



THE HANDBOOK OF HISTORICAL ECONOMICS

EDITED BY

ALBERTO BISIN

GIOVANNI FEDERICO



The Handbook of Historical Economics

This page intentionally left blank

The Handbook of Historical Economics

Edited by

Alberto Bisin

Giovanni Federico



ACADEMIC PRESS

An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1650, San Diego, CA 92101, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Copyright © 2021 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-0-12-815874-6

For information on all Academic Press publications
visit our website at <https://www.elsevier.com/books-and-journals>

Publisher: Brian Romer
Editorial Project Manager: Liz Heijkoop
Production Project Manager: Punithavathy Govindaradjane
Designer: Matthew Limbert

Typeset by VTeX



Contents

Merger or acquisition? An introduction to <i>The Handbook of Historical Economics</i>	xv
Alberto Bisin and Giovanni Federico	

PART 1 WHAT IS HISTORICAL ECONOMICS

CHAPTER 1	The economic history of economic history: the evolution of a field in economics	3
	Robert A. Margo	
CHAPTER 2	The two revolutions in economic history	17
	Martina Cioni, Giovanni Federico, and Michelangelo Vasta	
CHAPTER 3	History as evolution	41
	Nathan Nunn	

PART 2 SOURCES AND METHODS

CHAPTER 4	Historical data: where to find them, how to use them	95
	Paola Giuliano and Andrea Matranga	
CHAPTER 5	The use of archaeological data in economics	125
	Andrea Matranga and Luigi Pascali	
CHAPTER 6	Ethnographic and field data in historical economics	147
	Sara Lowes	
CHAPTER 7	Historical Econometrics: Instrumental Variables and Regression Discontinuity Designs	179
	Felipe Valencia Caicedo	
CHAPTER 8	Historical natural experiments: bridging economics and economic history .	213
	Davide Cantoni and Noam Yuchtman	
CHAPTER 9	Persistence – myth and mystery	243
	Hans-Joachim Voth	
CHAPTER 10	LATE for history	269
	Alberto Bisin and Andrea Moro	
CHAPTER 11	Dynamic general equilibrium modeling of long and short-run historical events	297
	Gary D. Hansen, Lee E. Ohanian, and Fatih Ozturk	
CHAPTER 12	Money, banking, and old-school historical economics	335
	Eric Monnet and François R. Velde	

CHAPTER 13	Institutional change and institutional persistence	365
	Daron Acemoglu, Georgy Egorov, and Konstantin Sonin	
CHAPTER 14	How institutions and cultures change: an evolutionary perspective	391
	Samuel Bowles, Jung-Kyoo Choi, Sung-Ha Hwang, and Suresh Naidu	
CHAPTER 15	State power and conflict driven evolution	435
	David K. Levine and Salvatore Modica	
CHAPTER 16	Culture, institutions, and policy	463
	Torsten Persson and Guido Tabellini	
CHAPTER 17	Phase diagrams in historical economics: culture and institutions	491
	Alberto Bisin and Thierry Verdier	

PART 3 TOPICS

CHAPTER 18	The economic history of commodity market development	525
	Giovanni Federico	
CHAPTER 19	Why Africa is not <i>that</i> poor	557
	Ewout Frankema	
CHAPTER 20	Religion in economic history: a survey	585
	Sascha O. Becker, Jared Rubin, and Ludger Woessmann	
CHAPTER 21	Persistent failure? International interventions since World War II	641
	Matt Malis, Pablo Querubin, and Shanker Satyanath	
CHAPTER 22	The ancient origins of the wealth of nations	675
	Quamrul H. Ashraf, Oded Galor, and Marc Klemp	
CHAPTER 23	Social mobility in historical economics	719
	Gregory Clark	
CHAPTER 24	The Industrial Revolution and the Great Divergence: recent findings from historical national accounting	749
	Stephen Broadberry	
CHAPTER 25	Attitudes, aptitudes, and the roots of the great enrichment	773
	Joel Mokyr	
CHAPTER 26	The interplay among wages, technology, and globalization: the labor market and inequality, 1620-2020	795
	Robert C. Allen	
CHAPTER 27	Debt and taxes in eight U.S. wars and two insurrections	825
	George J. Hall and Thomas J. Sargent	
CHAPTER 28	Biogeography, writing, and the origins of the state	881
	David Stasavage	

CHAPTER 29	The wife's protector: a quantitative theory linking contraceptive technology with the decline in marriage	903
	Jeremy Greenwood, Nezih Guner, and Karen A. Kopeccky	
Index		945

This page intentionally left blank

State power and conflict driven evolution[☆]

15

David K. Levine^{a,b} and Salvatore Modica^c

^a*Department of Economics, RSCAS European University, San Domenico di Fiesole, Italy*

^b*Department of Economics, WUSTL, St. Louis, MO, United States*

^c*Department of Economics, Business and Statistics (SEAS), Università di Palermo, Palermo, Italy*

15.1 Introduction and historical contest

The goal of this chapter is to examine the implications of the evolution of social organizations due to external competition. There are a variety of models of external competition. Models such as Ely (2002) examine voluntary migration - these models tend to efficient outcomes as people are drawn to locations with high per capita income. Historically, however, institutional success has not been through voluntary immigration into the arms of welcoming richer neighbors. Rather people and institutions have generally spread through invasion and conflict: the Carthaginians did not emigrate to Rome. Large institutional change has often occurred in the aftermath of the disruption caused by warfare and other conflicts. Hence it seems worthwhile studying external competition through conflict, which is the focus of this chapter.

It is common to develop fact driven theories: a historical fact or laboratory anomaly is observed and a theory is introduced to explain that fact. Here we focus on using theory to analyze facts and particularly facts it was not designed to explain. The theory as indicated is external competition through conflict. The theory itself tells us what facts and data to look for.

In a dynamic setting of a game or mechanism in which punishments and rewards are possible most social arrangements arise as equilibrium - this finds sharp definition in the folk theorem of repeated games (see Fudenberg and Maskin (1986)) but is a much broader observation. The goal of evolutionary game theory is to ask which of these many feasible institutions and arrangements are persistent, which are durable, which will we see in the long-run. Here we preview our results.

15.1.1 Geography

The starting point of any evolutionary theory of conflict must be to specify the geographical area over which conflict takes place. There is no point in arguing about why historically India did not conquer China or vice versa - even a brief glance at the map show the Himalayan mountains in between, and a closer look shows as well the dense jungles of Southeast Asia separating the two. Hence a theory of conflict must focus on a region in which societies are able to fight. Over the sweep of recorded history

[☆] We gratefully acknowledge support from the EUI Research Council and MIUR (PRIN 2017H5KPLL).

the bulk of human population has lived in the Eurasian continent (including North Africa). There are three distinct geographical subregions which undoubtedly had trading relations but were not in conflict: China, India, and Europe (including North Africa). For geographical as well as cultural reasons, in each of them history shows different societies competing on a more or less level playing field, and most of our historical data is from these three regions treated separately.

None of these areas, however, is completely isolated, and in addition to the contestants there have always been annoying “outsiders:” Mongols, Afghans, Vikings, English and so forth. Each of these groups has in common that while they can and do play a military role in the region of conflict they themselves are largely immune from the consequences. Hence Mongols in their deserts, Afghans in their mountain strongholds, and Vikings and English protected by their seas are all able to raid and fight and even conquer in China, India, and (continental) Europe, but are themselves largely immune to invasion. As the strength of these outsiders does not depend on events in the central area of conflict we model them as exogenous; and as their strength waxes and wanes - due to, for example, to climate change and migration - we examine what impact this has on the evolution of societies in the main areas of conflict.

15.1.2 Hegemony

In conflict luck matters, but success begets success. Conquering a city or a province strengthens the winner and weakens the loser. Hence as war and invasion unfold we expect that eventually one side gets lucky enough, strong enough, and faces opponents weak enough, that it wins outright. In this sense the conquest and destruction of Carthage by Rome (or vice versa) was inevitable. This argues that the natural state of mankind is hegemony, with a single society ruling over an isolated geographical area: that is China, India, and (continental) Europe should naturally and commonly be found under the rule of one strong central state.

To a surprising extent this is true. The fact is that hegemonies have been much more frequent and long lasting than one may think: over roughly two millennia prior to the industrial revolution, in India, China, and Europe (comprising 85% of world population during the period) about 37% of the time hegemonies prevailed.¹ On the other hand while hegemonies were common in China, they were infrequent in Europe and non-existent in India.² We find the answer in the role of outsiders - and ultimately with the fully developed theory - we can explain well not only when we see a hegemony and when not, but also why in the absence of hegemony we sometimes see competing societies that are extractive and sometimes competing societies that are inclusive.

On the grounds it is better to walk before running, we begin by explaining the simple theory: here weak outsiders lead to hegemony, but there is little to say about the structure of non-hegemonic regions when outsiders are strong. With the simple theory evolution not surprisingly selects societies with high state power - the ability to prevail in conflict over rivals. Never-the-less this simple theory has significant implications which we outline next.

¹ See Levine and Modica (2013) for data and sources. Here we take the population weights of India and China as 1 and Europe as 1/2.

² For unclear reasons in all the theorizing about China versus Europe few have thought to ask how their theory fares in the third (and second largest) center of world population, India. We do.

15.1.3 Malthus

One of the most fundamental economic aspects of a society is its population and prosperity. Here Malthus casts a long shadow. It is an amazing triumph of theory over fact that such a thoroughly discredited theory still maintains its grip on the imagination of the economics profession. Malthusian theory asserts that in the long-run population will adjust so that society should be at a subsistence level of prosperity. In our reading of economic history we have been unable to identify any study of any society in any place or at any time for which this is the case. Modern anthropological research argues that ancient hunter-gatherer societies were more prosperous than subsequent agricultural societies - so presumably above subsistence.³ Ancient agricultural societies may have largely consisted of peasants at subsistence level - but also had elites well above subsistence so on average they must have been above subsistence. The entire history of the industrial revolution - the well known fact of the demographic transition - all starkly contradict Malthusian theory.⁴ Yet despite the fact that every observed society is above subsistence some economic historians continue to assume that any society for which data is lacking are at subsistence.⁵ Historically, the subsistence level meant “the physical requirements to survive and reproduce.” To deal with the obvious contradictions in the data modern economic historians such as Clark (2007) have introduced the slippery idea that “subsistence” means “some socially determined level of per capita income above which population decreases and below which it increases.”⁶ This is somewhat awkward as the cross-sectional evidence is clear that rich countries reproduce at much lower rates than poor ones.

Rather than tweaking an obviously wrong theory we instead ask what happens when social evolution is driven by conflict. As Malthus himself recognized that there can be incentive compatible social arrangements that stabilize the population at a low level, he still thought that in some long-run evolutionary sense these low-population equilibria were unstable. Evolution through conflict argues the opposite - that societies at subsistence level are unstable. Specifically, our intuition can be captured by the following conceptual experiment. Imagine a “Malthusian” society with farmers living at the edge of subsistence. Next door live their less numerous but richer neighbors who control their population. What happens when the few but rich neighbors invade the nearly starving farmers? For the farmers to spend time fighting is to take time from farming - that is, to starve. The outcome of this conflict is easy to see. In place of the discredited Malthusian theory arguing that every technological improvement should be met with population increase driving income back to subsistence, we develop a theory of when and how per capita income increases and decreases with technological change.

³ See, for example, Bowles et al. (2021).

⁴ See for example Hansen and Prescott (2002).

⁵ See for example Maddison (2007).

⁶ A close reading of the literature reveals serious problems in the way data is collected to support these theories. The most central problem is that at best what is computed is median per capita income - that is, the typical income of a poor person. Of course, as we noted, the upper classes consume considerably more than subsistence so the mean must also be above subsistence. A typical example of this problem is in the classical and much cited Ladorie (1974) study of Languedoc peasants in France. Ignored in this study are the facts that the nobles live above subsistence; that the entire area made substantial payments to the King - and indeed the ability of France to conduct continual wars throughout this period indicates that substantial resources above subsistence were available. More serious students of historical per capita GDP such as Maddison (2007) point out the Malthusian bias implicit in conclusions of this type.

15.1.4 Theory motivated analysis of facts

As indicated our goal is to use theory to motivate the analysis of facts. Hence we ask: given the theory what should be true? For example: Why should Spain now more than 300 years after ceding it to the British still wish to regain Gibraltar? Why would the UK be concerned to retain the Falkland Islands and Scotland? Why should Spain care about Catalan succession? Nobody thinks to answer these questions because their theories do not suggest these facts are relevant. Our theory prompts looking at these facts and suggests that the answer to these question lies in the nature of societies selected by evolutionary pressure. In particular, we show that those societies that simply seek to defend their territory are doomed to domination by those who aggressively seek expansion. Hence successful societies such as Spain and England have survived and thrived precisely because they are aggressive.

15.1.5 Gradual versus stochastic evolution

The economics profession is often accused by ignoramuses in the popular press of “physics envy.” In this imaginary world of economic theory we supposedly think of economies as following predictable trajectories like planetary orbits. Nowhere is the contrast with such supposed theories greater than in modern evolutionary game theory. In conflict driven evolutionary theory the economy does not gradually converge to a long-run state - the economy is dynamic and stochastic. Random events play a central role, and eventually bad luck leads to too many things going wrong at once.

As a case in point: hegemonies are persistent, but they also rise and fall. The dynamics of the fall is interesting. Our model predicts that a hegemony does not gradually deteriorate as imagined by historians such as Gibbon (1776). It is rather under constant attack: there are rebellions, invasions - many of them. Most fail, a few succeed for a long period of time and then also fail. And eventually when the fall comes it is abrupt. This happens when the hegemony is at low ebb - in a recession or depression or suffering some other sort of temporary setback from which it has recovered many times. Success driven by luck enables rebels or invaders to overthrow the existing hegemony before it has time to regroup and assert its strength. The picture is of many failed attempts, most short, a few long, followed by a sudden and rapid success.

This theoretical description then directs us to the data. Is it true? What is the nature of the fall of hegemony? To answer this we look for a well documented case study and find it in the story of the fall of the Qing dynasty in China. Indeed, as we show, the history of that fall has precisely the characteristics predicted by the theory.

15.1.6 Balance of power

After taking the first steps with a simple model we next analyze more carefully the role of strong outsiders. Here a non-hegemonic state of affairs - a balance of power - can persist for a long period. The point is a simple one: while there is a natural tendency for one side to win a conflict, this tendency is offset by outsiders whose interests may conflict with the existence of a hegemon and may therefore intervene on the weaker side, or who may simply take advantage of foreign commitments to stab the hegemon in the back. We examine this theory in two stages.

We first present the basic result which establishes that outside intervention prolongs conflicts. We examine a large dataset on interventions and modern conflict and show that indeed, when there is no outside intervention (think of the US Civil War, or the two World Wars) conflict ends relatively quickly

- in four or five years - with one side winning. By contrast when there is outside intervention conflicts drag on for decades or even centuries, and often there is no clear victor.

Next we develop a careful model of the role of social institutions in creating and resolving conflict, examining the motivations of elites and masses and their ability to influence policy under more or less inclusive institutions. On the basis of economic incentives we develop a theory that takes into account the strength of outsiders but also the advancement of military technology to understand when we are likely to see hegemony, when we are likely to see a balance of power and in case there is a balance of power whether the competing societies are likely to be inclusive or exclusive. We then apply this to two millennia of historical data from the three great regions of China, India, and Europe. The theory provides a clear and compelling accounting for the facts. If we accept as economic historians say that it was the development of competing inclusive institutions in Europe in the late Medieval period that was responsible for the subsequent industrial revolution, our theory provides a clear answer to the question: why Europe and not China or India.⁷ The answer we offer is Ghengis Khan, the cannon, and the English. Ghengis Khan accomplished two things: he brought cannons to Europe and he depopulated Mongolia. In China the removal of the outside threat of the Mongols led - as the theory asserts - to an extractive hegemony. In Europe the advent of the cannon together with the preying of the English on the French and the Spanish resulted - as the theory asserts - in an inclusive balance of power. Here as is often asserted the English played a key role in the industrial revolution - albeit a rather different one than is commonly assumed.

15.2 A model

We start by presenting a formal model which yields predictions about institutional change brought about by conflict driven evolution. We call our competing social organizations *societies*, of which there are a finite number $j \in J$. Time is discrete, $t = 1, 2, \dots$. A society can be *inactive*, representing a template for a possible form of social organization, or it can be *active*, controlling resources. We assume that there are a positive integral number of resources L representing land, capital, people and so forth. For simplicity we simply refer to this as land. An active society j controls a positive amount of land $L_{jt} > 0$ while an inactive society controls no land $L_{jt} = 0$. Evolution will take place as land changes hands due to conflict. Total resources are fixed: $\sum_j L_{jt} = L$ at all t .

In the simplest version of the theory each society has a fixed structure. It does not evolve internally and institutions do not change, but evolution takes place as societies gain and lose land through conflict and perhaps are driven into extinction. In this scheme evolution follows a Markov process. A state $z_t \in Z$ is a list of land holding, $z_t = (L_{1t}, L_{2t}, \dots, L_{Jt})$. External evolution is modeled as changes of ownership over land due to conflict. We assume that at most one unit of land changes hands each period - this can be taken to mean that time periods are "short enough."

We initially use a mechanical model of conflict. We assume that there is a simple scalar index of the potential ability of a social organization to resist and influence other societies. This depends on institutional characteristics of the society, including stability of the government and the ability to collect taxes and conscript soldiers, which depend in turn on the law-abidingness of citizens, the efficiency of

⁷ Not that anyone besides us tries to account for India.

the courts, and the overall economic strength of the society. We refer to this as “state power” and denote it by γ_j . The overall ability to prevail in conflict will depend upon state power, but also upon resources L_{jt} .

We recognize as well that institutions and behavior must be learned. At any moment of time there is no reason to assume that a society is internally in equilibrium. There are models of varying complexity of how this may work (see Levine and Modica (2013)) but here we take a very simple approach. We assume that there are two types of societies indexed by a stability index $b_j \in \{0, 1\}$ with 1 indicating stability. Stable societies are those with robust incentive compatible institutions and behavior. By contrast unstable societies represent short-term behavior that may not be sustainable but tend nonetheless to generate short-term power. To wit, a strong leader may inspire fanatical soldiers, but as they age fanatical soldiers tend to become less interested in fighting in sketchy environments and more interested in wine and song - and surely that is true of their descendants. Our formal assumption is that the strongest unstable society is stronger than the strongest stable society: specifically we assume that $\max_{j|b_j=0} \gamma_j > \max_{j|b_j=1} \gamma_j$.

As indicated our model of evolution is a Markov model - indexed, however, by a parameter ϵ which in this context should be thought of as “the importance of luck in warfare.” Specifically there are transition probabilities $P_\epsilon(z_{t+1}|z_t)$. We are going to be interested in the case where ϵ is small, which will have the interpretation that a very weak society will have little chance of prevailing over a strong one. This enables us to make use of the mathematical concept of resistance. Specifically we make three basic assumptions about P_ϵ :

1. P_ϵ is ergodic
2. There exists $\lim_{\epsilon \rightarrow 0} P_\epsilon = P_0$
3. For all $z_{t+1}, z_t \in Z$ there is a resistance function $0 \leq r(z_t, z_{t+1}) \leq \infty$ and constants $0 < C < 1 < D < \infty$ such that $C\epsilon^{r(z_t, z_{t+1})} \leq P_\epsilon(z_{t+1}|z_t) \leq D\epsilon^{r(z_t, z_{t+1})}$.

Here $r(z_t, z_{t+1})$ is the *resistance* of the transitions from z_t to z_{t+1} ; observe that zero resistance means strictly positive probability for all ϵ hence positive probability with respect to P_0 ; on the other hand infinite resistance is zero probability in all P_ϵ 's. Resistance measures speed of convergence of $P_\epsilon(z_{t+1}|z_t)$ to zero; roughly speaking higher resistance means lower probability. In the final section we will see how such a resistance function can emerge from a model in which conflict is endogenous.

Our assumptions about conflict are stated in terms of resistances. The probability that society j loses a unit of land is given by a conflict resolution function with resistance $r_j(z) < \infty$ to j losing one unit of land - this is the resistance of a transition from z to a state where j has one less unit of land. If j loses land the probability the land goes to society k has *land gain resistance* $\lambda_{jk}(z)$ where $\lambda_{jj} = \infty$ but if $k \neq j$ then $\lambda_{jk} < \infty$. Note that when an active society loses a unit of land new institutions may be introduced in that land - that is to say, an inactive society may gain the land.

We make several specific assumptions about the conflict resolution and land gain resistance.

1. r and λ depend only on the land holdings, state power and stability of the different societies; since they are just templates for societies, we assume that the state power of inactive societies does not matter.
2. resistance to losing land r is monotone in the sense that the resistance of j losing land is increasing in its own state power and land, and decreasing in that of other societies.
3. for stable societies resistance to losing land r is strictly increasing in state power and land when non-zero.

4. the weakest stable society with positive land holding has zero resistance to losing a unit of land.
5. unstable societies have zero resistance to losing a unit of land.
6. for given land holding, resistance to losing land is greater when facing more than one opponent with positive land holding than when all enemy land is in the hands of the strongest land holding opponent - also this is strict if resistance is positive.
7. for land gain resistance active societies all have zero resistance to gaining land, that is if $k \neq j$ and $L_k > 0$ then $\lambda_{jk}(z) = 0$.

Our last assumption involves *outsiders* - forces which are not societies in J but which may still influence the evolution of the z_t process. Outsiders are assumed to be protected by geography, climate, or sheer strength from action by the region in question. One example is Great Britain with respect to continental Europe from roughly the 14th to 20th centuries. Currently the U.S. and Russia are outside forces with respect to the Middle East, being protected by distance, the ocean (in the case of the U.S.) and by military strength from Middle Eastern societies. The strength of the outsiders is modeled simply as a real number $0 \leq y \leq 1$. The idea is that outsiders are antithetical to concentration of power in J . To state this formally, for any stable society j let z_j be the *hegemonic* state in which $L_j = L$ and define the *hegemonic resistance* as $h(\gamma_j, y) = r_j(z_j)$. Under our assumptions $h(\gamma_j, y)$ is strictly increasing in γ_j when positive. We postulate that h depends on y as follows:

8. $h(\gamma_j, y)$ is continuous and strictly decreasing in y when positive. There is a \bar{y} sufficiently large that $h(\gamma_j, \bar{y}) = 0$ for all γ_j .

A stochastically stable state z is defined to be absorbing for $\epsilon = 0$ and such that the resistance of the unique limiting ergodic probability as $\epsilon \rightarrow 0$ is positive.⁸

15.2.1 A theory of hegemony

The next results, proven in Levine and Modica (2016), characterize the stochastically stable sets of the system.

Theorem 15.1. *If $y < \bar{y}$ then the stochastically stable states consist of hegemonies of stable societies with maximal state power. If $y > \bar{y}$ all states are positive recurrent for $\epsilon = 0$.*

This theory says we should often see hegemony. The idea of history being dominated by hegemonic states may seem a strange one, but the fact is that hegemonies have been much more frequent and long lasting than one may think: over roughly two millennia prior to the industrial revolution, in India, China, and Europe (comprising 85% of world population during the period) about 37% of the time hegemonies prevailed.⁹

Next, letting μ^ϵ denote the ergodic probability of P^ϵ (as in footnote 8), the next result says that hegemonies with higher state power (and hence higher hegemonic resistance) are more likely (hence more persistent).

⁸ It is proven in Young (1993) that the ergodic probability μ^ϵ (such that $\mu^\epsilon P^\epsilon = \mu^\epsilon$) has a unique limit μ^0 as $\epsilon \rightarrow 0$; formally, a state z is stochastically stable if $\mu^0(z) > 0$.

⁹ See Levine and Modica (2013) for data and sources. Here we take the population weights of India and China as 1 and Europe as 1/2.

Theorem 15.2 (Hegemonies with more state power are more persistent). *Let z_j and z_k be hegemonic states of j and k respectively. Then $\mu^\epsilon(z_j)/\mu^\epsilon(z_k) = 1/\epsilon^{h_j-h_k}$.*

15.3 The fall of Malthus¹⁰

The theory so far is one of evolution that favors higher state power. Our first goal is to examine what this implies about the organization of society. In general in order to disrupt neighbors or defend against disruption it is important to have resources that are not being used for other purposes: we refer to these as *free resources*. We first explore this idea in a setting with endogenous population and show how this leads to a theory of endogenous population size. In this discussion we focus on a single society hence we drop the j subscript.

In the *Malthusian Game* there are N families $i = 1, \dots, N$ and each family chooses family size $M^i \in \{1, 2, \dots, M\}$. Families utility comes from their children: utility is taken equal to family size. However, the society may act collectively to choose a social norm $v \in \{1, 2, \dots, M\}$ for the maximum size of families and impose a penalty P on any family that fails to adhere to the social norm. If $v \geq M - P$ the society is incentive compatible and family size will be v ; otherwise family size will be M . Define $\pi = Nv$ if $v \geq M - P$ and $\pi = MN$ otherwise. We take π as the immutable characteristic defining the society. Our goal is to relate state power γ to π .

Here there are many incentive compatible social norms: some with large populations and some with small. In real societies, long before the advent of birth control, population was controlled - largely, of course, by abstinence from intercourse. One possible implementation of this is: women are limited to a certain number of children, and anyone who attempts to violate the norm is put to death along with her children. As extreme as it may appear, in practice societies often used methods not so different than this. Marriage was limited and delayed through requirements of substantial accumulation of capital or side-payments as a prerequisite to get married, and unwed mothers were severely punished, in many cases through capital punishment. This seems to be understood by demographic historians such as Bacci (2006). We will show that evolution through conflict favors neither the poorest nor the richest societies, but those which generate in aggregate the greatest level of income above subsistence.

Our first economic assumption is that population produces output according to a production function $\alpha Y(\pi)$ where α is a scalar technology parameter. We assume that $Y(\pi)$ is smooth that $Y(0) = 0$ and $Y'(0) > 0$. We assume decreasing returns in the sense that when $Y'(\pi) \neq 0$ that $Y''(\pi) < 0$. As it seems compelling that only so many people can fit on a particular plot of land before production becomes impossible due to overcrowding we assume that output is bounded above and denote by \bar{Y} the least such upper bound. We allow but do not require that for large enough π output may actually be decreasing due to congestion.

Our second economic assumption is that output is needed to sustain the population so we assume that there is a subsistence level which we normalize to 1. That is, to sustain a population of π requires at least that amount of output. Here we eschew the modern slippery concepts of subsistence as used, for example, by Clark (2007) in favor of the traditional notion of subsistence as requirements for survival and reproduction.

¹⁰ Based on Levine and Modica (2013).

Note that since population is at least $\pi \geq M$ for sufficiently small α it will be impossible to meet the subsistence requirement. In particular if $\alpha\bar{Y} < M$ subsistence is impossible. Hence we define $\underline{\alpha} > 0$ to be the smallest value of α such that $\max_{\pi} \alpha Y(\pi) - \pi \geq 0$. We assume hereafter that $\alpha \geq \underline{\alpha}$.

Free resources are resources in excess of subsistence, that is $\alpha Y(\pi) - \pi$. As these can be used for warfare, assume that they give state power γ directly: $\gamma = \alpha Y(\pi) - \pi$. Then (unless outsiders are very strong) the long-run stochastically stable outcome will maximize free resources. By contrast recall the usual Malthusian case where population is so large that income per capita is at subsistence level, that is the value of π for which output per capita settles at subsistence level for each α : $\alpha Y(\pi)/\pi = 1$.

We can now highlight how different is evolution through conflict from Malthusian evolution. In Malthusian theory technological change is always squandered as population grows pushing per capita output back to subsistence. Hence improved technology in the long-run leaves per capita income unchanged and only leads to an increase in population. In our theory we have

Proposition 15.1. *The population size $\hat{\pi}$ that maximizes free resources is non-decreasing in α . As $\alpha \rightarrow \infty$, however, per capita output $Y(\hat{\pi})/\hat{\pi}$ grows without bound.*

The first result is a Malthusian one: improved technology does increase population. The second result is anti-Malthusian - population does not grow so much as to keep the population poor: sufficiently great technological improvement will increase per capita output. In the special case in which there is a satiation level for output, that is $Y(\bar{\pi}) = \bar{Y}$ the latter result is easy to establish: once the upper bound on population is reached there is no point in adding more people regardless of the state of technology. The only way to take advantage of improved technology to get more free resources is through increased per capita output.¹¹

Proof. For the first result the value $\hat{\pi}$ is defined implicitly by the equation $\alpha Y'(\hat{\pi}) - 1 = 0$, so $d\hat{\pi}/d\alpha = -Y'(\hat{\pi})/\alpha Y''(\hat{\pi})$. This is positive for $Y''(\hat{\pi}) < 0$.

For the second result, we have already dealt with the case where $\hat{\pi}$ is bounded above by $\bar{\pi}$. Suppose instead that $\hat{\pi} \rightarrow \infty$. If this is to be possible it must be that $Y(\pi) < \bar{Y}$ for all π so $\bar{Y} - Y(\pi) \rightarrow 0$ as $\pi \rightarrow \infty$ (by definition of \bar{Y}).

Observe that $\alpha Y(\hat{\pi}) - \hat{\pi} \geq \alpha Y(\hat{\pi}/2) - \hat{\pi}/2$ so that $\hat{\pi} \leq 2\alpha (Y(\hat{\pi}) - Y(\hat{\pi}/2)) \leq 2\alpha (\bar{Y} - Y(\hat{\pi}/2))$. Hence

$$\alpha Y(\hat{\pi})/\hat{\pi} \geq \frac{Y(\hat{\pi})}{2(\bar{Y} - Y(\hat{\pi}/2))}.$$

Since $\hat{\pi}$ is non-decreasing so is $Y(\hat{\pi})$. As $\hat{\pi} \rightarrow \infty$ so $\hat{\pi}/2 \rightarrow \infty$, so $\bar{Y} - Y(\hat{\pi}/2) \rightarrow 0$ so it must be that $\alpha Y(\hat{\pi})/\hat{\pi} \rightarrow \infty$. \square

Although high levels of technology will result in increased per capita income, it is by no means the case that the relation needs to be monotone. A simple example makes the point. Suppose that near $\pi = M$ we have $Y(\pi) = A + (\pi/K) - (\pi/K)^2/2$ where $A > 0$ and $K > M$.¹² Then we may easily

¹¹ The mechanism here is not dissimilar to that discussed in Hansen and Prescott (2002), where the exhaustion of land forces a change to a capital based technology that increases per capita income.

¹² Note that this cannot hold globally.

solve for $\hat{\pi} = K(1 - \alpha^{-1})$. Then in this region output per capita is

$$\frac{Y(\hat{\pi})}{\hat{\pi}} = \frac{\alpha A}{K(1 - \alpha^{-1})} + (\alpha/2 + 1/2).$$

Differentiating with respect to α we get

$$\frac{d(Y(\hat{\pi})/\hat{\pi})}{d\alpha} = \frac{A\alpha}{K(\alpha - 1)^2}(\alpha - 2) + 1/2,$$

which is negative for α near 1, that is for small α . This is consistent with the often made claim that farming societies were worse off than hunter gatherers, while of course for large α the theory predicts the certainly true fact that industrial societies are much better off.

15.3.1 Jewelry versus swords

Free resources, that is output above subsistence, can be used to improve utility and also to improve the chances of success in conflict. There can be complementarity between those goals - for example a good court system can both improve commerce and improve chances of success in conflict. There can also be substitutability between them. As an example, consider families who must choose whether to take their share of surplus output as either jewelry or as swords - the former representing things that raise individual utility, the latter things that improve the chances of success in conflict. Left to themselves individuals will clearly choose jewelry, however as is the case with family size societies can enforce social norms which require the acquisition of swords rather than jewelry.

The basic point is a simple one: greater state power requires free resources be spent on swords, and so evolution pushes in this direction. We shall see later that the time span over which societies fall is long - on the order of hundreds of years - so swords have little value even as a public good, the chances of losing a conflict being very low. This observation reflects a basic bias in conflict driven evolution. It is not benevolent. It favors systems that are better able to defend themselves regardless of the impact on individual utility. Consider for a moment defense expenditures in the USA versus what would actually be needed to defend the country from realistic threats. Given the geographical isolation from strong rivals the answer is surely very little. Yet evolutionary pressure is towards societies such as the USA that grossly overspend on the military.

Free resource maximization is a lot like profit maximization and indeed our formal models has a strong similarity with models of profit maximization. Moreover, from an empirical standpoint, this connection may explain the historical importance of monarchies that can only be described as profit maximizing. But as the jewelry and swords example makes clear, this connection is not perfect: the residual - profits if you will - may be turned to many uses and only some of these uses - swords not jewelry - enhance state power. Hence a profit maximizing monarchy that through social norms is bound to use its profit for fighting and conflict is the type we expect to survive: a profit maximizing monarchy that spends its profits on large and beautiful palaces has less of a future.

15.3.2 Wolves versus sheep

Societies vary by their inclination to export their ideas and social norms. The model we have presented is one of expansionary societies - always ready to grab some land from their neighbors. Expansion may

have many forms and motivations: two examples are conquest through warfare or religious conversion, but others are the desire to explore new territory, contact other societies and mix with them, propose values and possibly learn from outside communities. The Roman empire is a strong example of the first type of expansion; more modern expansions have often involved religious conversion - for example, the sending of religious missionaries, although this has often occurred in the context of warfare, for example the conversion of the South and Central American Indian populations to Christianity through a combination of conquest and missionary activity. Equally relevant is influence through exchange of goods spurred by explorations (think of Marco Polo), or the more modern culture spreading through the sale of goods ranging from Coca-Cola to television sets. Or going to the other extreme, we may think of the “curiosity”, that is the expansionism, of the primitive hunters-gatherers.

Regardless of the form of expansion, expansionary institutions are not universal.¹³ Religions such as Christianity and Islam have historically been expansionary trying actively to convert nonbelievers. By contrast since the diaspora Judaism has been relatively insular in this respect, and the same has been true of other groups such as the Old Believers in Czarist Russia. To our model we may add an additional type of society, non-expansionary ones which have resistance to losing land, but also very high resistance to gaining land.

Without going through the details, societies that are non-expansionary are never stochastically stable. Although they may be part of an absorbing state for $\epsilon = 0$ they face the problem that they can be whittled away: a unit of land is lost and the new situation is also an absorbing state. Ellison (2000) shows that absorbing states like these are not stochastically stable.

This observation helps us understand a puzzle. Why should Spain now more than 300 years after ceding it to the British still wish to regain Gibraltar. Why would the UK be concerned to retain the Falkland Islands and Scotland? Why should Spain care about Catalan succession? The answer to all these questions is: this is the nature of societies selected by evolutionary pressure. By contrast to the successfully independent UK and Spain the area currently comprising Czechia and Slovakia historically has not been independent - for centuries being part of the Habsburg Empire. Hence we should be less surprised at the division of Czechoslovakia into Czechia and Slovakia.

15.3.3 The weakest link

We have made assumptions that a society can only be in one state. In fact societies wax and wane, sometimes being strong and other times weak. We can add to the model internal states $\xi_{jt} \in \Xi$ for each society. As long as these evolve according to a fixed Markov process with transition probability $\Pi_j(\xi_{jt} | \xi_{j,t-1}) > 0$ independent of land holdings we can easily do the evolutionary analysis. State power will now depend upon internal state $\gamma_j(\xi_{jt})$. The overall state will now be $z_t = (L_{1t}, \xi_{1t}, L_{2t}, \xi_{2t}, \dots, L_{Jt}, \xi_{Jt})$. Instead of stochastically stable states in Z we will now have stochastically stable sets corresponding to the different points in Ξ . Theorem 15.1 remains intact except that instead of maximizing state power it is the weakest internal state that matters, that is, it is the hegemony of the largest value of $\underline{\gamma}_j = \min_{\xi} \gamma_j(\xi)$ that is stochastically stable.

Two examples illustrate what this is about. One case could be the model in Acemoglu and Robinson (2000) where the voting franchise is extended or contracted. As the system oscillates back and forth

¹³ Our notion of expansionism is connected to the theory of the transmission of innovations in Aoki et al. (2011).

between civilian and military rule what counts is state power in the weaker of those two internal states. In that model oscillations are driven by recessions - and recessions themselves have implications, state power presumably being greater in a boom than a recession. Hence from an evolutionary point of view what matters is that the worst possible recession be relatively mild. Even if from the point of view of welfare it might not be worth engaging in stabilization policy that lowers the variance of recessions because the reduction in mean consumption is too great, it could never-the-less be a strategy favored by evolution if the stabilization policy moderates the worse possible recession. Here perhaps lies an explanation for why government institutions are obsessed with recession and often seem to carry out stabilization policy to an excessive degree. Although it might be wrong from a welfare point of view, it may also be the strategy favored by evolutionary forces.

15.4 The fall of the Qing dynasty¹⁴

So far we have dealt with stochastically stable states or sets - behavior in the very long-run. As $\epsilon > 0$ the Markov process has rich dynamics and the theory provides a great deal of information about fluctuations near stochastically stable states and transitions between stochastically stable states. Suppose that \hat{z} is absorbing in the limit dynamic where $\epsilon = 0$. The main result proven in Levine and Modica (2016) is that the system will typically moving repeatedly out of \hat{z} and back, and finally exiting to reach the absorbing state \tilde{z} which is achievable with the least resistance.¹⁵ The likelihood of transition paths is greater the less their resistance. This implies that monotone paths in which a society repeatedly gains land without losing it are far more likely than one where it suffers temporary setbacks.

For our model the least resistance of leaving a hegemony is to have it attacked by a most powerful society which repeatedly gains land until it becomes hegemonic. We have already said the most powerful society is unstable, and we call it a society of *zealots*. Their hegemony has zero resistance to losing land, so the next least resistance phase is invasion of the zealots by a stable society which in turn conquers all land period after period.

With respect to fluctuations near a stochastically stable state \hat{z} Levine and Modica (2016) prove that the largest fraction of time will be spent at the stochastically stable state but there will be a large number of periods during which the system will move to nearby states that are not absorbing and falls back again. The relative likelihood of these states is given by the resistance of getting there. For example the hegemony may use a unit of land then get it back again; this happens many times, and many times more often than the number of times it loses two units of land before falling back. Eventually the system will breach what is called the inner basin - it will be sufficiently far from \hat{z} that there is no longer probability 1 of returning when $\epsilon = 0$. These states are the set of warring states. During this period there will be many societies which may rise and fall, and swap land back and forth - it is a chaotic and turbulent phase. Eventually a new hegemony will rise. The rise of the new hegemony is in some respects opposite of the fall. Once a stable society has enough land that it has positive resistance to land least resistance implies it can only gain land and not lose it.

¹⁴ Based on Levine and Modica (2016).

¹⁵ The resistance of a path is just sum of the resistances of its transitions.

Levine and Modica (2016) show that the entire period of transition is short relative to the length of the hegemony, and of course that must be true individually for each of the three phases. Under some additional assumptions Levine and Modica (2016) also show that the warring states phase where land trades chaotically back and forth is long relative to the initial invasion by zealots and the subsequent rise of a new hegemony.

Finally, Levine and Modica (2016) show that the longest departures from an absorbing \hat{z} that result in a return are typically much longer than the final transition to a new absorbing state.

As the model has many implications for the details of the dynamics it seems useful to examine a case study to see if these details are reflected in reality. The fall of the final Qing dynasty in China (early 20th century) is a good case study because we have good data about those details.

15.4.1 China

We compare the theoretical predictions of the model with the fall of the final Qing dynasty in China and the subsequent rise of the communist hegemony - a case study for which there is quite a bit of historical information.¹⁶ The basic fact, as we shall see, is that Chinese institutions that lasted from roughly the introduction of the Imperial Examination System in 605 CE until 1911 CE were swept away in less than a year. At the beginning of the nineteenth century, before the First Opium War, the Qing dynasty held a hegemony over China proper.¹⁷

Several independent sources of instability concurred to the fall of the hegemony. In the early 1800s China fell into a severe economic depression from which it did not recover prior to the fall of the hegemony. Outsiders, most notably the English, French, and Japanese actively intervened in China, sometimes fighting for and other times against the Qing, but in any case certainly piling on pressure. Opium consumption, induced by the English to correct trade imbalances, increased as well.

From 1839 to 1910 there were a series of unsuccessful attempts to overthrow the Qing dynasty including local rebellions and acts of defiance by committed revolutionaries. During this time the outlying territories were lost: Korea became independent, Indochina was lost to the French, and Taiwan to the Japanese further weakening the hegemon. Roughly speaking the state ξ_j became increasingly worse. However, each internal rebellion was successfully repressed, each war brought to an end, and in each case the Qing hegemony over China proper - tax collecting authority, control of institutions local and global - remained intact. Although it is hard to measure the relative frequency of failed rebellions before and after the economic weakness of the 19th Century, in the earlier periods there seem not to have been such dramatic episodes as the Boxer rebellion and the less known Duggan revolt (which lasted for fifteen years). As our discussion of the weakest link suggests, before the actual fall the state ξ_j is very bad, and there are many and probably increasing failed attempts at rebellion.

The actual fall of the Qing occurred in 1911 and it was very quick - as predicted the least-resistance results we mentioned above. There were again a series of revolts - now however they succeeded. Also as the theory suggests the length of the successful revolt - less than a year - is considerably shorter than the longest failed rebellions - the Boxer and Duggan rebellion lasting many years. The final successful revolt is coordinated by Sun Yat Sen. The groups carrying out the various revolts can reasonably be

¹⁶ There are of course many accounts of this period, and while they sometimes disagree on exactly who did what to whom when, all agree on the basic facts we describe below. One readable account by a journalist is that of Fenby (2008).

¹⁷ That is excluding the Korean Peninsula, Indochina, and Taiwan.

described as zealots: they share in common a dedication to overthrowing the Qing, they are willing to suffer severe risk and live under unpleasant circumstances in order to achieve that goal. Such behavior is power maximizing, but is not stable in the sense that no society has ever lasted very long based on the fanatical devotion of its members - nor was it the case in China. Hence the theoretical description of the fall of the hegemony is relatively accurate: zealots quickly capture the land, and do so without a serious setback. In some cases land is seized by other groups, but they quickly join Sun Yat Sen as the theory suggests. By the end of 1911 the Qing Emperor abdicated and Sun Yat Sen became the provisional President of China, which however no longer was hegemonic in any reasonable sense of the word.

Next is the period of warring states, both in theory and in fact. The theory says that there can be many competing societies, land may be lost and gained, zealots may or may not play a role. Again, this is an accurate description of the situation in China between 1911 and 1946. Sun Yat Sen was quickly deposed by a less fanatical and more materialistic warlord Yuan Shikai, but until about 1927, and even after, there are many warlords in various parts of China who rise and fall, many revolutions, some successful and other unsuccessful. Basically the theory predicts chaos (in a non-technical sense) and that is what we see. Beginning in about 1927 things settle down slightly with two relatively more powerful groups, the Nationalists and the Communists, fighting a civil war - but there remain many warlords who continue to rise and fall, at times forming alliances or professing allegiance to the two more significant groups. The two major groups, unlike the earlier revolutionaries, appear to have coherent and potentially stable institutions. Then in 1936 the Japanese seize control of most of the country, an occupation that lasts until 1945. Notice that as the theory suggests the length of this warring states period - 35 years - is much longer than either the fall (less than a year) and the rise (about three years).

The final stage of a least resistance transition is the rise of the new hegemon. Now we are in the basin of the new hegemony so the least resistance path consists of the hegemony gaining territory - without losing any - until hegemony is again established. Notice that since in this model once the basin is left there are zero resistance transitions to any particular hegemony breaching the threshold, the model makes no prediction about which hegemony eventually emerges - in particular there is a non-negligible probability that even a very weak hegemony emerges. In China, the threshold appears to be reached about 1946 when the Communists controlled about a quarter of the country and about a third of the population. They quickly overran the remaining areas held by the Nationalists, who retreated to Taiwan in 1949.

15.5 Peace¹⁸

While hegemonies are common in history, there are two glaring exceptions, namely India and continental Europe. Indeed the situation, especially in Europe, can better be described as a balance of power between competing societies. Clearly the theory presented so far is deficient: it says that as one side gains an advantage it becomes more likely to gain additional advantage. To see what might happen the example of the Korean war is useful. In September 1950 North Korea was on the verge of dominating the South. On the 15th of September the United Nations led by the United States launched an am-

¹⁸ Based on Levine and Modica (2018).

prohibious invasion reversing the situation. But rather than gaining resources and weakening the North Koreans the result of this success was the entry of China into the war on the side of North Korea - resulting in the United Nations forces being pushed back and ultimately in a stalemate. The key point is: gaining land leads to greater weakness of the opponent only if it does not draw outside intervention.

The intervention of outsiders - protected typically by their own strong geographic barriers so not at risk in the conflict - is common in history. In Europe following the fall of Rome and up to around 1066 we have the continued interference of northerners - the Vikings and later Swedes were especially well protected by their own geography. Following 1066 we have the constant interference of England - also safe behind a water barrier: during this period we observe that England constantly intervened in continental conflicts but always to support the weaker side, and eventually this policy of balance of power became explicit.¹⁹ India also was subject to repeated invasion from central Asia - protected not by water but by difficult desert and mountain terrain.²⁰

To further advance the theory and understand the role of outside intervention in the balance of power we need to model outside intervention more closely.

15.5.1 Two competing societies

With this aim we now focus on a conflict between two societies so that $j \in 1, 2$. Here we have $y_j = \varphi(\gamma_j, L_j)$. We write resistance of j losing a unit of land as $r(y_j, y_{-j})$ omitting the land argument; monotonicity in land is always assumed.

The purpose of the section is to examine how intervening outsiders change the course of the conflict, depending on the strength of their involvement. The starting point is of course to look at the no-intervention benchmark, so we shall start with that and introduce outsiders later. In that case, since victory in battle weakens the opposition making further victories easier, typically one side achieves enough success that it is ultimately able to win the war in a relatively short period of time. This leads to the peace of the strong ruling the weak. The problem with outside intervention is that supporting the weak may result in much longer and bloody conflicts (think of the Vietnam war or the more recent Syrian hell). The results we are going to present in this section provide some suggestions as to “how to intervene if you must,” *given* outsiders’ goals. The qualification is important since we do not intend to address the issue of the relative desirability of short term peace versus long-term conflict, but instead try to develop a useful model of the length and nature of conflict and how it depends upon outside intervention. If protecting the weak from the strong is a priority, we will see that the level of intervention is a relevant determinant of the nature and length of the conflict, with stronger intervention being generally preferable towards the goal of minimizing casualties.

¹⁹ It is not completely correct to view England and Scandinavia as “outsiders” as at various time they had continental interests and conversely, but the key point is that they had a core area relatively safe from invasion. In a different direction Hoffman (2013) argues a role also for the Western Catholic church which in Europe acted as a balancing force much akin to the outsiders of our model.

²⁰ The exact nature of the asymmetry in the physical geographical barrier is uncertain, but it is a fact that India has been invaded numerous times successfully from Central Asia, but there have been no successful conquests of Central Asia from India. Phil Hoffman in a private communication suggests that part of the answer may lie in the fact that the area of Central Asia is well suited for raising horses and India is not, and that horses play a central military role in conflict between Central Asia and India.

15.5.2 Outside intervention and the balance of power

Outside forces can reinforce society j and we focus on the case in which only one party to the conflict is the beneficiary of outside intervention at any moment of time. If society j is reinforced its *combined power* is $\phi_j = y_j + y$, otherwise it is $\phi_j = y_j$. Hence resistance is given by $r(\phi_j, \phi_{-j})$.

The behavior of outside forces is determined by an intervention policy which we initially take to be exogenous. We study a simple but important intervention: intervention on behalf of the weak. Specifically we assume there exist thresholds \bar{L}_j, \bar{L}_k with $\bar{L}_j + \bar{L}_k < L$ and such that if $L_{jt} \leq \bar{L}_j$ then outsiders reinforce society j . The inequality $\bar{L}_j + \bar{L}_k < L$ means that \bar{L}_j appears on the left of \bar{L}_k in the land line (measuring j 's land from left to right). The assumption that the size of intervention y is the same on both sides is a simplification enabling us to focus on \bar{L}_j as a measure of the strength of intervention.

It is useful at this point to denote combined power of j as $\phi_j(L_{jt})$ since state power γ_j is determined by j while intervention is determined by L_{jt} (because $L_{kt} = L - L_{jt}$).

With outside intervention in addition to hegemonic states that are stochastically stable there may also be stable *balance of power segments* consisting of a contiguous collection of states not including hegemonies.²¹ To analyze segments we extend the idea of an absorbing state to that of an *absorbing set* - meaning that as $\epsilon \rightarrow 0$ the probability of escaping from the set goes to zero, but the probability of moving about within the set remains positive. Hence segments are absorbing if at the left end society j has positive resistance; at the right end society k has positive resistance; and in the interior, if nonempty, both have zero resistance to losing land.²² From the general results of Young (1993) discussed above only absorbing segments can be stochastically stable.

There may be no absorbing segments at all, a situation we refer to as *hegemony*. In addition two (and only two) types of absorbing segments are possible that we refer to as *hot peace* and *prolonged war*. To motivate these terms we make two claims.

First, either the left endpoint of the segment is the intervention threshold for j or the right endpoint of the segment is the intervention threshold for k . To see this suppose the left endpoint is not an intervention threshold. Society j cannot have zero resistance otherwise we move left without resistance and we are not really at the left endpoint of an absorbing segment. Therefore society k must have zero resistance; then as it loses land it continues to have zero resistance at least until the intervention threshold is reached. If at the intervention threshold the outsiders are strong enough to protect society k then this terminates the absorbing segment - we cannot escape to the right. If the outsiders are not strong enough then resistance is zero until hegemony is reached so there is no balance of power segment at all.

Our second claim is that the segment must either run the entire length between the intervention thresholds or it must have length one. To see this, suppose that the segment starts at the left intervention threshold. If moving a second step to the right has resistance then the segment has length one. If it does not have resistance then moving further to the right continues to have no resistance at least until we encounter the right intervention threshold and the outsiders come in to prop up society k .

This gives rise then to the above classification of absorbing segments:

²¹ This is proved in Levine and Modica (2018).

²² We always measure the land of j from the left and the land of k from the right.

Hot peace is a segment of length one at one of the intervention thresholds - that is either $L_j = \bar{L}_j, \bar{L}_j + 1$ or $L_k = \bar{L}_k, \bar{L}_k + 1$ (inclusive). Here a single unit of land changes hands back and forth. As well as a single hot peace segment there can be a pair of hot peace segments, one at each intervention threshold.

Prolonged war is a segment running from one intervention threshold to the other, that is from $L_j = \bar{L}_j$ to $L_j = L - \bar{L}_k$ inclusive. We also require that the intervention thresholds not be adjacent, $\bar{L}_j + \bar{L}_k < L - 1$, so that the segment is longer than one and is not a hot peace.

The two types give rise to considerably different scenarios. Indeed, if we were to subdivide the units of land the length of hot peace segments would shrink while the length of prolonged war segments would not. Hot peace can be thought of concretely as a relatively low key and “peaceful” conflict, with border skirmishes going on without land actually being gained or lost - for example, the recent conflict between Israel and Lebanon which occasionally flares into the firing of rockets over the border or a small border incursion. The point is that while a hot peace is not peace the level of conflict and casualties are low as the fighting is extremely limited. The case of prolonged war is on the contrary a real, bloody war where the two sides fight back and forth losing and gaining substantial amounts of land and not merely skirmishing at the border. The civil war in Syria in the last years is a case in point.

Whether we see hot peace, prolonged war or hegemony depends on the strength of intervention. We distinguish four levels of intervention *on behalf of j*:

Definition 15.1 (Intervention strength). 1. *Strong*. Intervention takes place when resistance is positive in the absence of intervention: $r(y_j(\bar{L}_j), y_k(L - \bar{L}_j)) > 0$

2. *Ineffective*. Intervention is insufficient to give positive resistance: $r(y_j(\bar{L}_j) + y, y_k(L - \bar{L}_j)) = 0$; this includes the case where there is no intervention.

For the remaining cases we assume that #1 and #2 do not hold, that is $r(y_j(\bar{L}_j), y_k(L - \bar{L}_k)) = 0$ and $r(\phi_j(\bar{L}_j), \phi_k(L - \bar{L}_j)) > 0$:

3. *Medium*. When j gains a unit of land above the threshold (thus losing support) the opponent has zero resistance to losing land: $r(\phi_k(L - \bar{L}_j - 1), \phi_j(\bar{L}_j + 1)) = 0$ and $\bar{L}_j + \bar{L}_k < L - 1$.

4. *Weak*. When j gains a unit of land above the threshold the opponent has positive resistance to losing land: $r(\phi_k(L - \bar{L}_j - 1), y_j(\bar{L}_j + 1)) > 0$

Depending on the level of intervention some segments are absorbing and others are not. In Levine and Modica (2018) we characterize the relationship between intervention and the existence of absorbing segments of different types. The results are reported in the following

Theorem 15.3. *Existence, if any, of absorbing segments depending on the type of intervention on behalf of societies j and k can be summarized in Table 15.1 (where land is expressed in units of L_j).*

Stochastic stability of these segments depends on the strength of the outside forces. The formal result, see Levine and Modica (2018), is the following:

Theorem 15.4. *There exist $\infty \geq \bar{y} \geq \underline{y} > 0$ such that if $y > \bar{y}$ and if intervention thresholds are positive on both sides there are stochastically stable balance of power segments but not stochastically stable hegemonies, while if $y < \underline{y}$ there are stochastically stable hegemonies but not stochastically stable - or even absorbing - balance of power segments.*

Table 15.1 Intervention and Peace.

	Strong k	Medium k	Weak k	Ineffective k
strong j	hegemony	hegemony	hot peace at $L - \bar{L}_k$	hegemony of j
medium j	hegemony	prolonged war or hot peace at $\bar{L}_j = L - \bar{L}_k - 1$	hot peace at $L - \bar{L}_k$	hegemony of j
weak j	hot peace at \bar{L}_j	hot peace at \bar{L}_j	hot peace at both \bar{L}_j and $L - \bar{L}_k$	hot peace at \bar{L}_j
ineffective j	hegemony of k	hegemony of k	hot peace at $L - \bar{L}_k$	hegemony

15.5.3 Peace and war

Theorem 15.3 shows that there is a non-monotonicity in the consequences of intervention - roughly, one can see this by scanning the above table going up left. To understand this non-monotonicity it is useful to consider a simple case. Suppose that j and k are equally strong so that $\gamma_j = \gamma_k = \gamma$, and that the intervention policy is symmetric so that $\bar{L}_j = \bar{L}_k$. Hence intervention policy is indexed by a single scalar, the land threshold for intervention on behalf of both contenders. We assume for the present discussion that the number of units of land L is odd.²³ Finally we assume that the strength of the intervenor(s) φ_0 is high enough that strong intervention is possible, but that it is ineffective for \bar{L}_j sufficiently small.

Now increase the intervention threshold for both sides at the same time. Start with \bar{L}_j small. In this case as we have noted intervention is ineffective - there is no point in intervening when j has become so weak that they have lost even with outside help. In this case there is no balance of power segment, but rather a hegemony of one society: we refer to this as the *peace of the strong over the weak*. As \bar{L}_j increases, eventually the point is reached where intervention is weak. As we indicated we now have a hot peace in which the weaker side survives by virtue of outside support and the stronger side by virtue of their strength.

The key transition to understand is that from weak to medium intervention, since it is medium intervention that leads to a prolonged war. Why is this? As the intervention threshold increases the side receiving support is propped up when it is relatively strong: eventually strong enough that the opposition no longer has resistance to losing land. At this point intervention becomes medium and when launched from behind the shield of foreign protection success is now possible and may sometimes range until intervention occurs on the side of the opponent. As an example of this we might consider the second Vietnam war until the withdrawal of the United States in 1973: here we have the United States intervening to prevent the fall of South Vietnam and the Soviet Union intervening to prevent the fall of North Vietnam. The war ranged for nearly twenty years with substantial battle deaths and loss and gain of territory on both sides and no doubt would have gone longer had the United States not withdrawn its intervention.

As the strength of intervention \bar{L}_j increases further the length of the prolonged war segment shrinks reducing the scale of the conflict until eventually \bar{L}_j reaches the center and we are again at a hot peace. As an example of this we might consider the intervention of the United States on both sides of the Israel/Egypt conflict at the Camp David accords in 1978: in effect the United States provides arms

²³ So it is feasible for the two thresholds \bar{L}_j, \bar{L}_k to be adjacent; this would be ruled out by symmetry if L is even.

and support to both armies to stare at each other across a border that will bring quick intervention in response to a violation.

We want to emphasize the non-monotonicity of the consequences of intervention in its strength: a weak or strong intervention leads to hot peace, but a medium intervention leads to prolonged war and it is the costliest in terms of lives and distress to the peoples and economies involved. No intervention brings about peace relatively quickly but with the strong dominating the weak. What the model suggests is that if the goal of protecting the weak is predominant, then to minimize the costs of war intervention should be strong enough to avoid going back and forth between states where one part in turn is considerably stronger than the other, and reduce the war to what we have called a “hot peace” - which can be thought of as “border skirmishes”, and hopefully end in reaching an unarmed negotiation stage.

We may ask: why do we see prolonged war at all? Should not the participants knowing that fighting will simply rage back and forth between the intervention thresholds just skip the conflict? We refer the reader to section 7 of Levine and Modica (2018) for analysis of costs and benefits of intervention and the game being played when there are two symmetric intervenors in competition with one another, the equilibrium of which may result in a prolonged conflict between the two societies involved in the original conflict.

15.5.4 History of modern war

We shortly turn to details of different configurations and illustrate them with examples. In addition to discussing specific cases, we gather the substantial post World War II conflicts in the form of tables. Cases where one combatant did not occupy any land are excluded as the theory does not apply. For the rest we examined each postwar conflict in the Uppsala database. We excluded those marked as insignificant, those involving military coups, those involving invasions of minor powers by major powers (for example: 1956 invasion of Hungary) and guerrilla conflicts where the guerrillas did not control land and resources (for example: the Basque region). We examined each remaining conflict and believe that we have included the most significant. In some cases there were several intervention regimes: we discuss those separately. The data about individual conflicts is taken from Wikipedia. The tables show the region, the year in which the conflict began, and the number of years it lasted. Casualties (including civilian casualties) are reported in deaths per 100,000 per year which is the standard unit for reporting, for example, murder rates.²⁴ To put these numbers in context, note that the overall murder rate for Europe and Asia is about 3, for the entire world about 6, and for Africa about 12 and for the Americas about 16. So, for example, the death rate of 20 in the Sri Lankan civil war (a hot peace) is comparable to the murder rate in the Americas, while the death rate of 380 in the Syrian civil war (a prolonged war) is an order of magnitude higher. Following the casualty rates we list the parties and outside intervenors. In cases when war ended due to the withdrawal of intervention we report the “collapse” as the number of subsequent years until one side achieved victory. Entries in the table are arranged in chronological order.

Before examining the details, it is worth taking a broad overview of our findings. Intervention that either is designed to preserve the balance of power or which does so because of conflicting interests of

²⁴ Civilian casualties are the bulk of casualties and there are a wide range of estimates. We used the middle of the range of estimates.

the intervenors can lead either to a hot peace or a prolonged war. There is a large discontinuity in the amount of harm done in a hot peace and a prolonged war: in a hot peace death rates are on the order of relatively high murder rates, or in some cases lower, while in a transitional or prolonged war they are an order of magnitude larger than very high murder rates. Taking the Sri Lankan civil war as an example of hot peace we see that for 26 years the death rate was about 20, comparable to the highest murder rates in the world. Taking the breakup of India and Pakistan after the British withdrawal as an example of a transitional war it was vastly bloodier - the death rate was about 250. However, the Sri Lankan civil war lasted 26 years so the total is about 520, more than double that in India and Pakistan where the transitional war lasted only a year. Overall a hot peace does not seem to represent much of savings in terms of casualties over non-intervention and a transitional war - but it does protect the weak. From a policy point if we were to take the point of view that, say, Lebanon posed a threat, then keeping it a bloody mess for three decades would surely neutralize that threat - but from a humanitarian point of view it represents a catastrophe. If we are to take a very cynical view of the conflict between Shia and Sunni, especially the current war in Syria, as a Western effort to preserve a balance of power that neutralizes the Arab world as a threat - the wave of refugees descending on Europe with the consequent social and political problems shows that such an effort can have pretty heavy unintended consequences.

We compare prolonged war - brought about by medium intervention - to hot peace. As we have indicated, this seems the least justifiable form of intervention. Data are collected in Table 15.2. The only rationale for medium intervention we can think of is that a region poses a particular danger and hence the importance of keeping it weak offsets the bloody harm of prolonged war. Yet, if we look at the record, Vietnam, Sudan, Angola, Lebanon, and Syria do not appear to have ever presented any great danger to the intervening powers. It is interesting that while the US intervention in Vietnam is widely criticized outside the US, it seems to be so for mostly the wrong reasons. Surely there was nothing wrong with supporting the South, for, despite all the shortcomings of its government, there was no popular desire to be ruled by the equally bad or worse government in the North. Nor can there be much moral doubt about opposing the spread of communism: one need not look further than North

Table 15.2 Medium Interventions with Prolonged Wars.

Region	Start	Duration	Casualties	Parties	Intervenors	Collapse
Vietnam	1955	20	170	North South	Soviets US	1
Sudan	1955	60	330	North Sudan Southern Sudan	Egypt Ethiopia/Uganda	
Angola	1975	27	86	MPLA UNITA	Soviets South Africa	1 [1]
Lebanon	1975	31	400	Shia Christian/Druze	Syria Israel	1
El Salvador	1979	12	138	Government FMLN	United States Soviet Union	
Syria	2011	5+	380	Government Insurgents	Russia, Iran West	

1. The date at which intervention on behalf of UNITA ceased is unclear. We dated it to May 2001 when DeBeers - the main source of funding and illicit weapons shipments to UNITA - ceased operation in Angola.

Korea and Cuba - two of the most miserable places in the world - to see that. Nor is it clear why the direct involvement of the US is worse than indirect Russian involvement. From our point of view the US should be rather criticized for creating a prolonged and costly conflict by attempting to maintain a balance of power in the South.

Table 15.3 shows, by contrast, that hot peace has much lower cost.

Region	Start	Duration	Casualties	Parties	Intervenors	Collapse
Iron Curtain	1945	46	0	Eastern Europe Western Europe	Soviets US	1
Sinai	1948	68+	1	Israel Egypt	West Soviets [1]	
Korea	1950	66	8	North South	China US	
Nagorno-Karabakh	1988	28	26 [2]	Armenia Azerbaijan	Turkey Russia	

*1. It should be noted that originally the Soviets supported Israel.
2. It is unclear in which population the casualties occurred. Virtually all deaths occurred during the six years of active war beginning in 1988. It is estimated that 28,000–38,000 died in that conflict. The population of Nagorno-Karabakh is only 147,000, but it is highly unlikely the bulk of casualties occurred among that population. We used the average of the population of Azerbaijan and Armenia as our base population.*

15.6 Prosperity²⁵

We lastly build on the understanding of evolution of institutions we have gained in the previous sections to address the highly debated issue of why the industrial revolution took place in the West rather than, say, China. To do so we use a model where some of the basic conclusions reached above are taken for granted in order to go into more details about institutional arrangements and warfare technology. The specification of the model is motivated by the widely held belief that technological progress was favored by competition between relatively inclusive institutions such as those in Western Europe, and hindered by the relatively extractive hegemonies found in China (see, for example, Landes (2003), Lin (1995), and Liu and Liu (2007)). Accepting this basic conclusion we ask: why was there competition between relatively inclusive institutions in Europe while in China we find an extractive hegemony? Why did India - made up of competing societies not an extractive hegemony - generate relatively little innovation? To address the issue we have to somehow formalize the ideas of extractive and inclusive societies, and to be able to talk about “who may include whom” we will have two groups in a society. Assuming for simplicity that there are only two societies we take for granted that the long run possibilities are hegemony and balance of power - but which now may feature extractive and inclusive societies. The puzzle of the emergence of the inclusive balance of power in Europe is that since extractive institutions generally levy higher taxes and have larger armies, if evolution is driven by conflict

²⁵ Based on Levine and Modica (2019).

how come these “strong” extractive institutions do not predominate over “weaker” inclusive institutions? Outsiders - whose influence we have stressed in the previous sections - will still be an important part of the picture but we now want to open the model to the different possible institutional equilibrium arrangements, and to do this we have to be more specific about how wars develop. As we shall report in this section, we find that when outsiders are weak extractive hegemonies will predominate, while when outsiders are strong military technology matters: if war outcomes are insensitive to defensive strength - fortifications are well able to resist siege - an extractive balance of power will predominate, while good siege technology will result in the predominance of an inclusive balance of power. The last possibility may be thought of as “survival of the weakest”. In brief, in Europe around 1500 CE we argue that outsiders were strong and military technology (think of the cannon) was already well developed; while in China the Mongolian diaspora greatly weakened the strength of outsiders, resulting in the extractive hegemony from the Yuan dynasty onward. We now turn to details.

15.6.1 Model basics

The basic elements are a stripped down version of the previous ones: there are two societies and two units of land, one for each society; and there are two possible configurations: a *balance* of power in which each society occupies its own unit of land and *hegemony* in which one society, the *occupier*, occupies both units of land and the other society is referred to as the *occupied*.

The novel element is that there are two groups in each society: the *commercial elites* and the *military elites*. There are two types of institutions, inclusive institutions w and extractive institutions s . Roughly speaking with inclusive institutions the commercial elites have the upper hand, while with extractive institutions the military elites have the upper hand. Depending on circumstances either society may have either type of institution. There are five possible states of the system: $z \in Z = \{w, s, ww, sw, ss\}$. The first two correspond to hegemony in which the occupier has inclusive and extractive institutions respectively and the remaining three correspond to a balance of power in which both have inclusive, one has extractive, the other inclusive, and in which both have extractive institutions.

Conflict between societies takes place over time $t = 1, 2, \dots$. At the beginning of period t there is a status quo given by the state from the previous period z_{t-1} . A game between the two groups in the two societies is played and the outcome determines the state z_t in the current period. The particular game depends upon the status quo z_{t-1} and a iid random shock. It takes place in two stages. In the first stage only one of the four groups is *active* and may decide to initiate a conflict to achieve a particular *goal*. The decision is based on a stochastic utility shock. If the active group is part of an occupied society the conflict is a *rebellion* to liberate their land and the goal is to install particular institutions there; thus if the rebellion is successful the hegemonic state will transit to a balance of power. If the active group is part of a balance of power the conflict is to *attack* the other society and the goal is to occupy their land; in this case success will result in hegemony. If the active group chooses not to initiate a conflict the status quo remains unchanged and $z_t = z_{t-1}$. If the active group initiates a conflict a second stage simultaneous move game is played. The active group initiating the conflict is designated as the *aggressor* and one group from the opposing society is the *defender*. Each simultaneously decides the level of effort to devote to the conflict and these effort levels stochastically determine the new state. All of the groups are myopic in the sense that they care only about the consequences of their actions in the current period.

In addition to the four decision making groups of insiders there are a number of *outsiders* whose strength relative to the insiders is denoted as before by $y > 0$. Our basic hypothesis remains that outsiders are disruptive of hegemony but supportive of a balance of power.

15.6.2 The initiation of conflict

We now describe in greater detail the game and payoffs.²⁶ In a balance of power each society has an equal chance of being active. The goal is to occupy the land belonging to the other society and install the active group's institutions there. In a society with inclusive institutions the active group is the commercial elite; in a society with extractive institutions the active group is the military elite.

In an inclusive hegemony the active group is the occupied commercial elite. In an extractive hegemony the active group is the occupied military elite. There are two possible goals: either to revolt and install inclusive institutions, or to revolt and install extractive institutions. It is possible, for example, that the commercial elites would agree to extractive domestic institutions in return for liberation from foreign domination by inclusive institutions.

Once the active group and goal are determined an iid random utility shock \tilde{u} occurs. The active group then decides whether or not to initiate conflict - to attack or revolt. If the active group decides not to initiate a conflict the game ends and the state remains unchanged. In this case $z_t = z_{t-1}$ and the utility of all groups is that in the status quo. If the active group decides to initiate conflict the utility of the active group is increased by \tilde{u} , the current state z_t is randomly determined through conflict resolution, and the utility of all groups is determined by the current state (as specified shortly) minus the costs of conflict plus the utility shock for the active group.

As conflict - at least in the sense of an all-out revolt or attempt to occupy a foreign nation - is rare, we assume that the utility shock is with high probability negative. If \tilde{u} is very negative the active group will not choose to initiate a conflict, so it is only the upper tail of this random variable that matters. We assume this has an exponential form given by three parameters $U > 1$, $0 < P < 1$, and $\sigma > 0$ so that if $v \geq -U$ then $\Pr(\tilde{u} \geq v) = P e^{-\sigma(v+U)}$. Observe in particular that $\Pr(\tilde{u} \geq -U) = P$. With probability $1 - P$ the shock is smaller than $-U$ and no conflict is initiated. The parameter σ is a scale parameter for the utility shock distribution. If it is large the probability of a shock much bigger than $-U$ is very small. We will be interested in the case in which σ is large.

15.6.3 Conflict resolution

The simultaneous move game following the decision of the active group to initiate a conflict is as follows. If conflict takes place the active group - now called the *aggressor* - determines the level of effort $1 \geq x_a \geq 0$ to devote to the conflict. In a balance of power the defender is the commercial elite if the society under attack has inclusive institutions and the military elite if the society under attack has extractive institutions. In an inclusive hegemony the defender is the occupier commercial elite; in an extractive hegemony the defender is the occupier military elite. The defender determines a level of effort $1 \geq x_d \geq 0$ to devote to the conflict. Each contestant group $i \in a, d$ faces a quadratic cost of effort provision $C(x_i) = (\gamma/2)x_i^2$ where $\gamma \geq 1$. The two groups who are neither aggressor nor defender do not bear any cost of conflict. Note also our base assumption that every society regardless of its type

²⁶ The model is simple and stylized; robustness to more general formulations is discussed in Levine and Modica (2019).

faces the same cost of raising resources: our explanation of social outcomes - in contrast to Hoffman (2013)'s theory of the great divergence after 1600 - does not rest on the idea that there are systematic differences in the cost of raising resources due to social organization.

Let $\zeta \in \{h, b\}$ be an indicator of whether the state is hegemonic or a balance of power. The probability the aggression succeeds depends on the resources committed by the contestant groups and is given by a conflict resolution function

$$\pi(x_a, x_d) = \Pi_\zeta(y) + \alpha(x_a - [(1 - \varphi)x_d + \varphi])$$

where $0 \leq \varphi \leq 1$, $\Pi_\zeta(y) > 0$ is continuous with $\Pi_h(y)$ strictly increasing in y , $\Pi_b(y)$ strictly decreasing in y and $\alpha > 0$, which is consistent with our view of that outsiders help rebels chances of success - $\Pi_h(y)$ increasing - but hurt those of an aggressor in a balance of power - $\Pi_b(y)$ decreasing. We also assume that there is a unique value y^* such that $\Pi_h(y^*) = \Pi_b(y^*)$. The parameter α measures the sensitivity of the outcome to the differential effort of the two combatants. We have assumed that this is not too large. The parameter φ measures the sensitivity of the outcome to defensive effort. The coefficient on x_d is $(1 - \varphi)x_d + \varphi$ a weighted average of the defensive effort and 1. Our interpretation is that φ measures the value of fixed fortifications. The reason for this is that the benefit of fortifications is that they enable a small army to hold off a much larger force. In Masada in 66 CE, for example, a group of roughly 1,000 men women and children held off the Roman Empire for about seven years before being overcome by a military force of around 15,000. On the other hand, effective fortifications reduce the benefit of a larger defending force: it is unlikely that a Jewish force of 2,000 or 3,000 would have had much more success against the Romans than 1,000. Here the effectiveness of defense is measured by $(1 - \varphi)x_d + \varphi$ where φ captures the basic idea that with effective fortifications the defense is strong but not particularly sensitive to defensive strength. Hence our interpretation of φ as the effectiveness of fortifications.

If the aggression fails the status quo remains unchanged, $z_t = z_{t-1}$ and the utility of all groups is that in the status quo less the effort and plus the utility shock. If the status quo is a balance of power and the aggression succeeds the new state is hegemony with the institutions of the aggressor. If the status quo is hegemony and the aggression succeeds the new state is a balance of power in which the defender institutions are unchanged and the aggressor institutions are those determined by its goal. In all success cases the utility of all groups is that of the new state less the effort and plus the utility shock.

15.6.4 Incentives: transfers and utility

In addition to the random utility shock and conflict costs the utility of groups is determined by the current state z_t . From the economic point of view the two groups in each society are characterized by a *transfer* of resources from the commercial elites to the military elites. Extractive societies are defined so that this transfer is larger than in inclusive ones. So there are two possible transfer levels representing a transfer from the commercial elites to the military elites on each unit of land; we normalize high transfers to 1, and low transfers are $0 < \tau < 1$. In a balance of power the military elites receive the transfers from their own land, so for example in a w -type society we can write the transfer vector as $(-\tau, \tau)$. In hegemony the occupier military elites receive the transfers from both units of land and high transfers are always taken from the occupied commercial elites; and the occupier commercial elites pay τ in an inclusive hegemony and 1 if the hegemony is extractive. The transfers are summarized in the following table (for example the $-\tau, 1 + \tau$ entry says that in an inclusive hegemony the occupier

commercial elite pays τ and the occupier military elite gets τ from them plus 1 from the occupied commercial elite)

	s	w
balance	-1, 1	$-\tau, \tau$
occupier	-1, 2	$-\tau, 1 + \tau$
occupied	-1, 0	-1, 0

The key features are that there is a preference against foreign rule by both commercial elites and military elites and that occupier military elites benefit from occupation at the expense of the occupied military elites.

15.6.5 Equilibrium

An equilibrium is the stochastic process in which a Nash equilibrium occurs within each period. It is shown in Levine and Modica (2019) that equilibrium is unique and it is an aperiodic and ergodic Markov process on the state space $Z = \{w, s, ww, sw, ss\}$. From what we have seen the possible transitions other than remaining at the status quo are the following:

$$w \rightarrow ww, sw \quad s \rightarrow ss, sw \quad ww \rightarrow w \quad sw \rightarrow s, w \quad ss \rightarrow s$$

We denote the unique ergodic probability distribution over the state space by μ_σ . As before as $\sigma \rightarrow \infty$ the ergodic distributions μ_σ have a unique limit μ and the stochastically stable states are those for which $\mu(z) > 0$. These are the states which are observed “most of the time” when σ is large - the case of interest given that serious conflict is infrequent.

15.6.6 Stochastic stability

The “typical” institutional configurations are characterized in the following result.²⁷

Theorem 15.5 (Prosperity). *Only one of s, ss, ww can be stochastically stable (w and sw cannot). There exist τ^*, φ_τ such that:*

1. *if $\tau > \tau^*$ or $\varphi > \varphi_\tau$: for $y < y^*$ the stochastically stable state is s , for $y > y^*$ it is ss*
2. *if $\tau < \tau^*$ and $\varphi < \varphi_\tau$ there is a function $y(\varphi) \leq y^*$ such that: for $y < y(\varphi)$ the stochastically stable state is s , for $y > y(\varphi)$ it is ww (“survival of the weakest”).*

Let us see what the result says. If $\tau > \tau^*$ - inclusive institutions do not differ much from extractive ones - the only stable institutions are extractive. Roughly: the commercial elites are unwilling to make much effort to defend inclusive institutions that are not all that inclusive. Its interesting implication is that we will not often see “somewhat inclusive” institutions, only extractive or “strongly inclusive” institutions. Turning to the case $\tau < \tau^*$, first observe that if $y < \min_\varphi y(\varphi)$ - sufficiently weak outsiders - then only extractive hegemony is stochastically stable, regardless of military technology. By contrast

²⁷ More complete statement and proof in Levine and Modica (2019).

with stronger outsiders, that is larger values of y , we will see a balance of power - but military technology determines which type: large φ - effective fortifications, large defensive armies not needed - favors extractive institutions, while small φ - good siege technology - favors inclusive technology.

As we mentioned, our reading of this result is that the invention of gunpowder led to a great reduction in the effectiveness of fortifications. Small φ and strong outsiders made stable inclusive balance of power possible in Europe; in China extractive hegemony was favored by weak outsiders; and in India an extractive balance of power emerged as a consequence of strong outsiders but primitive siege technology (the cannon arriving late). We look at history in the next section.

15.6.7 History

We report here historical data from Maddison (2007) Table 15.4 below covers the three major regions: India, China, and Europe up to the period of globalization beginning around 1500 CE and beginning - for reasons of data reliability - about 0 CE. For each region epochs reported by historians are given and the strength of outsiders and fortifications reported. The penultimate column reports the actual state and the final column the stochastically stable state predicted by the theory. Eyeballing the data, the theory does quite well, missing only one epoch, the Sui/Tang period in China where the theory says we should have seen an extractive balance of power, but in fact see an extractive hegemony. The states that the theory says are not stochastically stable are never seen in the data.

We can subject this data to a formal analysis. Recall that $z \in \{ss, s, ww\}$ are the possible observed states and let $\hat{z}(\eta, \varphi)$ the stochastically stable state determined by the exogenous variables. Our theoretical model says that the stochastically stable state should be observed with high probability over sufficiently long periods. We take “long periods” to mean 2.2 to 5.6 centuries. We model “high probability” empirically by assuming that with probability β the stochastically stable state $\hat{z}(\eta, \varphi)$ is observed and that with probability $1 - \beta$ the observed state is drawn randomly with probabilities $\alpha(z)/(1 - \beta)$ where $\sum_{z \in \{ss, s, ww\}} \alpha(z) = 1 - \beta$. This gives rise to the linear conditional probability model $\Pr(z|\eta, \varphi) = \alpha(z) + \beta \cdot 1(z = \hat{z}(\eta, \varphi))$.

Maximum likelihood estimation of this model gives estimates $\bar{\alpha}(ss) = \bar{\alpha}(ww) = 0$ and $\bar{\beta} = 0.91$ with a standard error 0.08.²⁸ This indicates that the model does well in both economic terms - $\bar{\beta}$ is close to 1 and far from 0 - and in statistical terms - $\bar{\beta}$ is estimated with a high degree of reliability.

To understand better the importance of sampling error we draw $\lambda(\beta)$ the log likelihood function as a function of β in the graph below. The solid vertical line in the graph is the left limit of the 95% confidence interval based on the estimated standard error of $\bar{\beta}$. Also, twice log of the likelihood ratio $-2(\lambda(\beta) - \lambda(\bar{\beta}))$ is approximately chi-squared with one degree of freedom. Above the horizontal line is the 5% acceptance region of the likelihood ratio test, which thus shows that the range $[0.60, 0.99]$ are those values of β that cannot be rejected at the 5% level.²⁹

²⁸ From Table 15.4 we see that the log-likelihood is equal to $8 \ln(\alpha(ss) + \beta) + 4 \ln(\alpha(s) + \beta) + \ln \alpha(s) + 2 \ln(\alpha(ww) + \beta)$. The solution to the constrained maximization was computed using *R* (code available).

²⁹ The acceptance region is $-2(\lambda(\beta) - \lambda(\bar{\beta})) \leq \bar{\chi} = 3.841$ or $\lambda(\beta) \geq \lambda(\bar{\beta}) - \bar{\chi}/2$: the chi-squared line in the top graph above plots $\lambda(\bar{\beta}) - \bar{\chi}/2$ so the 5% acceptance region is above the line, and rejection region below it.

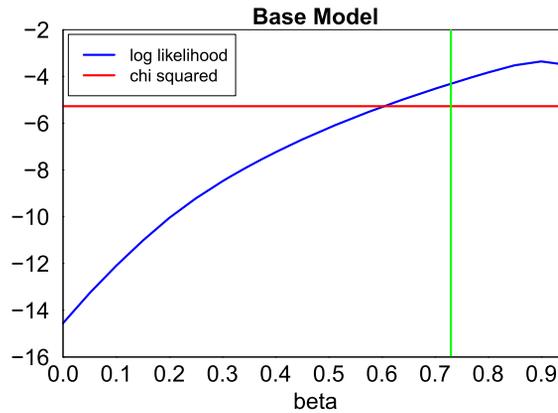


Table 15.4 Data. The duration of most episodes is from 2.2 to 5.6 centuries.

Region	Description	Period	Duration	Outsiders η	Fortifications φ	Actual state	Stable state
India	Classical/Medieval	200 BCE - 320	5.2	strong	strong	ss	ss
India	Classical	321 - 650	3.3	strong	strong	ss	ss
India	Early Medieval	651-1200	5.5	strong	strong	ss	ss
India	Late Medieval	1200 - 1525	3.3	strong	strong	ss	ss
India	Mogul	1526 - 1748	2.2	weak	weak	s	s
China*	Han	202 BCE - 220	4.2	weak	strong	s	s
China	Warring Kingdoms	221-588	3.7	strong	strong	ss	ss
China	Sui/Tang	589 - 906	3.2	strong	strong	s	ss
China*	Song	960 - 1279	3.2	strong	weak	ww	ww
China	Yuan/Ming/Qing	1280-1839	5.6	weak	weak	s	s
Europe	Roman	149 BCE - 329	4.8	weak	strong	s	s
Europe	Byzantine	330 - 628	3.0	strong	strong	ss	ss
Europe	Medieval	629 - 1054	4.3	strong	strong	ss	ss
Europe	Middle Ages	1054 - 1292	2.4	strong	strong	ss	ss
Europe*	Renaissance	1293-1607	3.1	strong	weak	ww	ww

Asterisks denote technological progress.

References

- Acemoglu, D., Robinson, J.A., 2000. Why did the West extend the franchise? *The Quarterly Journal of Economics* 115, 1167–1199.
- Aoki, K., Lehmann, L., Feldman, M.W., 2011. Rates of cultural change and patterns of cultural accumulation in stochastic models of social transmission. *Theoretical Population Biology* 79, 192–202.
- Bacci, M., 2006. *A Concise History of World Population*. Wiley Blackwell.
- Bowles, S., Choi, J.-K., Hwang, S.-H., Naidu, S., 2021. How institutions and cultures change: an evolutionary perspective. In: Bisin, A., Federico, G. (Eds.), *The Handbook of Historical Economics*. Elsevier. Chapter 14 (in this book).

- Clark, G., 2007. *A Farewell to Alms*. Princeton University Press.
- Ellison, G., 2000. Basins of attraction, long run stochastic stability and the speed of step-by-step evolution. *The Review of Economic Studies* 67, 17–45.
- Ely, J., 2002. Local conventions. *The B.E. Journals in Theoretical Economics* 2, 1–32.
- Fenby, J., 2008. *Modern China: The Fall and Rise of a Great Power, 1850 to the Present*. Penguin Press.
- Fudenberg, D., Maskin, E., 1986. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54, 533–554.
- Gibbon, E., 1776. *The History of the Decline and Fall of the Roman Empire*.
- Hansen, G., Prescott, E., 2002. Malthus to Solow. *The American Economic Review* 92, 1205–1217.
- Hoffman, P.T., 2013. *Why Did Europe Conquer the World?* Princeton University Press.
- Ladurie, L.R., 1974. *The Peasants of Languedoc*. University of Illinois Press.
- Landes, D.S., 2003. *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge University Press.
- Levine, D., Modica, S., 2013. Anti-Malthus: conflict and the evolution of societies. *Research in Economics* 76, 289–306.
- Levine, D.K., Modica, S., 2016. Dynamics in stochastic evolutionary models. *Theoretical Economics* 11, 89–131.
- Levine, D., Modica, S., 2018. Intervention and peace. *Economic Policy*, 361–402.
- Levine, D., Modica, S., 2019. Survival of the weakest: why the West rules. Working paper. www.dklevine.com.
- Lin, J., 1995. The Needham puzzle: why the industrial revolution did not originate in China. *Economic Development and Cultural Change* 43 (2), 269–292.
- Liu, Y., Liu, C., 2007. Diagnosing the cause of scientific standstill, unravelling the Needham puzzle. *China Economist* 10, 83–96.
- Maddison, A., 2007. *Contours of the World Economy 1-2030 AD*. Oxford University Press.
- Young, P., 1993. The evolution of conventions. *Econometrica* 61, 57–84.