sustained by a soaring economic growth rate, China has been steadily swelling its defense budget since the mid-1990s, with an average of 11.6% annual spending increase between 1996 and 2006. The People's Liberation Army has been actively modernizing both its conventional and nuclear arsenals, and investing heavily in research and development of high-tech military devices. Perhaps the clearest indication of China's preparation to contest U.S. supremacy has taken place in the naval area. While the

⁸⁸¹ Brooks and Wohlforth, *World*, 80-83.

⁸⁸² Robert J. Art, "Correspondence: Striking the Balance," *International Security* 30.3 (Winter 2005/06), 181.

⁸⁸³ Posen, "Policy," 149-151.

Chinese Navy has traditionally focused on coastal defense and been a somewhat neglected branch of the Chinese military, it has made headlines in recent years by showing interest in developing blue-water capabilities. The Chinese Navy has displayed several nuclear-powered submarines in offshore maneuvers in 2009 and 2010, and Chinese officials have repeatedly announced China's intention to build an aircraft carrier, which was confirmed by the Pentagon, as well as other power projection tools such as large warships and stealth submarines. Because maritime East Asia is largely dominated by the U.S. Navy, China's power projection is bound to collide with the United States. Moreover, there have been suggestions that the Chinese may be in the process of permanently securing their nuclear arsenal by developing a mobile, second-strike force, which would allow China to stand up more forcefully to U.S. pressures.⁸⁸⁴

In addition to its internal efforts, China has also enhanced its position vis-à-vis the United States by entering into a 'strategic partnership' with Russia in 2001, which includes extensive arms sales. While the partnership was meant in part to create mutually beneficial commercial exchanges and solve regional issues they share such as separatism and Islamic radicalism, both Russia and China have also repeatedly insisted that the collaboration is "an expression of their preference for a multipolar world," a clear hint at a balancing purpose. Another manifestation of China's contestation of U.S. primacy is its energy and arms deals with Iran, an open challenge to U.S. policy efforts to isolate the blacklisted Iranian regime. Although the collaboration predates the U.S. accession to unipolarity and thus was not originally aimed at balancing the United States, the exchanges have been renewed since the early 1990s and have included the issue that has been most irritating to the United States, nuclear technology. And

⁸⁸⁴ U.S. Department of Defense, *Annual Report to Congress: Military Power of the People's Republic of China* (Office of the Secretary of Defense, 2009), 18, 36, 40; Keir A. Lieber and Daryll G. Press, "U.S. Nuclear Primacy and the Future of the Chinese Deterrent," *China Security* 5 (Winter 2007), 72.

even though the United States has managed to pressure China into renouncing some of its dealings with Iran since the late 1990s, some arms sales have undeniably continued.⁸⁸⁵

Just like with the CSDP, some scholars argue that China's dealings with Russia and Iran do not constitute balancing but are prompted by other rationales such as regional security or economic gain and simply account for normal great power politics.⁸⁸⁶ However, collaboration can occur for a variety of reasons and resisting U.S. primacy does not preclude a balancer from simultaneously pursuing other goals. States rarely have monolithic interests and in many ways balancing and regional security or economic gains may be complimentary, rather than mutually exclusive, rationales. Moreover, it is not surprising that a balancer would pursue several interests at once to disguise or cushion its balancing efforts given the risks and potential costs of balancing under unipolarity. Thus, the fact that China has not openly provoked the United States does not prove it is failing to balance. On the contrary, there seems to be a clear correlation between the occurrence of unipolarity and China's military strengthening efforts, which have dramatically accelerated in the early-to mid-1990s as the United States acquired its sole superpower position. China's military reforms, like its foreign collaborations, naturally serve several purposes, such as establishing the country as the regional hegemon and diffusing real and perceived threats from Taiwan, Korea, Japan, and Southeast Asia, but they cannot be seen in isolation. Indeed, some of China's military upgrades have been directly aimed at countering specific U.S. advantages and can therefore only constitute a countervailing effort against U.S. primacy.⁸⁸⁷ In the end, the seemingly contradictory nature of China's behavior—military reforms, foreign collaborations and pro-multipolarity rhetoric undermining the United States,

⁸⁸⁵ Bates Gill, "Chinese Arms Exports to Iran," *Middle East Review of International Affairs* 2.2 (May 1998); Brooks and Wohlforth, *World*, 72-5.

⁸⁸⁶ Brooks and Wohlforth, *World*, 69-71, 95-96.

⁸⁸⁷ Walt, *Taming*, 132-139

combined with simultaneous cooperation with and cajoling of the United States—is clearly indicative of a balancing strategy under unipolarity: resist but surreptitiously because it is too costly to be upfront about it.

Besides internal and external military strengthening efforts, which touch the core of balancing, some countries have used more nuanced, non-military tools to counter the United States, an approach sometimes called "soft-balancing." Soft-balancing strategies include coordinated diplomatic, economic, and institutional actions that aim to raise the cost of American unilateral action and create hurdles restraining U.S. freedom of action around the world. A number of scholars look critically upon the notion of soft-balancing because it does not directly shift the distribution of capabilities and to them should thus not be considered balancing. However, soft-balancing offers a safe option to balancers under the constraints and costs of unipolarity. Moreover, although it does not immediately alter relative power positions, soft-balancing has an indirect effect on U.S. primacy because it "limits the ability of the United States to impose its preference on others."⁸⁸⁸

One example of soft-balancing is Western Europe's use of international institutional rules and norms to delay U.S. war plans in Iraq before the beginning of the 2003 Operation Iraqi Freedom, by attempting to tie the United States to multilateral UN decisions. The purpose, for France, Germany, and Russia, was to obtain "more say over, and more independence from, American policies." At the same time Turkey and Saudi Arabia engaged in another type of soft-balancing by denying or threatening to deny use of their territory for the U.S. war effort, rendering the war effort significantly costlier for the United States. In the Asian theater, South

⁸⁸⁸ T.V. Paul, "The Enduring Axioms of Balance of Power Theory and their Contemporary Relevance," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004), 1-4; Robert Art, "Europe Hedges Its Security Bets," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004), 199; Art, "Correspondence," 178-83; Page 9-10; Walt, *Taming*, 126-7.

Korea and China have at times used their diplomatic leverage to rein in U.S. actions against North Korea, and the European Union, China and Russia are currently doing the same with respect to Iran.⁸⁸⁹ Such actions constitute balancing because they make it harder for the United States to use its military around the world. Soft-balancing can also diminish U.S. influence, by publicly denouncing U.S. policies, encouraging protests both domestically and abroad, and discouraging countries to cooperate with the United States. It can additionally impose real military costs on the United States. For example, France and Germany's diplomatic barrage led several countries to decline participating in the Iraq War, resulting in the United States bearing the quasi entire financial and human burden for the war. The same would have likely happened in even greater proportions if the United States had decided to take military action against Iran. The strength of the anti-Iraq War coalition ultimately forced the United States to show restraint in its dealing with Iran, as Walt points out. Soft balancing, although it has not affected the distribution of power, has thus had very real consequences for the unipole and has thus become accepted as a safe balancing option under unipolar constraints.⁸⁹⁰

Besides military strengthening and soft-balancing, a few balancers have also used one last type of anti-American intervention, asymmetric balancing. Unlike soft-balancing, asymmetric balancing is undertaken by secondary powers or non-state actors that are markedly weak and would under normal circumstances easily subject to U.S. pressure. They possess too little influence or are too isolated to engage even in soft-balancing, so their only possible balancing method is to develop tools that artificially inflate their power and allow them to interdict American influence on them. Unconventional warfare is one example of a successful asymmetric tool that some actors, like Serbia in the mid-1990s and Afghan and Iraqi insurgents

⁸⁸⁹ Art, "Europe," 204; Walt, *Taming*, 131; Paul 11-17.

⁸⁹⁰ Page 9-10, 36-43; Walt, *Taming*, 131.

since 2001, have used against the unipole. These actors know that the United States possesses an unmatched advantage in conventional warfare, but that it has shown some difficulties with urban warfare and guerilla-type insurgency, which are both cheap to put in place and render the United States' highly sophisticated conventional arsenal virtually powerless. NBC weapons are another asymmetric tool embraced by weak powers intent on resisting the United States. North Korea, Iraq, and Iran have all developed or sought to develop NBC weapons because just a few of those weapons are sufficient to inflict massive damage on a target, regardless of the enemy's own capabilities. They are therefore particularly cost-efficient to develop against a much more powerful enemy like a unipole. Terrorism is a final asymmetric tool used to fend off U.S. primacy. As with NBC weapons, one terrorist attack can inflict vast amounts of damage on a significantly more powerful enemy like the United States, and when perpetuated by non-state actors has the advantage of rendering retaliation particularly difficult. Asymmetric balancing works similarly to soft-balancing, not by shifting the distribution of power but by raising the cost of U.S. action and in effect limiting its reach into certain regions, except that asymmetric tools multiply the potential cost exponentially because they focus on counter-value targets.⁸⁹¹

Thus, in the end, the Mongols are poised to remain the sole global hegemon to date. The United States is unlikely to follow in their footsteps because it faces circumstances very different from the Mongols', chief among them the presence of numerous potential balancers that are more powerful both militarily and economically and that dominate their respective regions. While trust issues are always present among the potential balancers, they seem much less salient than in Mongol times given the extent and depth of current cooperative efforts and alliance patterns. There is no indication that the United States is attempting to deceive potential balancers

⁸⁹¹ Waltz, "Structural Realism," 64; Walt, *Taming*, 132-139, 187; Paul 11-17; Posen, "From Unipolarity to Multipolarity," 14-15.

or conceal its real power, and there is no evidence either of potential balancers' misperception of American power or intentions. While except in a few cases, the potential balancers do not consider the United States an immediate physical danger to them, the overwhelming power of the unipole nevertheless remains a constraint and latent threat of which are aware.⁸⁹² And while current balancing efforts are "nowhere near adequate to ... affect the United States' overall military primacy and its control of the commons," as many scholars have pointed out, such efforts have managed to diminish U.S. influence and render U.S. control harder to achieve and sustain in some areas, and we can anticipate that they will continue to do so and perhaps pave the way for more decisive balancing in the future.⁸⁹³ Thus, rather than hegemony or lasting unipolarity, we should expect multipolarity to resume, but because of the extent of the power gap intrinsic to unipolarity and the constraints it generates for balancers, a fundamental change in the distribution of power can only take place in the medium to long term.

⁸⁹² Walt, "Alliances," 95-6.

⁸⁹³ Brooks and Wohlforth, *World*, 35-6, 43.

Bibliography

Chapter 1: Introduction

- The White House, *The National Security Strategy of the United States of America*, Washington, D.C. (September 2002).
- Polybius, *The Histories*, transl. W.R. Paton (Cambridge, MA: Loeb Classical Library, 2010).
- Sun Tzu, *The Art of War*, transl. Lionel Giles (El Paso, TX: El Paso Norte Press, 2009), XI.30.
- Thucydides, *History of the Peloponnesian War*, transl. Rex Warner, (New York: Penguin Books, 1954).
- "U.S. Fears Grow over China Military," *BBC News* (25 May 2007), available at <http://news.bbc.co.uk/2/hi/6691691.stm> (accessed Aug. 10, 2010).
- "President Hu: PLA Budget to Rise with the Economy," *China Daily* (2 August 2007), available at http://www.chinadaily.com.cn/china/2007-08/02/content_5447330.htm (accessed Aug. 10, 2010).
- Butterfield, Herbert, "Balance of Power," in Philip P. Wiener, ed., *Dictionary of the History of Ideas: Studies of Selected Pivotal Ideas* (New York: Charles Scribner's Sons, 1973).
- Finnemore, Martha, *The Purpose of Intervention: Changing Beliefs about the Use of Force* (Cornell: Cornell University Press, 2004).
- Jervis, Robert, *Systems Effect: Complexity in Political and Social Life* (Princeton: Princeton University Press, 1997).
- Layne, Christopher, "Kant or Cant: The Myth of the Democratic Peace," *International Security* 19.2 (Autumn 1994).
- Li, Dun J., *The Ageless Chinese: A History* (New York: Charles Scribner's Sons, 1978).
- Waltz, Kenneth N., *Theory of International Relations* (Boston: McGraw Hill, 1979).

Chapter 2: A Conceptual Framework for Balance of Power Failure

- Art, Robert J., "Europe Hedges Its Security Bets," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004).
- Art, Robert J., "Correspondence: Striking the Balance," *International Security* 30.3 (Winter 2005/06).
- Biddle, Stephen, *Military Power: Explaining Victory and Defeat in Modern Battle* (Princeton, NJ: Princeton University Press, 2004).
- Brooks, Stephen G. and Wohlforth, William C., "Hard Times for Soft Balancing," *International Security* 30.1 (Summer 2005).

- Butterfield, Herbert, "The Balance of Power," in Herbert Butterfield and Martin Wight, eds., *Diplomatic Investigations: Essays in the Theory of International Politics* (Cambridge: Harvard University Press, 1996).
- Chase-Dunn, Christopher, et. al., "Semi-Peripheral Development and Empire Upsweeps Since the Bronze Age," *University of California-Riverside Institute for Research on World-Systems Working Paper #56*, presented at the annual meeting of the International Studies Association, New Orleans, LA (Feb. 20, 2010).
- Claude, Inis, *Power and International Relations* (New York: Random House, 1962).
- Crawford, Timothy W., "Wedge Strategy, Balancing, and the Deviant Case of Spain, 1940-1941," *Security Studies* 17.1 (January 2008).
- Doyle, Michael, "Balancing Power Classically: An Alternative to Collective Security?" in George Downs, ed., *Collective Security Beyond the Cold War* (Ann Arbor: University of Michigan Press, 1994).
- Doyle, Michael, *Ways of War and Peace* (New York: W.W. Norton & Company, 1997).
- Elman, Colin, "Horses for Courses: Why Not Neorealist Theories of Foreign Policy?," *Security Studies* 6.1 (Autumn 1996).
- Fortman, Michael, Paul, T.V., and Wirtz, James J., "Balance of Power at the Turn of the New Century," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004).
- George, Alexander and Bennett, Andrew, *Case Studies and Theory Development in the Social Sciences* (Cambridge: BCSIA Harvard University, 2005).
- Gilpin, Robert, "The Theory of Hegemonic War," *Journal of Interdisciplinary History* 18.4 (Spring 1988).
- Grieco, Joseph, Powell, Robert, and Snidal, Duncan, "The Relative Gains Problem for International Cooperation," *The American Political Science Review* 87.3 (Sept. 1993).
- Gulick, Edward Vose, *Europe's Classical Balance of Power: A Case History of the Theory and Practice of One of the Great Concepts of European Statecraft* (Wesport, CT: Greenwood Press, 1955).
- Haas, Ernst, "The Balance of Power: Prescription, Concept, or Propaganda," *World Politics* 5.4 (July 1953).
- Jervis, Robert, *Perception and Misperception in International Politics* (Princeton, NJ: Princeton University Press, 1976).
- Jervis, Robert, "Cooperation Under the Security Dilemma," *World Politics* 30.2 (January 1978).
- Kaplan, Morton, *System and Process in International Politics* (New York: John Wiley & Sons, Inc., 1957).
- King, Gary, Keohane, Robert, and Verba, Sidney, *Designing Social Inquiry—Scientific Inference in Qualitative Research* (Princeton: Princeton University Press, 1994).
- Kissinger, Henry, *A World Restored: Metternich, Castlereagh and the Problems of Peace 1812-1822* (London: Victor Gollancz, 1973).
- Levy, Jack S. and Thompson, William R., "Hegemonic Threats and Great-Power Balancing in Europe, 1495-1999," *Security Studies* 14.1 (Jan.-March 1995).

- Levy, Jack S., "Power Transition Theory and the Rise of China," in Robert Ross and Zhu Feng, eds., *China's Ascent: Power, Security, and the Future of International Politics* (Ithaca, NY: Cornell University Press, 2008).
- Levy, Jack S. and Thompson, William R., "Balancing at Sea: Do States Ally against the Leading Global Power?", paper presented at the MIT Security Studies Program seminar, Feb. 24, 2010.
- Lobell, Steven E., Ripsman, Norrin M., and Taliaferro, Jeffrey W., eds., *Neoclassical Realism, the State, and Foreign Policy* (Cambridge University Press, 2009).
- McNeill, William H., *The Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000* (Chicago: University of Chicago Press, 1982).
- Mearsheimer, John, *The Tragedy of Great Power Politics* (New York: W.W. Norton, 2001).
- Mill, John Stuart, *A System of Logic—Ratiocinative and Inductive* (Toronto: University of Toronto Press, Routledge, 1978).
- Morgenthau, Hans J., *Politics Among Nations: The Struggle for Power and Peace* (New York: Alfred A. Knopf, 2nd Ed., 1956).
- Morgenthau, Hans J., *Politics Among Nations: The Struggle for Power and Peace* (New York: McGraw Hill, Brief Ed., 7th Reprint, 1993).
- Nye, Joseph, *Understanding International Conflicts* (New York: Pearson Longman, 5th edition, 2005).
- Organiski, A.F.K. and Kugler, Jacek, *The War Ledger* (Chicago: University of Chicago Press, 1980).
- Organski, A.F.K., *World Politics* (New York: Alfred A. Knopf, 1988).
- Pape, Robert, "Soft Balancing Against the United States," *International Security* 30.1 (Summer 2005).
- Paul, T.V., "The Enduring Axioms of Balance of Power Theory and their Contemporary Relevance," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004).
- Plagge, Arnd, "Patterns of Deception: Why and How Rising States Cloak their Power," paper presented at Yale University's International Security Studies Colloquium in International History and Security (Sept. 29, 2009).
- Ragin, Charles C., Berg-Schlosser, Dirk, and de Meur, Gisèle, "Political Methodology: Qualitative Methods," in Robert E. Goodin and Hans-Dieter Klingemann, eds., *A New Handbook of Political Science* (Oxford: Oxford University Press, 1998).
- Reiter, Dan, "Learning, Realism, and Alliances: The Weight of the Shadow of the Past," *World Politics* 46.4 (July 1994).
- Rock, Stephen, *Why Peace Breaks Out: Great Power Rapprochement in Historical Perspective* (Chapel Hill: University of North Carolina Press, 1989).
- Rose, Gideon, "Neoclassical Realism and Theories of Foreign Policy," *World Politics* 51.1 (Oct. 1998).
- Rosecrance, Richard and Stein, Arthur, eds., *The Domestic Bases of Grand Strategy* (Ithaca: Cornell University Press, 1993).
- Rosecrance, Richard, "Overextension, Vulnerability, and Conflict: the 'Goldilocks Problem' in International Strategy," *International Security* 19.4 (Spring 1995).

- Ross, Robert, "Bipolarity and Balancing in East Asia," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power. Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004).
- Russett, Bruce and Oneal, John, *Triangulating Peace: Democracy, Interdependence, and International Organizations* (New York: W.W. Norton, 2001).
- Schroeder, Paul, "Historical Reality v. Neo-Realist Theory," *International Security* 19.1 (Summer 1994).
- Schweller, Randall L., "Bandwagoning for Profit: Bringing the Revisionist State Back In," *International Security* 19.1 (Spring 1994).
- Schweller, Randall L., *Unanswered Threats: Political Constraints on the Balance of Power* (Princeton: Princeton University Press, 2006).
- Hui, Victoria Tin-bor, *War and State Formation in Ancient China and Early Modern Europe* (New York: Cambridge University Press, 2005).
- Vasquez, John A., "The Realist Paradigm and Degenerative versus Progressive Research Programs: An Appraisal of Neotraditional Research on Waltz's Balancing Proposition," in John A. Vasquez and Colin Elman, eds., *Realism and the Balance of Power: A New Debate* (Upper Saddle River: Prentice Hall, 2003).
- Walt, Stephen M., "Alliance Formation and the Balance of Power," *International Security* 9.4 (Spring 1985).
- Walt, Stephen M., *The Origins of Alliances* (Ithaca, NY: Cornell University Press, 1990).
- Walt, Stephen M., *Taming American Power. The Global Response to U.S. Primacy* (New York: W.W. Norton, 2005).
- Waltz, Kenneth N., "Theory of International Relations," in Fred Greenstein and Nelson Polsby, eds. *Handbook of Political Science Volume 8: International Politics* (Reading, MA: Addison Wesley Publishing Co., 1975).
- Waltz, Kenneth N., *Theory of International Politics* (Boston: McGraw Hill, 1979).
- Watson, Adam, *The Evolution of International Society: A Comparative Historical Analysis* (New York: Routledge, 1992).
- Wendt, Alexander, *Social Theory of International Politics* (Cambridge: Cambridge University Press, 1999)..
- Wight, Martin, "The Balance of Power," in Herbert Butterfield and Martin Wight, eds., *Diplomatic Investigations: Essays in the Theory of International Politics* (Cambridge: Harvard University Press, 1996).
- Wohlforth, William, Kaufman, Stuart, and Little, Richard, "Introduction: Balance and Hierarchy in International Systems," in William Wohlforth, Stuart Kaufman, and Richard Little, eds., *The Balance of Power in World History* (New York: Palgrave MacMillan, 2007).

Chapters 3 and 4: The Mongol Conquests

Primary Sources (by original language)

1. Mongol/Uighur sources:

- Anonymous (most likely a member of Genghis Khan's Borjigin clan), *Yüan Ch'ao pi-shih* or *Tobchi'an* (ca. 1228), transl. into English (*The Secret History of the Mongols*) by Urgunge Onon (Richmond, UK: Curzon Press, 2001).
This is the major source detailing the rise of Genghis Khan, specifically up to his unification of the Mongol tribes. The Mongol manuscript (written sometime after Genghis Khan's death likely in Uighurs characters) does not exist anymore; all modern translations are based on a 14th century Chinese transcript of the text.
- Sagan Setsen, *Erdeniin Tobchi (Precious Summary)* (ca. 1662), transl. into German (*Geschichte der Ost-Mongolen und Ihres Fürstenhauses*) by Isaac Jacob Schmidt (St. Petersburg: N. Gretsck, 1829; reprint by Europe Printing, The Hague, 1961).
Reputable Mongol chronicle of the life of Genghis Khan written in the 17th century by a Mongolian Prince of Ordos, translated into German by a Moravian missionary.

2. Chinese sources:

- Kuang, Ssu-ma et. al., *Tong Kien Kang Mou (Annals of the Empire)*, Vol. 9 (11th - 13th century), transl. into French (*Histoire Générale de la Chine ou Annales de Cet Empire*, Vol. 9) by Père Joseph-Anne-Marie de Moyriac de Mailla (Paris: Abbé Grosier, 1779; reprint by Ch'eng-wen Publishing Company, Taipei, 1969).
Very complete Chinese records in 12 volumes, started by Chinese historian Ssu-Ma Kuang (Sima Guang) in the 11th century, continued by Confucian scholar Chu Hsi (Zhu Xi) in the 12th century, and finished by others, translated into French by a Jesuit missionary in Beijing.
- Wang Kuo-wei, *Cheng-wou Ts'in-tcheng Lou (Tale of the Campaigns of Genghis Khan)* (ca. 1285), transl. into French (*Histoire des Campagnes de Gengis Khan*, Vol. 1) by Paul Pelliot and Louis Hambis (Leiden, Netherlands: E.J. Brill, 1951).
History of Genghis Khan written by a Chinese scholar under Khubilai Khan using historical documents available at the time, including a now lost Mongol chronicle contemporary to the *Secret History*, the *Altan Däbtär (Golden Book)*, which was presumably the official book of the Mongol Imperial family and is a key source because it is written in a more historical, and much less legendary, style than the *Secret History* (the *Altan Däbtär* was also consulted by Persian chronicler Rashid al-Dīn, below).
- Li, Chih-Ch'ang, *Ch'ang-Ch'un Chen-jen Hsi-yu-chi (The Journey of the Adept Ch'ang Ch'un to the West)* (ca. 1228), transl. into English (*The Travels of an Alchemist – The Journey of the Taoist Ch'ang Ch'un from China to the Hindukush at the Summons of Chingiz Khan*) by Arthur Waley (London: George Routledge & Sons, 1931; reprint by Greenwood Press, Westport, CT, 1976).
Tale of the travels of Taoist master Ch'ang Ch'un, who went to Genghis Khan's court at Samarkand and in Afghanistan at the Khan's request, between 1220 and 1223. Written and narrated by Ch'ang Ch'un's disciple Li Chih-Ch'ang, who accompanied him on the journey.

3. Persian sources:

- Juvaini, 'Ala-ad-Din 'Ata-Malik, *Ta'rīkh-i Jahān Gushā (The History of the World Conqueror)* (ca. 1252-3), transl. into English by John Andrew Boyle, 2 volumes (Manchester, UK: Manchester University Press, 1958).

Detailed account of the rise of the Mongol Empire by an official of the Mongol Empire, originally from Khwarezm, who accompanied Hulegu on the campaigns against the Ismaelites and the Caliphate and was then appointed Governor of Baghdad.

- Tabīb, Rachid al-Dīn, *Jami 'al-Tawarikh (Compendium of Chronicles)* (ca. 1307-1316), transl. into French (*Histoire des Mongols de la Perse*) by Etienne Marc Quatremère (Paris, 1836; reprint by Oriental Press, Amsterdam, 1968).
Historical chronicle with particular emphasis on Hulegu and the Ilkhan branch of the Mongols, who commissioned the work, written by a Persian physician and Grand Vizier to the Ilkhan court who was of Jewish descent.
- Juzjani, Minhaj al-Siraj, *Tabaqāt-i-Nāsirī* (ca. 1259), transl. into English (*A General History of the Muhammadan Dynasties of Asia, 2 Vol.*) by Major H.G. Raverty (New Dehli: Asiatic Society of Bengal, 1881; reprint by Oriental Books Reprint Corporation, New Dehli, 1970), Vol. 2.
Annals of Islam from its origins to A.D. 1259 written by an imam and judge originally from Khwarezm, who immigrated to the Sultanate of Dehli (India). It relies partly on earlier, and some lost, sources.

4. Arabic sources:

- Ibn al-Athīr al-Jazari, Ali 'Izz al-Dīn, *Al-Kāmil fī'l-Ta'rīkh (The Complete History)* (ca. 1231), partly transl. into English (*The Chronicle of Ibn al-Athīr for the Crusading Period, Parts 1, 2, and 3*) by D.S. Richards, Part 3: "The Years 589-629/1193-1231 – The Ayyubids after Saladin and the Mongol Menace" (Oxford: Ashgate, 2008).
A key Islamic historical work by an Arab scholar and historian, originally from the Syrian-Turkish border area, who lived in Mosul.
- Al-Nasawi, Shihāb al-Dīn Muhammad, *Sīrat al-sultān Jalāl al-Dīn Minkbrati* (ca. 1241), transl. into French (*Histoire du Sultan Djelal ed-Din Mankobirti, Prince du Kharezm*) by O. Houdas (Paris: Ernest Leroux, 1895).
Written by the secretary of Shah of Khwarezm Jelal ad-Din, whom he followed during the entire conflict with the Mongols – first-hand account of battles, but often distorted to present Jelal ad-Din under favorable light.

5. Armenian sources:

- Of Akanc', Grigor (a.k.a. Maghak'ia the Monk), *History of the Nation of the Archers – the Mongols* (ca. 1271), transl. into English by Robert P. Blake and Richard N. Frye (Cambridge, MA: Harvard University Press, 1954).
Chronicle of Armenian history at the time of the Mongol conquests written by an Armenian monk, based on oral history and possibly on the manuscript (which has not survived) of an eye-witness of the invasions.

6. Hungarian sources:

- Bela, King of Hungary, *Brief König Belas IV and den deutschen König Konrad IV (Letter from King Bela IV to German king Konrad IV)* (1241), transl. into German (*Der Mongolensturm – Berichte von Augenzeugen und Zeitgenossen 1235-1250*) by Hansgerd Göckenjan and James R. Sweeney (Graz, Austria: Verlag Styria, 1985).

7. Russian sources:

- Anonymous (most likely Russian monks), *The Chronicle of Novgorod* (ca. 1016-1471, by successive writers), transl. into English by Robert Mitchell and Nevill Forbes (London: Camden Society, 1914; reprint by Academic Intl., Hattiesburg, MS, 1970). Oldest surviving chronicle from the Republic of Novgorod, a vast principality in Northern Russia repeatedly attacked by the Mongols and the Golden Horde between the 13th and 15th century.

8. Syriac sources:

- Sauma, Rabban Bar, *History of Rabban Sâwmâ and Yahb-Allâhâ III* (ca. 1290-1294), transl. into English (*The Monks of Kûblâi Khân, Emperor of China or the History of the Life and Travels of Rabban Sâwmâ*) by E.A. Wallis Budge (London: The Religious Tract Society, 1928; reprint by AMS Press, New York, 1973). Travel memoirs of a Mongol diplomat of Nestorian (Christian) faith who was sent as an envoy to Western Europe by Khubilai Khan.

9. Latin sources:

- da Pian del Carpine, Giovanni (a.k.a. John de Plano Carpini), *Historia Mongalorum quos nos Tartaros Appellamus (History of the Mongols which we Call Tartars)* (ca. 1250), transl. into English by a nun of Stanbrook Abbey, UK and edited (*The Mongol Mission*) by Christopher Dawson (New York: Sheed and Ward, 1955). Record by one of the first Europeans to visit the Mongol court. Sent by Pope Innocent IV to the Mongol court from 1245 to 1247, he witnessed the kuriltai (election assembly) and coronation of Guyuk Khan near Karakorum. This text includes the letter of Innocent IV to Guyuk Khan, and Guyuk's response.
- de Rubruquis, Willielmi (a.k.a. Willian of Rubruck), *Itinerarium Fratris Willielmi de Rubruquis de Ordine Fratrum Minorum, Galli, Anno gratia 1253 ad partes Orientales (Journal of Friar William of Rubruck of the Order of the Minorite Friars into the East, 1253)* (ca. 1256), transl. into English by a nun of Stanbrook Abbey, UK and edited (*The Mongol Mission*) by Christopher Dawson (New York: Sheed and Ward, 1955). Report by a Flemish Franciscan missionary who, after participating to the 7th Crusade (1248) along Louis IX of France (St. Louis), was sent by the French king on a mission to the Mongol court of Mongke Khan, from 1253 to 1255.
- de Torre Maggiore, Rogerius de Apulia (a.k.a. Roger of Torre Maggiore), *Carmen Miserabile super Destructione Regni Hungariae per Tartaros (Sad Song about the Destruction of the Kingdom of Hungary by the Tartars)* (ca. 1243), transl. into German (*Der Mongolensturm – Berichte von Augenzeugen und Zeitgenossen 1235-1250*) by Hansgerd Göckenjan and James R. Sweeney (Graz, Austria: Verlag Styria, 1985). Description of Batu's 1241 Hungarian campaign by a Catholic monk from Italy, future archbishop of Croatia, who was made prisoner by the Mongols and managed to escape.
- Parisiensis, Matthaei (a.k.a. Matthew Paris), *Chronica Majora* (ca. 1240-1253), edited by Henry Richard Luard, Rolls Series, Vol. 3 (London: Longman & Co., 1876). Illustrated chronicle of medieval history (1240-1253) by an English Benedictine monk at St. Albans, England.
- Dlugosz, Jan, *Annales Seu Cronici Incliti Regni Poloniae* (ca. 1480) (Annals or Chronicles of the Famous Kingdom of Poland), transl. into English (*The Annals of Jan Dlugosz*) and edited by Maurice Michael (IM Publications, 1997).

Thirteenth century chronicle by a Polish priest and ambassador relating the historical highlights of Eastern Europe from 965 to 1480.

10. French/Italian sources:

- Polo, Marco, *Milione – Le Divisament Dou Monde (The Description of the World, generally titled The Travels of Marco Polo)* (ca. 1298), original Old French and Italian versions published by Arnoldo Mondadori Editore, Milan, Italy, 1982. Transl. into English by Manuel Komroff (New York: Random House/The Modern Library, 1926). The various original manuscripts of Marco Polo's *Travels* were written in Italian and in French but none is similar. Komroff's edition attempts to fill these gaps and correct mistakes from previous translations.

Secondary Sources

- Allsen, Thomas T., *Mongol Imperialism: the Policies of the Grand Qan Möngke in China, Russia, and the Islamic Lands 1251-1259* (Berkeley: University of California Press, 1987).
- Blair, Sheila, "East Meets West under the Mongols," *The Silk Road Journal* 3.2 (Dec. 2005), available online at <http://www.silk-road.com/newsletter/vol3num2/>.
- Boyle, John Andrew, *The Mongol World Empire 1206-1370* (collection of articles) (London: Variorum Reprints, 1977).
- Chambers, James, *The Devil's Horsemen: The Mongol Invasion of Europe* (London: Phoenix Press, 1979).
- Christian, David, *A History of Russia, Central Asia, and Mongolia-Volume 1: Inner Eurasia from Prehistory to the Mongol Empire* (Oxford: Wiley Blackwell, 1998).
- Dalintai, Zhu, *Menggu Bingxue Yanjiu: Jian Lun Chengjisihan Yong Bing Zhi Mi* (Beijing: Di Ban, 1990), transl. by Urunge Onon (*The Mongol Art of War: the Strategy of Genghis Khan*) and published in *The Secret History of the Mongols* (Richmond, UK: Curzon Press, 2001).
- Grousset, René, *L'Empire des Steppes: Attila, Gengis-Khan, Tamerlan* (Paris: Payot, 1939).
- d'Ohsson, Mouradja, *Histoire des Mongols depuis Tchinguiz-Khan jusqu'à Timour Bey*, 4 Vol. (Amsterdam, 1824).
- Hoang, Michel, *Genghis Khan*, transl. from the French (*Gengis-Khan*) by Ingrid Cranfield (London: Saqi Books, 1990).
- Phillips, E.D., *The Mongols* (London: Thames and Hudson, 1969).
- Saunders, J.J., *The History of the Mongol Conquests* (London: Routledge & Paul Keegan, 1971).
- Turnbull, Stephen, *Genghis Khan and the Mongol Conquests 1190-1400* (Oxford: Osprey Publishing, 2003).
- Weatherford, Jack, *Genghis Khan and the Making of the Modern World* (New York: Crown Publishers, 2004).

Chapters 5 and 6: The Rise of Rome

Primary Sources

- Polybius. *The Rise of the Roman Empire* (transl. by Ian Scott-Kilvert) (New York: Penguin Books, 1979).
One of the most reputable sources for mid-Republican history. As a friend of consul Scipio Aemilianus, Achean historian Polybius (ca. 203-120 B.C.) personally witnessed many of the events of and surrounding the Third Punic War.
- Livy. *The War with Hannibal* (Books XXI-XXX of *The History of Rome From Its Foundation*, transl. by Aubrey de Selincourt) (New York: Penguin Books, 1965).
Pro-republican Civil Wars and early Principate historian (59 B.C -17 a.d.). Work is highly detailed but known for inaccuracies. Used mostly when confirmed by other sources.
- Livy. *Rome and the Mediterranean* (Books XXXI-XLV of *The History of Rome From Its Foundation*, transl. by Henry Bettenson) (New York: Penguin Books, 1976).
- Tacitus, Publius. *Anales (Annals, transl. by A.J. Woodman)* (Indianapolis: Hackett Publishing Company, Inc., 2004).
Early Principate provincial Roman senator, consul, governor, and literary author (ca. 56-118 A.D.).
- Appianus (Appian). *Roman History Vols. I-III* (transl. by Horace White) (Cambridge, MA: Harvard University Press, Repr. 1982).
Roman historian and lawyer of Greek ethnicity (ca. 95-175 A.D.). Reputable, particularly for Civil Wars era work.
- Vegetius Renuus, Publius Flavius. *De Re Militari (The Military Institutions of the Romans*, transl. by Lieut. John Clark, ed. By Brig. Gen. Thomas R. Phillips, U.S.A.) (Westport, CT: Greenwood Press Publishers, Repr. 1985).
Roman military writer from the late Roman Empire, dates unknown. Compilation of various sources; known for inconsistencies.
- Josephus, Titus Flavius. *The War of the Jews* (transl. by G.A. Williamson) (New York: Penguin Books, Repr. 1970).
Early Principate Jewish historian (ca. 37-100 a.d.) who witnessed and documented the Jewish rebellion and the sack of Jerusalem by the Romans.
- Cesar, Gaius Julius. *Commentariorum De Bello Gallico Liber Primus (Gallic War*, transl. by W. A. McDevitte and W. S. Bohn) (New York: Harper & Brothers, 1st Ed., 1869).
Roman general, proconsul of Gaul, and dictator (100-44 B.C.). Participated in the Civil Wars and the downfall of the Republic. Detailed account of his conquests.
- Frontinus, Sextus Julius. *Stratagemata (The Stratagems*, transl. by R.I. Ireland) (Leipzig: Teubner, 1990).
Roman engineer, praetor and governor of Britain (ca. 40-103 a.d.). Vast knowledge of Roman public works.

Secondary Sources

- Adcock, F.E. *The Roman Art of War Under the Republic* (Cambridge: W. Heffer & Sons Ltd., 5th Ed. 1981, or. Harvard University Press, 1940).
- Amit, Moché. "Les Moyens de Communication et la Défense de l'Empire Romain." *La Parola del Passato – Rivista di Studi Antichi* Vol. 20 Issue 102 (1965).

- Badian, E. *Foreign Clientelae (264-70 B.C.)* (Oxford, UK: Oxford University Press, 1958).
- Badian, E. *Publicans and Sinners: Private Enterprise in the Service of the Roman Republic* (Ithaca, NY: Cornell University Press, 1972).
- Baradez, Jean L. "Organisation Militaire Romaine de l'Algérie Antique et l'Evolution du Concept Défensif de ses Frontières." *Revue Internationale d'Histoire Militaire* No. 13 (1953).
- Biddle, Stephen, *Military Power: Explaining Victory and Defeat in Modern Battle* (Princeton, NJ: Princeton University Press, 2004).
- Charlesworth, Martin Percival. *Trade-Routes and Commerce of the Roman Empire* (Chicago: Ares Publishers Inc., 2nd Ed., 1926).
- Chevallier, Raymond. *Roman Roads* (Berkeley, CA: University of California Press, 1976).
- Collingwood, R.G. "The Purpose of the Roman Wall." *The Vasculum* Vol. 8 (1920).
- Cornell, T.J. *The Beginnings of Rome: Italy and Rome from the Bronze Age to the Punic Wars (c. 1000-264 B.C.)* (London: Routledge, 1995).
- Couissin, Paul. *Les Armes Romaines: Essai sur les Origines et l'Evolution des Armes Individuelles de Légionnaire Romain* (Paris: Librairie Ancienne Honoré Champion, Ed., 1926).
- Eadie, John W. "The Development of Roman Mailed Cavalry." *The Journal of Roman Studies* Vol. 57 No. 1/2 (1967).
- Eckstein, Arthur M. *Mediterranean Anarchy, Interstate War, and the Rise of Rome* (Berkeley, CA: University of California Press, 2006).
- Errington, R.M. *The Dawn of Empire: Rome's Rise to World Power* (Ithaca, New York: Cornell University Press, 1972).
- Ferrill, Arther. *Roman Imperial Grand Strategy* (Lanham, MD: Publications of the Association of Ancient Historians 3—University of America Press, 1991).
- Goldsworthy, Adrian Keith. *The Roman Army at War 100 B.C.-A.D. 200* (Oxford: Clarendon Press, 1996).
- Goldsworthy, Adrian Keith. *The Complete Roman Army* (London: Thames & Hudson Ltd., 2003).
- Harkness, Albert. *The Military System of the Romans* (New York: D. Appleton and Company, 1887).
- Harris, William V. *War and Imperialism in Republican Rome 327-70 B.C.* (New York: Oxford University Press, 1979).
- Lintott, Andrew. *Imperium Romanum: Politics and Administration* (New York: Routledge, 1993).
- Luttwak, Edward N. *The Grand Strategy of the Roman Empire From the First Century A.D. to the Third* (Baltimore, MD: Johns Hopkins University Press, 1976).
- MacMullen, Ramsay. "Notes on Romanization." *Bulletin of the American Society of Papyrologists* Vol. 21 Issue 1-4 (1984).
- MacMullen, Ramsay. "How to Revolt in the Roman Empire." *Rivista Storica dell'Antichità* Vol. 15 (1985).
- Mommsen, Theodor. **Römische Geschichte – Erster Band: Bis zur Schlacht von Pydna, Drittes Buch** (Berlin: Weidmannsche Buchhandlung, 8th Ed., 1888).

- Piganiol, André. "La Notion de Limes." *International Congress of Roman Frontier Studies* 5th (1961).
- Prevas, John. *Hannibal Crosses the Alps: the Invasion of Italy and the Punic Wars* (Cambridge: Perseus/Da Capo Press, 1998).
- Ramsay, A.M. "The Speed of the Roman Imperial Post." *The Journal of Roman Studies* Vol. 15 (1925).
- Sands, Percy Cooper. *The Client Princes of the Roman Empire under the Republic* (New York: Arno Press, 1975).
- Sherwin-White, A.N. *The Roman Citizenship* (Oxford, UK: Oxford University Press, 2nd Ed., 1973).
- Smith, William. *A Dictionary of Greek and Roman Antiquity* (London: John Murray, 1875).
- Starr, Jr., Chester G. *The Emergence of Rome as Ruler of the Western World* (Ithaca, NY: Cornell University Press, 1950).
- Stevenson, G.H. *Roman Provincial Administration Till the Age of the Antonines* (Oxford, UK: Basil Blackwell, 2nd Ed., 1949).
- Watson, G.R. *The Roman Soldier* (Ithaca, NY: Cornell University Press, 1969).
- Webster, Graham. *The Roman Imperial Army of the First and Second Centuries A.D.* (University of Oklahoma Norman Press, 1998).
- Wells, Colin. *The Roman Empire* (Stanford, CA: Stanford University Press, 1984).

Chapter 7: The Qin Unification of China

Primary Sources

- *Chan-kuo Ts'e (Intrigues of the Warring States)*, transl. and ed. by J.I. Crump, *Michigan Monographs in Chinese Studies Series 77* (Ann Arbor, MI: The University of Michigan Press, 1996).
- Chia Yi, "The Faults of Ch'in," transl. by Burton Watson, in Cyril Birch, ed., *The Anthology of Chinese Literature – Vol. 1: From Early Times to the Fourteenth Century* (New York: Grove Press, 1994).
- *Guliang Zhuan (The Commentary of Guliang on the Spring and Autumn Annals)*, attr. Guliang Xi or his disciple Zi-Xia, ca. 3rd cent. A.D., in *Harvard-Yenching Institute Sinological Index Series* (Beijing: Harvard-Yenching Institute, 1937).
- Han Fei Tzu, *Basic Writings*, ca. mid-3rd cent. B.C., transl. Burton Watson (New York: Columbia University Press, 1964).
- Sima Qian, *Records of the Grand Historian*, ca. 109-91 B.C., transl. Burton Watson (Hong Kong: The Chinese University of Hong Kong and Columbia University Press, 1993).
- *The Book of Lord Shang: A Classic of the Chinese School of Law*, attr. Shang Yang, ca. mid-4th cent. B.C., transl. and ed. by J.J.L. Duyvendak (Chicago: University of Chicago Press, 1963).
- *Yün-meng Manuscripts*, in Anthony François Paulus Hulsewé, ed., *Remnants of Ch'in Law: An Annotated Translation of the Ch'in Legal and Administrative Rules of the 3rd*

Century B.C., Discovered in Yün-meng Prefecture, Hupei Province, in 1975 (Leiden: E.J. Brill, 1985).

Secondary Sources

- Bau, Tzong-Ho, *The Stability of International Systems: A Study of the Warring States System of Ancient China* (Ph.D. diss., The University of Texas at Austin, 1986).
- Bodde, Derk, *China's First Unifier—A Study of the Ch'in Dynasty as Seen in the Life of Li Ssu* (Leiden: E.J. Brill, 1938).
- Chen, Samuel Shih-Tsai, "The Equality of States in Ancient China," in *The American Journal of International Law* 35.4 (Oct. 1941).
- Fairbank, John K., "Introduction: Varieties of the Chinese Military Experience," in John K. Fairbank and Frank A. Kiernan, *Chinese Ways in Warfare* (Cambridge, MA: Harvard University Press, 1974).
- Hui, Victoria Tin-Bor, *War and State Formation in Ancient China and Early Modern Europe* (New York: Cambridge University Press, 2005).
- Keightley, David N., "Where Have All the Swords Gone? Reflections on the Unification of China," *Early China* 2 (1976).
- Kiser, Edgar and Cai, Yong, "War and Bureaucratization in Qin China: Exploring an Anomalous Case," in *American Sociological Review* 68.4 (Aug. 2003).
- Lewis, Mark Edward, "Warring States Political History," in Michael Loewe and Edward L. Shaughnessy, eds., *The Cambridge History of Ancient China – From the Origins of Civilization to 221 B.C.* (Cambridge, UK: Cambridge University Press, 1999).
- Lewis, Mark Edward, *The Early Chinese Empires: Qin and Han* (Cambridge, MA: The Belknap Press of Harvard University Press, 2007).
- Li, Dun J., *The Ageless Chinese: A History* (New York: Charles Scribner's Sons, 1978).
- Sawyer, Ralph D., *The Seven Military Classics of Ancient China*, transl. by Ralph D. Sawyer (Boulder, CO: Westview Press, 1993).
- Trousdale, William, "Where All the Swords Have Gone," *Early China* 3 (1977).
- Van Evera, Stephen, *Causes of War: Power and the Roots of Conflict* (Ithaca, NY: Cornell University Press, 1999).
- Wagner, Donald B., *Iron and Steel in Ancient China* (Leiden: E.J. Brill, 1996).
- Walt, Stephen, "Keeping the World 'Off-Balance': Self-Restraint and U.S. Foreign Policy," in G. John Ikenberry, ed., *America Unrivaled: The Future of the Balance of Power* (Ithaca, NY: Cornell University Press, 2002).
- Yates, Robin D.S., "Siege Engines and Late Zhou Military Technology," in *Explorations in the History of Science and Technology in China*, "Special Number of the 'Collections of Essays on Chinese Literature and History' (1982).

Chapter 8: U.S. Expansion into the Western Hemisphere

Primary Sources

- U.S. Dept. of State, *Papers Relating to the Foreign Relations of the United States – Transmitted to Congress Dec. 2, 1902* (Washington, D.C.: Government Printing Office, 1903).
- U.S. Dept. of the Navy - Office of Naval Intelligence, *The Spanish-American War: A Collection of Documents Relative to the Squadron Operations in the West Indies – Arranged by Rear Admiral Pascual Cervera Y Topete* (Washington, D.C.: Government Printing Office, 1899).
- German Foreign Ministry, *Die Grosse Politik der Europäischen Kabinette, 1871-1914: Sammlung der Diplomatischen Akten des Auswärtigen Amtes - hrsg. von Johannes Lepsius, Albrecht Mendelssohn Bartholdy, Friedrich Thimme* (Berlin: Deutsche Verlagsgesellschaft für Politik, 1922).
- McKinley, William and Everett, Marshall, *Exciting Experiences in our Wars with Spain and the Filipinos* (Chicago: The Educational Co., 1900).

Secondary Sources

- Art, Robert J., "The Influence of Foreign Policy on Seapower: New Weapons and Weltpolitik in Wilhelminian Germany," in Robert J. Art and Kenneth N. Waltz, eds., *The Use of Force: International Politics and Foreign Policy, Second Ed.* (Lanham, MD: University Press of America, Inc., 1983).
- Art, Robert J., *A Grand Strategy for America* (Ithaca, NY: The Century Foundation, Inc. and Cornell University Press, 2003).
- Berstein, Serge and Milza, Pierre, "L'Impérialisme Colonial," in Serge Berstein and Pierre Milza, eds., *Histoire du XIXè Siècle* (Paris: Hatier, 1996).
- Blaustein, Albert P., "The U.S. Constitution: America's Most Important Export," *Issues of Democracy* 9.1 (March 2004).
- Bourne, Kenneth, *Britain and the Balance of Power in North America 1815-1908* (Berkeley, CA: University of California Press, 1967).
- Finnemore, Martha, *The Purpose of Intervention* (Ithaca, NY: Cornell University Press, 2003).
- Gallagher, John and Robinson, Ronald, "The Imperialism of Free Trade," *The Economic History Review* VI.1, 2nd Series (1953).
- Gardner, Lloyd C., LaFeber, Walter F., and McCormick, Thomas J., *Creation of the American Empire: U.S. Diplomatic History* (Chicago, IL: Rand McNally & Company, 1973).
- Grenville, John A.S. and Young, George Berkeley, *Politics, Strategy, and American Diplomacy: Studies in Foreign Policy 1873-1917* (New Haven, CT: Yale University Press, 1966).
- Healy, David, *U.S. Expansionism: The Imperialist Urge in the 1890s* (Madison, Wisc.: The University of Wisconsin Press, 1970).
- Knowlton, Robert J., "The Early Influence of the United States Constitution in the Western Hemisphere: The Cases of the Mexican Constitution of 1824 and Bolivar's Ideas," *Discussion Paper No. 78*, University of Wisconsin-Milwaukee Center for Latin American and Caribbean Studies Publication (March 1988).

- LaFeber, Walter F., *The New Empire: An Interpretation of American Expansion 1860-1898* (Ithaca, NY: Cornell University Press, 1963).
- Layne, Christopher, "Kant or Cant: The Myth of the Democratic Peace," *International Security* 19.2 (August 1994).
- May, Ernest R., *Imperial Democracy: The Emergence of America as a Great Power* (Chicago, IL: Imprint Publications, 1961, repr. 1991).
- Mearsheimer, John J., *The Tragedy of Great Power Politics* (New York: W.W. Norton & Company, 2001).
- Munro, Dana Gardner, *The United States and the Caribbean Area* (Boston, MA: World Peace Foundation, 1934).
- Lippmann, Walter, *U.S. Foreign Policy: Shield of the Republic* (Boston, MA: Little, Brown, and Company, 1943).
- Perkins, Dexter, *The Monroe Doctrine 1867-1907* (Gloucester, MA: Peter Smith, 1966).
- Robinson, Ronald and Gallagher, John, *Africa and the Victorians: The Climax of Imperialism in the Dark Continent* (New York: St. Martins Press, 1961).
- Russett, Bruce and Oneal, John R., *Triangulating Peace: Democracy, Interdependence, and International Organizations* (New York: W.W. Norton & Company, 2001).
- Spykman, Nicholas John, *America's Strategy in World Politics: The United States and the Balance of Power* (Archon Books, 1942, repr. 1970).
- Waltz, Kenneth N., *Theory of International Politics* (Boston, MA: McGraw-Hill, 1979).
- Zakaria, Fareed, *From Wealth to Power: The Unusual Origins of America's World Role* (Princeton, NJ: Princeton University Press, 1998).

Chapter 9: Conclusion

- U.S. Department of Defense, *Annual Report to Congress: Military Power of the People's Republic of China* (Office of the Secretary of Defense, 2007).
- Eurostat, "World Market Share of High Tech Exports, Leading High Tech Trading Countries," (2006), available at http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/High-tech_statistics (accessed Aug. 7, 2010).
- International Monetary Fund, "World Economic Outlook Database" (April 2010), available at <http://www.imf.org/external/pubs/ft/weo/2010/01/weodata/index.aspx> (accessed Aug. 7, 2010).
- Archer, Christon I., et. al., *World History of Warfare* (Lincoln, NE: University of Nebraska Press, 2002).
- Art, Robert J., "Europe Hedges Its Security Bets," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004).
- Art, Robert J., "Correspondence: Striking the Balance," *International Security* 30.3 (Winter 2005/06).
- Ashley, Maurice, *The Age of Absolutism 1648-1775* (Springfield, MA: Merriam Company Publishers, 1974).

- Atkinson, C.T., "The War of the Spanish Succession," in Lord Acton, A.W. Ward, G.W. Prothero, and Stanley Leathes, eds., *The Cambridge Modern History Vol. 5: The Age of Louis XIV* (New York: The MacMillan Company, 1908).
- Brooks, Stephen G. and Wohlforth, William C., *World Out of Balance: International Relations and the Challenge of American Primacy* (Princeton, NJ: Princeton University Press, 2008).
- Elliott, J.H., *Imperial Spain 1469-1716* (London: Penguin, 2nd ed., 2002).
- Esdaile, Charles, *Napoleon's Wars: An International History 1803-1815* (New York, NY: Viking, 1997).
- Finnemore, Martha, *National Interests in International Society* (Ithaca, NY: Cornell University, 1996).
- Fisher, H.A.L., "The Codes," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907).
- Fletcher, Richard, *Moorish Spain* (Berkeley: University of California Press, 2003).
- Gill, Bates, "Chinese Arms Exports to Iran," *Middle East Review of International Affairs* 2.2 (May 1998).
- Grant, A.J., "The Government of Louis XIV (1661-1715)," in Lord Acton, A.W. Ward, G.W. Prothero, and Stanley Leathes, eds., *The Cambridge Modern History Vol. 5: The Age of Louis XIV* (New York: The MacMillan Company, 1908).
- Guillard, Anton, "France and Her Tributaries 1801-3," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907).
- Gulick, Edward Vose, *Europe's Classical Balance of Power: A Case History of the Theory and Practice of One of the Great Concepts of European Statecraft* (Wesport, CT: Greenwood Press, 1955).
- Gya, Giji, "ESDP and EU Mission Update," *European Security Review* 36 (December 2007).
- Hassall, Arthur, "The Foreign Policy of Louis XIV (1661-97)," in Lord Acton, A.W. Ward, G.W. Prothero, and Stanley Leathes, eds., *The Cambridge Modern History Vol. 5: The Age of Louis XIV* (New York: The MacMillan Company, 1908).
- Heckscher, Eli F., *The Continental System: An Economic Interpretation* (Oxford: Clarendon Press, 1922).
- Ikenberry, G. John, Mastanduno, Michael, and Wohlforth, William C., "Unipolarity, State Behavior, and Systemic Consequences," *World Politics* 61.1 (January 2009).
- Jervis, Robert, *Perception and Misperception in International Politics* (Princeton, NJ: Princeton University Press, 1976).
- Kamen, Henry, *Empire: How Spain Became a World Power 1492-1763* (New York: Harper Collins, 2003).
- Keohane, Robert O., *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton, NJ: Princeton University Press, 1984).
- Keohane, Robert O. and Martin, Lisa, "The Promise of Institutional Theory," *International Security* 20.1 (Summer 1995).
- Knecht, Robert J., *The French Religious Wars 1562-1598* (Wellingborough, UK: Osprey Publishings, 2002).

- Krauthammer, Charles, "The Unipolar Moment," *Foreign Affairs* 70.1 (Winter 1990-1991).
- Layne, Christopher, "The Unipolar Illusion: Why New Great Powers Will Arise," *International Security* 17.4 (Spring 1993).
- Lieber, Keir A. and Press, Daryll G., "U.S. Nuclear Primacy and the Future of the Chinese Deterrent," *China Security* 5 (Winter 2007).
- Lynn, John Albert, *The Wars of Louis XIV 1667-1714* (Harlow, UK: Addison Wesley Longman, 1999).
- McNeill, William H., *The Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000* (Chicago, IL: University of Chicago Press, 1982).
- Mearsheimer, John, "The False Promise of International Institutions," *International Security* 19.3 (Winter 1994-5).
- Mettam, Roger, *Power and Faction in Louis XIV's France* (Oxford: Blackwell Publishers, 1988).
- Oren, Ido, "The Subjectivity of the 'Democratic' Peace: Changing U.S. Perceptions of Imperial Germany," *International Security* 20.2 (Autumn 1995).
- Ostwald, Jamel, *Vauban Under Siege: Engineering Efficiency and Martial Vigor in the War of the Spanish Succession* (Leiden, NL: Brill, 2007).
- Pape, Robert, "Soft Balancing against the United States," *International Security* 30.1 (Summer 2005).
- Pariset, Georges, "The Consulate 1799-1804," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907).
- Georges Pariset, "France under the Empire 1804-14," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907).
- Parker, Geoffrey, *The Dutch Revolt* (London: Penguin, Revised Edition, 1990).
- Parker, Geoffrey, *The Grand Strategy of Philip II* (New Haven, CT: Yale University Press, 2000).
- Paul, T.V., "The Enduring Axioms of Balance of Power Theory and their Contemporary Relevance," in T.V. Paul, James Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford: Stanford University Press, 2004).
- Posen, Barry R., "Command of the Commons: The Military Foundation of U.S. Hegemony," *International Security* 28.1 (Summer 2003).
- Posen, Barry R., "European Union Security and Defense Policy: Response to Unipolarity?" in *Security Studies* 15.2 (April-June 2006).
- Posen, Barry R., "From Unipolarity to Multipolarity: Transition in Sight?" (forthcoming chapter, April 2010 draft).
- Ross, Robert S., "Bipolarity and Balancing in East Asia," in T.V. Paul, James J. Wirtz, and Michael Fortmann, eds., *Balance of Power: Theory and Practice in the 21st Century* (Stanford, CA: Stanford University Press, 2004).
- Ruggie, John G., "Multilateralism: The Anatomy of an Institution," *International Organization* 46.3 (Summer 1992).
- Russett, Bruce and Oneal, John, *Triangulating Peace: Democracy, Interdependence, and International Organizations* (W.W. Norton & Company, 2001).

- Scoville, Warren C., *The Persecution of Huguenots and French Economic Development 1680-1720* (Berkeley, CA: University of California Press, 1960).
- Spiro, David E., "Give Democratic Peace a Chance? The Insignificance of the Liberal Peace," *International Security* 19.2 (Autumn 1994).
- Von Clausewitz, Carl, *The Campaign of 1812 in Russia* (Cambridge, MA: Da Capo Press, 1995).
- Walt, Stephen M., *Taming American Power: The Global Response to U.S. Primacy* (New York: W.W. Norton & Company, 2005).
- Walt, Stephen M., "Alliances in a Unipolar World," *World Politics* 61.1 (January 2009).
- Waltz, Kenneth N., "Structural Realism after the Cold War," in G. John Ikenberry, *America Unrivaled: The Future of the Balance of Power* (Ithaca: Cornell University Press, 2002).
- Waltz, Kenneth N., "Evaluating Theories," in John Vasquez and Colin Elman, eds., *Realism and the Balance of Power: A New Debate* (Upper Saddle River: Prentice Hall, 2003).
- Wickham-Legg, L.G., "The Concordats," in Lord Acton, Adolphus W. Ward, George W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907).
- Wilson, H.W., "The Command of the Sea 1803-15," in Lord Acton, Adolphus W. Ward, G.W. Prothero and Stanley Leathes, eds., *The Cambridge Modern History Vol. IX: Napoleon* (London: Cambridge University Press, 1907).
- Wohlforth, William C., "The Stability of a Unipolar World," *International Security* 24.1 (Summer 1999).
- Wolfe, John B., *Louis XIV* (New York: W.W. Norton & Company, 1968).