NBER WORKING PAPER SERIES

THE PARADOX OF CIVILIZATION: PRE-INSTITUTIONAL SOURCES OF SECURITY AND PROSPERITY

Ernesto Dal Bó
Pablo Hernández
Sebastián Mazzuca

Working Paper 21829
http://www.nber.org/papers/w21829

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2015

We thank Demian Pouzo, Santiago Oliveros, Bob Powell, Alvaro Sandroni and David Schönholzer for valuable discussion, as well as seminar and conference participants for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Paradox of Civilization: Pre-Institutional Sources of Security and Prosperity
Ernesto Dal Bó, Pablo Hernández, and Sebastián Mazzuca
NBER Working Paper No. 21829
December 2015
JEL No. D74,N4,Z1

ABSTRACT

The rise of civilizations involved the dual emergence of economies that could produce surplus ("prosperity") and states that could protect surplus ("security"). But the joint achievement of security and prosperity had to escape a paradox: prosperity attracts predation, and higher insecurity discourages the investments that create prosperity. We study the trade-offs facing a proto-state on its path to civilization through a formal model informed by the anthropological and historical literatures on the origin of civilizations. We emphasize pre-institutional forces, such as physical aspects of the geographical environment, that shape productive and defense capabilities. The solution of the civilizational paradox relies on high defense capabilities, natural or manmade. We show that higher initial productivity and investments that yield prosperity exacerbate conflict when defense capability is fixed, but may allow for security and prosperity when defense capability is endogenous. Some economic shocks and military innovations deliver security and prosperity while others force societies back into a trap of conflict and stagnation. We illustrate the model by analyzing the rise of civilization in Sumeria and Egypt, the first two historical cases, and the civilizational collapse at the end of the Bronze Age.

Ernesto Dal Bó
University of California, Berkeley
Haas School of Business
545 Student Services Building #1900
Berkeley, CA 94720-1900
and NBER
dalbo@haas.berkeley.edu

Pablo Hernández
New York University Abu Dhabi
PO Box 903
New York City, NY 10276-0903
pablo.hernandez@nyu.edu

Sebastián Mazzuca
Johns Hopkins University
Department of Political Science
338 Mergenthaler Hall
3400 N. Charles Street
Baltimore, MD 21218
smazzuca@jhu.edu

1 Introduction

Anatomically modern humans have lived in subsistence and stateless societies for roughly 97% of their 200,000-year-long history. If there is a Big Bang in human history, it occurred as recently as 5,000 years ago, when the first civilizations emerged. Civilization meant fundamental transformations: systematic surplus production, urbanization, public architecture, writing, and states. Although the rise of civilization is arguably more of a qualitative change than the Industrial Revolution, modern political economy has paid much less attention to it.

According to an influential view in archaeology, the rise of civilizations is primarily driven by an exceptional potential for food production, both in terms of endowments and technology. For V. Gordon Childe, the key features of Lower Mesopotamia, the "cradle of civilization," were an extremely fertile alluvial soil, an abundance of edible animals, and irrigation technology. Identical factors were emphasized for the rise of Egypt, the first pristine civilization after Sumer. Both in Lower Mesopotamia and Egypt "*irrigation agriculture could generate a surplus far greater than that known to populations on rain-watered soil*" and "*as productivity grew, so too did civilization*" (Mann 1986: 80, 108).

Without a substantial surplus, it was not possible to fund the tangible components of civilizations. However, surplus production was only a necessary condition for civilization, not a sufficient one. In fact, prosperity could be self-defeating. Primitive food producers were surrounded by nomadic tribes for whom agricultural surpluses were a most tempting target for looting. The resulting clash is a primordial conflict shaping the civilizational process. According to McNeill (1979, p. 71), "*Soon after cities first arose ... the relatively enormous wealth that resulted from [their economic activity] made such cities worthwhile objects of attack by armed outsiders.*" For anthropologists, intergroup violence had been prevalent since before civilization (Keeley 1996), but the emergence of large surpluses intensified the potential for conflict. According to Michael Mann, "*the greater the surplus generated, the more desirable it was to preying outsiders*" (1986: 48).

Since civilization entailed the joint achievement of prosperity and security, its emergence is a fundamental paradox. Primitive societies that held production close to subsistence levels could hope to mitigate predation, but stagnation would foreclose the civilization process. To reach civilization, primitive societies with the capacity for surplus production had to overcome the dangers of self-defeating prosperity without relying on the relative safety of stagnation. A proper balance was needed between surplus production and surplus protection.

The contextual conditions allowing for such a balance are rare, as evidenced by the fact that, out of thousands of primitive societies, only a handful could develop independent civilizations, starting with Sumer and Egypt. In this paper we develop a model to identify the logical conditions for successful civilization, and examine for the first time the historical record for the rise of Sumer and Egypt under the perspective of the civilizational paradox. The historical cases illustrate the logic of the model, and the model allows for a richer interpretation of the cases. Sumer and Egypt provide evidence that the potential for surplus emphasized by archaeologists and geographers was only half of the story of successful civilization. The other half was surplus protection. In addition to their historical preeminence as cases achieving the right balance between prosperity and security, Sumer and Egypt illustrate that protection occurs in two contrasting ways—defense can be natural as in Egypt, or man-made as in Sumer.

The rise of civilization, and its intrinsic paradox, can be usefully compared to the rise of the modern state in the post-Westphalia context, another major turning point in history. The rise of the modern state also involves a paradox. European rulers striving for a monopoly of violence were able to reach unprecedented levels of power but in the process they undermined their own ability to credibly commit to respecting private rights. Unstable property rights in turn diminished the capacity of the underlying society to grow, and ultimately damaged the ruler's own power. The standard insight is that the solution for the modern state was "institutions" understood as rules of the political game: checks and balances, as well as the expansion of political rights, helped the ruler to solve its credibility problem either vis-à-vis society at large or vis-à-vis competing factions within the elite (North and Weingast 1989, Acemoglu and Robinson 2005, Lizzeri and Persico 2004).

In contrast to the solution to the paradox of the modern state, the solution to the paradox of civilization in our approach does not involve institutions. The joint achievement of order and prosperity in the context of pristine civilizations is a pre-institutional process, involving tangible assets and technologies, of either economic or military nature. Pristine civilizations emerged in areas with exceptional natural endowments for food production, and the man-made contributions to the civilizational breakthrough were not political rules, as institutional theories would emphasize, but productive and defense equipment. Two massive engineering accomplishments are the mark of the Ancient Near East: irrigation infrastructure in both Egypt and Sumer, and perimetral walls in cities throughout Mesopotamia and the Levant. Each public good had a single, well-defined mission: surplus production and surplus protection. The prominence of the two types of public works reflects the centrality of the

production-protection tension in the process of civilization building.

Our pre-institutional theory on the joint achievement of security and prosperity can help improve our understanding of the rise of the first civilizations, and shed light on the problem of state formation more generally. A broader goal is to generate insights for a wide class of development trajectories in which a potentially prosperous region, being surrounded by predatory threats, may fall in the traps of security-enhancing stagnation or self-defeating prosperity. This class includes the interaction between a large number of proto-cities and barbarian invaders from the steppes across the Eurasian continent throughout the Middle Ages; the long struggle in 19th century Latin America between elites from port-cities engaged in nation-building and rural warlords, "caudillos," dominating the periphery; as well as contemporary state-building efforts in failed states of Sub-Saharan Africa and the Middle East, in which international economic aid, if not coupled with military buildup, may have counter-productive effects by inducing voracity among neighbors. Echoing concerns in history and anthropology about the reversibility of gains in social complexity that lead to statehood, our theory provides an account for civilization collapse, and more generally for economic or military reversals in societies that had achieved prosperity and security. For illustration, we will use the model to account for the End of the Bronze Age, a much-debated process in which dozens of civilization centers collapsed rather quickly throughout the Eastern Mediterranean, ushering in the first "dark ages" in the historical record.

1.1 Overview of the model

In our model, a population in control of an economy with the potential to create surplus (the "incumbent") faces potential attacks by a predatory group (the "challenger"). In this setting, the incumbent has the opportunity to invest and grow future income, which would lead to "prosperity." However, the incumbent may prefer to spend resources in consumption and defense if future flows may be lost after a successful attack by the challenger. Three parameters are fundamental in the model: initial productivity, 'defense capability' and 'growth capability.' The two capabilities are the rates at which current consumption can be transformed into defense and future income, respectively. The three parameters regulate fundamental tradeoffs. Higher initial productivity helps finance more defense, but it also attracts stronger predation. If sufficient defense can be financed, the challenger can be deterred. If attacks cannot be deterred, the rate of return to economic investment may not be high enough to justify surplus production. Moreover, productive investment, a pre-

condition for systematic surplus production and the construction of the durable structures associated with civilization, may intensify predatory challenges by raising income. The result is a tradeoff between investment-led growth and security.

Our analysis has two main parts. In the first part, the incumbent's defense capability is exogenous, and in the second part defense capability can be improved. Both parts help to account for different aspects of the rise of ancient civilizational states, and more generally the variety of development paths taken by societies that have a potential for growth but face threats.

The analysis with exogenous defense capability begins by characterizing the unique equilibrium of the game. Since productive investment fosters attacks, it is not obvious that civilizational breakthrough is possible. Therefore, the key formal questions are whether some combination of parameter values allows for productive investment, and whether productive investment requires fully deterring the challenger. A key result is that the parameter space is partitioned into four regions corresponding to the four "prosperity/security" combinations.

The relationship between security and prosperity has been a perennial concern in the social sciences. A dominant view, inspired by Hobbesian philosophy, is that state-provided security is a precondition for prosperity (Lane 1958; Olson 2000; Bates 2001; see Boix 2015 for a contrasting approach). But the state itself has to be explained and the Hobbesian view provides no clear message on whether state formation requires a modicum of prosperity in the first place. In our model, when both defense and growth capabilities are low, neither prosperity nor security are possible, and societies remain locked in the situation of economic stagnation and conflict characterized by Keeley (1996), which corresponds to the Hobbesian "state of nature." If investment returns are high relative to defense capability, prosperity becomes possible even in the face of attacks. Although anti-Hobbesian, the possibility of prosperity without full security is consistent with a widespread occurrence in the history of humanity: populations that prefer to grow their economies rather than attaining full deterrence despite threats of predation from neighboring plunderers, like the Chinese with the Mongolians and the Saxons with the Vikings in the 10th century. Lastly, when both defense and growth capabilities are high and "balanced," the society can manage both to grow and deter predators. The latter two analytic possibilities are the key to our explanation for the emergence of civilizations. Civilizations occur in cases where high enough returns to productive investment allow the economy to grow, and where the incumbent manages to deter attacks by the challenger, or, if attacks occur, to repel them with reasonably high probability (complete security is an elusive feature in the historical record).

The case of Egypt can be explained in terms of natural endowments for both growth and (exogenous) defense. Growth capabilities were given by rich alluvial soils that could be quickly improved through productive investments, and defense was provided by the surrounding deserts, which protected dwellers along the Nile from most types of attack (Bradford 2001).

The model allows us to study how shocks (natural, or policy-originated) to defense and growth capabilities can generate transitions from one area to another in the security-prosperity outcome space. A rich picture of transitions where the effect of shocks depends on initial conditions emerges. Such shocks can account not only for the rise of civilizations, but also for their fall and the emergence of dark ages where security and prosperity are lost. We exemplify this type of application with a study of the end of the Bronze Age around 1200BC. Our model shows that enhanced defense capabilities are a necessary condition for achieving security and prosperity, but expanding growth capability, while valuable, is not strictly necessary. Moreover, under certain conditions, an imbalanced mix may worsen outcomes.

The rise of civilization in Southern Mesopotamia poses a challenge to our model with exogenous defense capability, however, since the Sumerian settlements, in contract to Egypt, did not have natural protection. Rather, as widely attested in the archaeological record, they faced challenges from various pastoralist groups. How could these first city-states ever emerge? The picture put forward by the anthropological literature is that the settled groups who developed pristine states exploited an agrarian "staple finance", which, being highly rewarding, would fund their defense (Johnson and Earle 2000, p. 305-306). These groups had enough of a material advantage that could be turned into a military one, by relying on walls, weaponry, and numbers, all of which could be used to deter or defeat their enemies. This process of endogenous improvement of defense capabilities can be accounted for in our extended model, where the incumbent can make investments not only to expand production but also to upgrade its defense. The key result is that when the initial productivity is high enough the incumbent can fund its way out of the region without security or prosperity into a region with high levels of both.

While higher initial productivity always exacerbates conflict in the model with exogenous defense capability, in the model with endogenous defense it may pave the way to security and prosperity. That this should happen is not obvious, since improvements in defense capability are an investment, and as such they are discouraged by the insecurity associated with higher initial productivity. When stronger defense capability is put in place, a Hobbesian effect is observed: the enhanced security yields a higher effective return to productive investment

and it fosters prosperity.

1.2 Plan for the paper

In the next section we relate our contribution to two broad literatures, one in political economy of state formation and political sources of prosperity, and one in history and anthropology on the origin of civilizations. In section 3 we present the model with exogenous defense capability, and use it in section 4 to analyze the rise of Egypt and the end of the Bronze Age. In section 5 we extend the model to allow for endogenous defense capability, and in section 6 we use the extended model to account for the rise of Sumeria. We conclude in section 7.

2 Related Literature

Historians, archaeologists and anthropologists have emphasized a variety of factors to explain the rise of civilizations, including natural endowments for food production and the construction of physical defense against outsiders. Archaeologists like V. Gordon Childe (1936), who first conceptualized the advent of the Neolithic era as an "agricultural revolution," focused on the innovations in the means and relations of production while abstracting from the necessary accompanying innovations in military protection. On the other hand, several archaeologists have noted the paramount role of investments in protection, such as fortifications, walls, and moats, in the erection of the first cities (Service 1975, 299). According to Near Eastern archaeologist Volkmar Fritz, "*in the Jordan Valley, settlements were surrounded by a wall even before it is possible to speak of the city proper*" (1997 II: 19). Other authors, like Mann (1986) and McNeill (1979), explicitly connected food production with protection needs, as mentioned earlier. However, we are not aware of any account that has explicitly focused on the interplay of surplus production and surplus protection to point out a solution to the civilizational paradox. As we will show, the interplay is subtle and perhaps profitably analyzed through a formal model.

Our solution to the civilizational paradox highlights the interaction of growth and defense capabilities, natural or man-made. Our approach to civilization builds on, but departs from, historical accounts that emphasize the geographic and institutional sources of economic prosperity and income differences across societies. Historical approaches emphasizing the availability of domesticable plants and animals to explain why some regions generated

surpluses while others did not (e.g., Diamond 1999) contribute a necessary building block for understanding the prosperity of the first settled societies. However, a purely geographic approach is incomplete, for it misses the role of incentives and strategic action that is at the core of any viable socio-political account of the origin of civilizations. Our approach incorporates both strategic actors and an environment where fundamentals can be interpreted as reflecting geographic factors such as food production potential or features of the terrain that afford protection against attacks.

Our approach makes progress at the cost of abstracting from other aspects that have been considered in anthropological theories of the state. We will not review these theories comprehensively, but comment on those featuring relevant parallels or departures from ours.[1] An influential view in anthropology, which refines classic Marxist insights, is that the state emerged as an instrument to sustain and expand economic inequality in a context of increasing social stratification (Fried 1960, 728). However, we abstract from social hierarchy and the varieties of forms of rule: in our model the incumbent can be interpreted as an egalitarian community, a benevolent ruler representing the population, or a perfectly despotic dictator acting as the residual claimant. This is not because we think political stratification is unimportant, but because it helps to focus attention on the incumbent-challenger interaction. For Carneiro (1970), early states originated in pockets of fertile land surrounded by areas less suitable for food production. Population growth induced groups to fight for the scarce fertile land, and losers who were unable or unwilling to flee had to accept political domination. The ensuing political stratification is the basis of the state. The Nile valley, surrounded by deserts, is a good example of circumscribed productive land. Our model generates a similar empirical implication; however, it is not driven by intra-societal exploitation but by the fact that surrounding land of very low quality can act as a barrier against challengers. Unlike Carneiro's theory, our model does not appeal to population pressure, an assumption that has been challenged by some writers (see Allen 1997).

It is customary in the social sciences in general and in political sociology in particular to view the state as the monopoly on violence. Departing from contemporary theories of state formation, and adapting from Weber, we define the state not in binary terms but as a matter of degrees (Weber 1978: ch. I, s. 16). In particular, state formation involves attaining higher degrees of protection from attacks. We deliberately focus on the state as "sovereignty," defense of a surplus-producing society from threats by non-producers (inside or outside the territory), and abstract from "rulership," the creation of a political hierarchy

[1]For a review of anthropological theories of early states see for example Claessen and Skalnik (1978).

within a society. Our definition of the state is pre-institutional in the sense that "state" in our analysis is pure military force and abstracts from the rules that regulate the access, the exercise or the division of its power, which even critics of the institutional approach include in their definition (see Boix 2015: 66-77). The exclusion of both rulership and political institutions from our model helps identify how early civilizations could resolve the security-prosperity paradox. Civilization is the intersection of surplus production and statehood seen as sufficient surplus protection.

Our work is related to both theories of state formation (Tilly 1975, 1992, Spruyt 1996) and theories of the political sources of prosperity (North and Weingast 1989; Olson 2000, Bates 2001; Boix 2015). In contrast to our model, theories of state formation do not place the state in the context of the "security-prosperity" tradeoff, and theories of the political sources of prosperity focus on rules of the political game once the state is already in place rather than on pre-institutional forces.

Olson (2000) and Boix (2015) are noteworthy in that they relate state formation to long run development. According to Olson (2000) states emerged when roving bandits were replaced by stationary bandits who preferred to limit rent extraction and provide productivity-enhancing public goods.[2] Our focus is very different for we abstract from rulership. We share with Boix (2015) an interest in mechanisms of state formation and economic prosperity that extend back into prehistoric times, as well as a focus on "hard" causes like geographic factors, military and economic capabilities shaped by the physical environment. Our model complements Boix's in several ways. It is especially built to understand the emergence of pristine civilizations, which is not as central to Boix's analysis. Although Boix finds sources of pre-institutional cooperation under conditions of anarchy (absence of state), he conceives of state formation as the selection of either republican or monarchic institutional settings. By contrast, we focus on the properties of state formation that allow for productive investment before political institutions become central. If Boix's distinct focus is on the link between state formation and economic inequality, ours is on the link between state formation and the civilizational paradox. Our model is explicitly focused on the trade-off between security and prosperity, which allows for a distinct partition of the parameter space capturing biogeographic and technological features of the context.

Our work has important complementarities with the work by Mayshar, Moav and Nee-

[2]See Sanchez de la Sierra (2014) for an investigation of how these incentives shaped decisions by armed gangs in the Eastern Congo, who responded to the presence of taxable resources by seeking to monopolize violence.

man (2013), and Mayshar, Moav, Neeman and Pascali (2015). They also combine a focus on early states, an emphasis on geographic drivers, and the use of formal theory. For us, geography matters because it defines both productive and defense capabilities, while for them it determines the observability of production (the former paper) or its appropriability (the latter). Mayshar, Moav and Neeman (2013) use a principal-agent model to show how monitoring capabilities shape the extent of political centralization, and account for contrasting trajectories in Sumeria and Egypt, where observability of the Nile allowed for a more unified and lasting state. Our focus is not on the form of states, but on the conditions for their emergence. This is also the focus of Mayshar, Moav, Neeman and Pascali (2015), who focus on the appropriability of different crops instead of general productivity. The production of cereals, which in contrast to tubers is highly appropriable, creates a demand for protection and makes taxation feasible. They equate the state with the political hierarchy that results from appropriability and assume it results in the full prevention of conflict. We abstract both from appropriability differentials and from issues of internal hierarchy, but investigate the conditions under which conflict can be reduced or eliminated.

A recent literature studies the incentives of rulers to make investments in state capacity in situations where the ruler may lose control of the polity to a competing faction (Besley and Persson 2011), or to a foreign power (Gennaioli and Voth 2015). Gennaioli and Voth (2015) formalize Tilly's (1990) argument that modern European states formed as a result of the competitive pressures of military conflict, which created a need to centralize fiscal control. (They also find empirical support for the idea that fiscal centralization was conducive to successful state building.) There are some differences in terms of modeling: unlike in Besley and Persson's model, investments in our model can augment the virulence of challenges, and we abstract from the competitive dynamics between states that are at the core of Gennaioli and Voth's analysis. There are differences in substantive focus as well. In the state capacity literature there is a pre-existing state, while we focus in pre-state societies that move towards statehood by attaining some degree of deterrence. Our paper is also related to the formal study of state consolidation. Powell (2012) offers a treatment where state consolidation happens exogenously, while Powell (2013) considers endogenous consolidation. The key difference is that in our model consolidation is studied in relation to investment and growth.

3 The Basic Model

3.1 Setup

Our baseline model features the incentives to raise an army to protect wealth from usurpers at the cost of resources for consumption or productive investment. Later on we introduce the decision to invest in defense capability.

Players

An "incumbent" controls a productive asset that yields a nonstorable flow $v_t > 0$ every period. The asset can be a piece of land, a port, or any bundle of productive resources including people. The initial level v_1 tracks properties of the environment (e.g., weather, quality of the soil, topography) that affect the quantity and value of goods that the economy can produce, i.e., productivity. A "challenger" receives an exogenous income flow from nature that we normalize to zero, and is interested in wresting control of the productive asset away from the incumbent. This interaction captures the large class of cases of inchoate urban centers (agricultural settlements, city-ports, markets at the crossing of interior roads) where food-producing populations and/or trading elites face the threat of predatory attacks by nomadic tribes or plundering warlords. Although we emphasize external threats, the challenger can also include recalcitrant elements within the incumbent population who undertake predatory or subversive activities.

Actions, resources and technology

There are two periods $t = 1, 2$. We describe actions generically although it should be kept in mind that there is no future in period 2. In each period the incumbent can spend its flow v_t on consumption, productive investment i_t or mobilizing resources to defend its asset. One dollar of productive investment i_t costs one dollar of consumption and it adds $\rho > 1$ dollars to the yield of the productive asset in the future.[3] That is, productivity evolves according to the relation $v_{t+1} = v_t + \rho i_t$; we abstract from depreciation and discounting for simplicity. ρ captures anything that affects the returns to productive investments in the asset controlled by the incumbent. For example, ρ could, like v_1, reflect climatic conditions, soil fertility, and other features of the environment, or the price of goods sold.[4]

[3]Given that for $\rho < 1$ investment is never worthwhile, failure to obtain it in equilibrium is obvious and uninteresting. Hence our assumption $\rho > 1$ which makes investment at least a possibility.

[4]If the value of what the incumbent produces follows a standard price $\times$ quantity formulation we can write $v_1 = p.q$, and $v_2 = v_1 + \rho.p.i = p.q + \rho.p.i$, where q and i are physical units. Then, changes in p will cause changes in both v_1 and ρ. Changes in the baseline physical capacity of production q will be captured through changes in v_1 exclusively, and changes in the physical returns to investment as changes in ρ only.

The effectiveness of the incumbent's defense (or "army") is denoted a_t and such an army costs the challenger an amount $\frac{a_t}{\kappa_t}$ where $\kappa_t \geq 0$ is the value of the incumbent's defense capability. The higher the defense capability of the incumbent, the higher the "firepower" a_t attained by a given conflict effort $\frac{a_t}{\kappa_t}$. In this section κ_t is exogenous and we will derive implications for conflict and prosperity stemming from different values of κ_t. The expanded version of the model in section 5 will be devoted to endogenizing κ_t. Thus, in period t the incumbent must observe a budget constraint,

$$v_t - i_t - \frac{a_t}{\kappa_t} \geq 0. \tag{1}$$

κ_1 captures anything that yields the incumbent an advantage at producing defense or military firepower at a given expense, such as a rugged terrain or better military technology or expertise.[5]

The challenger observes the choices of a_t and i_t by the incumbent and chooses its own conflict effort b_t.[6] If victorious in the first period the challenger captures control of the productive asset in the second period. Whenever the challenger attacks ($b_t > 0$), it prevails with probability $\frac{b_t}{a_t+b_t}$ and it gains nothing with the complementary probability (i.e., we adopt the typical Tullock contest success function).[7] If the incumbent is defeated it obtains an outside payoff normalized to zero; the challenger becomes the new incumbent. If the challenger selects $b_t = 0$ we say the incumbent has successfully deterred the challenger, and this lack of challenge to the authority of the incumbent results in full security. As we explain later, we associate the degree of security–be it in terms of internal order or external sovereignty–with the degree of statehood, so full security corresponds to full statehood.

Timing

In each period the incumbent selects a_t and i_t. After observing (a_t, i_t) the challenger selects b_t. If $b_1 = 0$, the players retain their positions in period 2. If $b_1 > 0$, then there is conflict at the end of period 1. The winner becomes the incumbent in period 2.[8]

[5]Some types of infrastructure (e.g., roads) may affect both ρ and κ_1: a road may increase the returns to investing in a port, and it may also make the incumbent's army more (or less, depending on circumstances) effective.

[6]Assuming that the challenger's war expense is basically effort is equivalent to assuming that the challenger's income is sufficient to finance the optimal war effort b_t^*. Because the effects of interest are not driven by a budget constraint on the challenger being binding, we follow the most parsimonious approach of not making explicit a resource constraint on the challenger.

[7]Generalized versions of the ratio-based contest success function exist but are less tractable. Hirshleifer (2001) explores some of the difficulties.

[8]In the two period model it makes no difference whether we assume that the new incumbent faces a new

Payoffs

Both challenger and incumbent are risk neutral and care linearly about income and units of effort. The incumbent acts as a Stackelberg leader, choosing a_1 and i_1 to maximize the value of the game for an incumbent V_t:

$$V_t = v_t - \frac{a_t}{\kappa_t} - i_t + \frac{a_t}{a_t + b_t} V_{t+1}. \tag{2}$$

The challenger chooses b_t to maximize the expression

$$W_t = \frac{b_t}{a_t + b_t} V_{t+1} - \frac{b_t}{\kappa_c}, \tag{3}$$

where $V_{t+1} = v_t + \rho i_t$, and where κ_c is the military capability of the challenger. We could also parametrize the challenger's objective with a factor h and write the expected benefit as $\frac{b_t}{a_t+b_t} h V_{t+1}$, so as to capture different levels of "hunger" by the challenger.[9] Although capturing a different substantive aspect, the parameter h would be mathematically redundant since the challenger's problem could be rewritten as involving a military capability of $h\kappa_c$ instead. Therefore we will abstract from the parameter h. To simplify notation, we will develop the model normalizing $\kappa_c = 1$, and will comment later on how the solution changes with variations in κ_c. An additional simplification is we do not consider here the realistic possibility that conflict destroys part of the asset. Our results do not change qualitatively by assuming that conflict is destructive.[10]

We will solve for a Subgame Perfect Nash Equilibrium by backward induction.

3.2 Solution

Second period

The rewards from conflict accrue in the next period if any, so the challenger does not fight in the second and last period and $b_2 = 0$. Anticipating this the incumbent chooses $a_2 = 0$

challenger in period 2, since in that period there are not incentives to fight.

[9] This parameter could also track the differential ability of the challenger at "operating" the asset. One issue we do not take up here is the case where a challenger has a high valuation for the stream of production (as when looting animals and food) but a low valuation for the asset due to an inability to operate it. These are interesting variations that go into the finer issue of modes of challenge that may be costly to the incumbent but do not pose a replacement threat. The study of these variations is left for future research.

[10] The model presented here represents the limit case of a more general model where a fraction $\sigma \in [0, 1]$ of the asset survives the war. The solution to the expanded model is similar and continuous in σ, so the solution we focus on remains qualitatively similar when σ dips below 1 (proof available upon request).

and selects i_2 to maximize the value of consumption in the second period $V_2 = v_2 - i_2$, yielding $i_2 = 0$ and $V_2 = v_2$.

First period

The challenger observes the pair (a_1, i_1) and chooses b_1 to maximize W_1 as given by expression (3). Since the first order condition is $\frac{a_1}{(a_1+b_1)^2} v_2 = 1$, and $v_2 = v_1 + \rho i_t$, the best response function of the challenger is,

$$b_1(a_1, V_2) = \begin{cases} \sqrt{a_1 (v_1 + \rho i_1)} - a_1 & if\ a_1 < V_2 \\ 0 & otherwise \end{cases}. \tag{4}$$

This expression exhibits a key trade-off of the model: productive investments i_1 raise the value of the productive asset. Thus, conditional on maintaining control of the asset, investment is a good idea for the incumbent since $\rho > 1$; however, the future control of the asset is not a forgone conclusion. Investment raises the incentives of the challenger to arm itself since it makes it more attractive to become the incumbent. Therefore, while productive investments increase the value of future incumbency, they may lower the chance that the current incumbent gets to reap that value. This is the civilizational paradox: future prosperity raises insecurity, which in turn depresses incentives to invest and undermines the creation of that future prosperity. The civilizational paradox is analytically related to Hirshleifer's (1991) paradox of power, according to which the poorer contender acts more aggressively. Against this backdrop, our task is to understand whether there are any parameter values v_1, κ_1, and ρ that map into security and prosperity. To answer this question we must study the problem of the incumbent.

The incumbent maximizes V_1 as given by (2) subject to the budget constraint (1) and anticipating the challenger's best response in (4). The latter indicates that if $a_1 \geq v_1 + \rho i_1$ the challenger will choose not to fight, and therefore the incumbent would never choose a_1 beyond the point $v_1 + \rho i_1$, which attains deterrence. This can be incorporated into the incumbent's problem as an additional, deterrence constraint. The incumbent's problem in period 1 can then be written as,

$$\max_{a_1, i_1} \left\{ v_1 - \frac{a_1}{\kappa_1} - i_1 + \frac{a_1}{a_1 + b_1}(v_1 + \rho i_1) \right\} \tag{5}$$

subject to

$$
\begin{aligned}
v_1 - \frac{a_1}{\kappa_1} - i_1 &\geq 0 \ (BC) && (6)\\
v_1 + \rho i_1 - a_1 &\geq 0 \ (DC) && (7)\\
a_1 &\geq 0 && (8)\\
i_1 &\geq 0, && (9)
\end{aligned}
$$

where (BC) is the incumbent's budget constraint and (DC) is the deterrence constraint. The Lagrangian, which expresses the expected utility of the incumbent, is:

$$
\begin{aligned}
\mathcal{L} &= v_1 - \frac{a_1}{\kappa_1} - i_1 + \frac{a_1}{a_1 + b_1}(v_1 + \rho i_1) \\
&\quad + \lambda_{BC}(v_1 - \frac{a_1}{\kappa_1} - i_1) + \lambda_{DC}(v_1 + \rho i_1 - a_1) + \lambda_a a_1 + \lambda_i i_1, \qquad (10)
\end{aligned}
$$

where λ_{BC}, λ_{DC}, λ_a and λ_i are the Lagrange multipliers for each constraint (1)-(9). We will characterize the solution $(a_1, i_1, \lambda_{BC}, \lambda_{DC}, \lambda_a, \lambda_i)$ to this problem for each parameter combination (ρ, κ_1, v_1). The first order and complementary slackness conditions that characterize the optimum are given by,

$$
\frac{\partial \mathcal{L}}{\partial a_1} = \frac{1}{2}\sqrt{\frac{v_1 + \rho i_1}{a_1}} - \frac{1}{\kappa_1} - \frac{\lambda_{BC}}{\kappa_1} - \lambda_{DC} + \lambda_a = 0; a_1 \geq 0, \lambda_a \geq 0, \lambda_a a_1 = 0 \text{ c.s.} \qquad (11)
$$

$$
\frac{\partial \mathcal{L}}{\partial i_1} = \frac{\rho}{2}\sqrt{\frac{a_1}{v_1 + \rho i_1}} - 1 - \lambda_{BC} + \lambda_{DC}\rho + \lambda_i = 0; i_1 \geq 0, \lambda_i \geq 0, \lambda_i i_1 = 0 \text{ c.s.} \qquad (12)
$$

$$
\lambda_{BC}(v_1 - \frac{a_1}{\kappa_1} - i_1) = 0 \text{ c.s.}, \ \lambda_{DC}\left(v_1 + \rho i_1 - a_1\right) = 0 \text{ c.s.} \qquad (13)
$$

Solving the program (10) requires checking which combinations of values for the endogenous variables $(a_1, i_1, \lambda_{BC}, \lambda_{DC}, \lambda_a, \lambda_i)$ constitute the optimum for different regions of the parameter space $(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+$. Note from (11) that the marginal benefit of a_1 goes to infinity as a_1 goes to zero (a typical feature of contests), so the optimum must feature $a_1 > 0$ and $\lambda_a = 0$. Beyond this, the method for solving the problem is tedious: it requires checking which combinations of values for the endogenous variables are consistent with the constraints for each parametric region and also yield the highest value for the program. The following proposition summarizes the solution, the details of which can be found in the appendix.

Proposition 1 *Optimal behavior by the incumbent yields a partition of the parameter space* $(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+$ *into four distinct regions:*

Region 1 *(R1):* $\left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | \kappa_1 > \rho,\ \rho > \kappa_1/(\kappa_1 - 1)\right\}$ *Security and prosperity*
In **R1** *the solution is:* $\left\{a_1 = v_1 \frac{\kappa_1(1+\rho)}{\kappa_1+\rho}, i_1 = v_1 \frac{(\kappa_1-1)}{(\kappa_1+\rho)}, V_1 = v_1 \frac{\kappa_1(1+\rho)}{(\kappa_1+\rho)}\right\}$
Region 2 *(R2):*$\left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | \rho > \kappa_1,\ \rho > 4/\kappa_1\right\}$ *Prosperity without security*
In **R2** *the solution is:* $\left\{a_1 = \frac{\kappa_1 v_1}{2}\left(1 + \frac{1}{\rho}\right), i_1 = \frac{v_1}{2}\left(1 - \frac{1}{\rho}\right), V_1 = \frac{v_1}{2}\left(1 + \frac{1}{\rho}\right)\sqrt{\rho\kappa_1}\right\}$
Region 3 *(R3):*$\left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | 2 > \kappa_1\,,\ \rho < 4/\kappa_1\right\}$ *Neither prosperity nor security*
In **R3** *the solution is:* $\left\{a_1 = v_1\left(\frac{\kappa_1}{2}\right)^2, i_1 = 0, V_1 = v_1\left(1 + \frac{\kappa_1}{4}\right)\right\}$
Region 4 *(R4):*$\left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | \kappa_1 > 2,\ \rho < \kappa_1/(\kappa_1 - 1)\right\}$ *Security without prosperity*
In **R4** *the solution is:* $\left\{a_1 = v_1, i_1 = 0, V_1 = v_1\left(2 - \frac{1}{\kappa_1}\right)\right\}$.

The following figure contains a graphical representation of the solution.

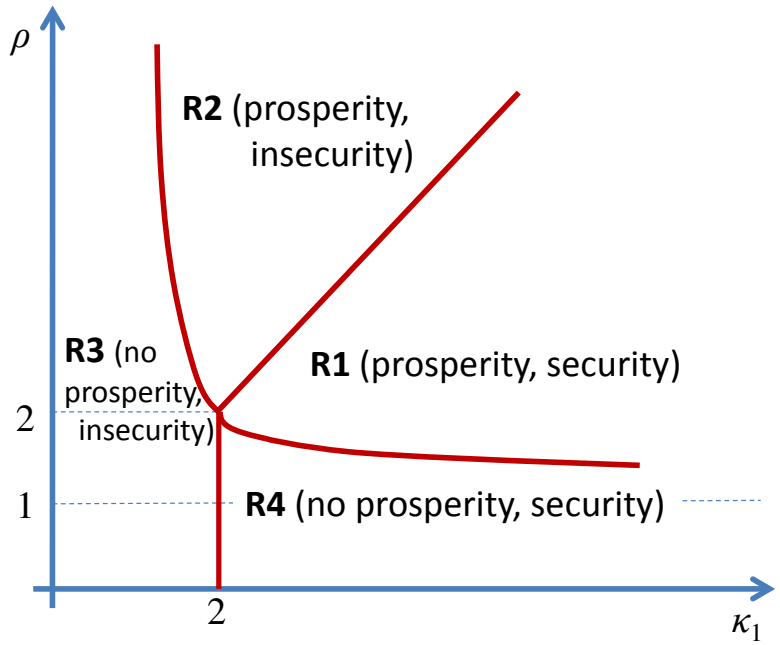


Figure 1: Equilibrium partition of the parameter space (Proposition 1)

A convenient feature of this model is that the optimal decisions by the incumbent on defense effort a_t and productive investment i_t are invariant in v_1. This feature greatly simplifies the characterization of emerging "regimes" with exogenous defense capability, as we can restrict attention to the bidimensional space (κ_1, ρ).

The main feature of the solution is that all four combinations of security and prosperity can be observed depending on the values of the parameters (κ_1, ρ). For low values of both defense capability and yield of investment, the incumbent will be stuck in a situation of economic stagnation and conflict ($R3$). In $R3$ the prospect of conflict lowers the rate of return to investment, preventing investment and hence growth. If defense capability κ_1 is higher but investment returns ρ are still low, the incumbent will be in region **R4**, where the challenger is deterred but there is no investment. In this area growth is foreclosed not by existing conflict but by the fact that if growth were attempted the challenger would become more aggressive, which would raise the costs of maintaining deterrence. If returns ρ are relatively high and defense capability κ_1 relatively low–i.e., in **R2**–growth occurs despite the fact that full security is not attained. If, starting from **R2** or **R4**, defense capability κ_1 were to become sufficiently higher, the incumbent would be in **R1**. Relative to **R2**, the added defense capability helps attain deterrence, which increases net investment returns and then expands growth potential.[11] Relative to **R4**, the added defense capability makes is cheaper to attain deterrence and releases resources for growth.

Inspection of Figure 1 yields the following,

Remark 1 *From a situation of no prosperity nor security (***R3***), a large enough increase in defense capability κ_1 is a necessary and sufficient condition for attaining both prosperity and security; in contrast,*

Remark 2 *Increases in the growth capability ρ are not necessary nor sufficient for attaining both security and prosperity.*

Natural shocks could increase or decrease parameters like ρ and κ_1. An incumbent that enjoys security and prosperity in **R1** could, through a reduction in κ_1, be plunged into stagnation and conflict in **R3**. A reduction in κ_1 could be thought of as a negative shock to the incumbent's defense technology.

As said earlier, the (security, prosperity) regimes characterized in Proposition 1 are invariant in initial income v_1; that is, whether investment i_1 and arming by the challenger b_1 are positive or zero does not depend on v_1. But changes in income v_1 do affect the particular values of all endogenous variables whenever positive. In particular, we have the following,

[11] Productive investment is higher in **R1** than in **R2** whenever $v_1 \frac{\kappa_1 - 1}{\kappa_1 + \rho} > \frac{v_1}{2}\left(1 - \frac{1}{\rho}\right)$, which is always the case for $\kappa_1 > \rho$, a condition characterizing **R1**.

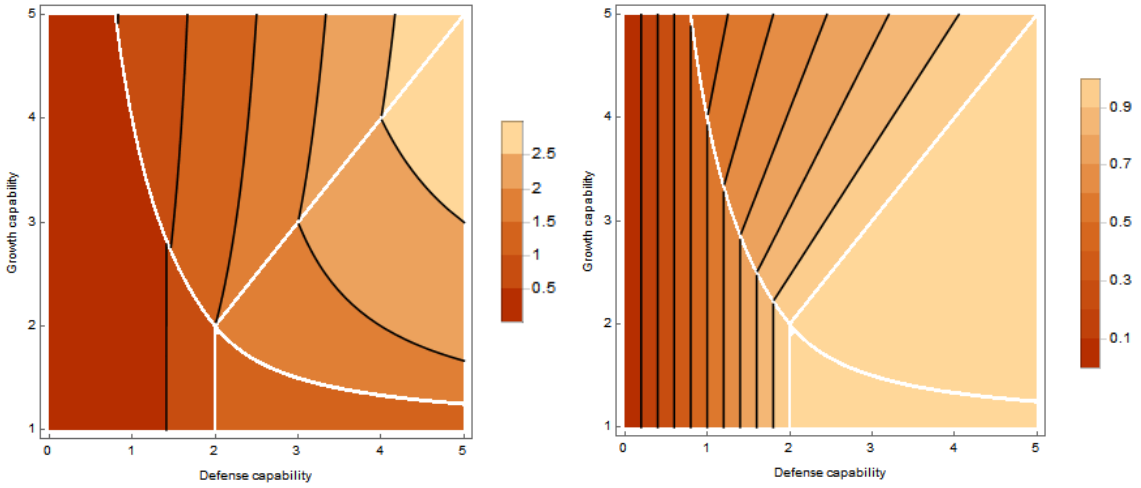
Proposition 2 *Increases in initial income v_1 exacerbate conflict; that is, in regimes where (either or both) a_1 and b_1 are positive, they increase with v_1.*

Proof: see appendix.

This result highlights one of the central forces in the prosperity-security paradox, namely the fact that a more productive incumbent that cannot fully deter its enemies will be engulfed in more virulent conflict.

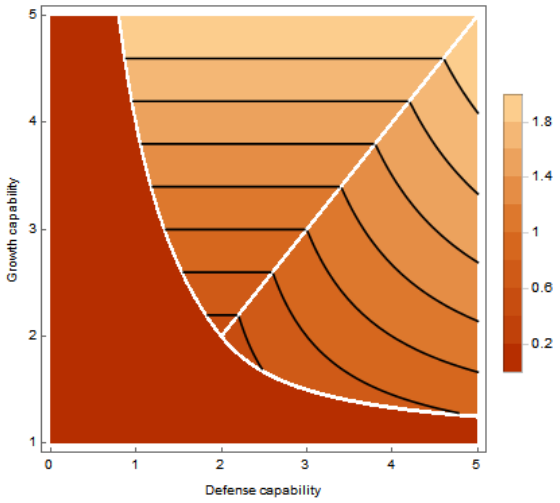
In order to connect the model to the historical record, we now relate the regions in Figure 1 to the event of a civilization rising. We defined stateness as a relative high degree of security. Figure 2 displays contour plots of relevant equilibrium magnitudes. The continuous lines within each region represent level curves, and the ligher shades of color represent higher values of the respective magnitude. Figure 2(a) shows that the arming effort of the incumbent increases as defense capability is higher, and this contributes to increasing security. Figure 2(b) represents security as proxied by the probability that the incumbent will prevail. This probability is 1 in **R4** and **R1**, and it decreases in **R2** as defense capability goes down or growth capability goes up (as this fires up the challenger). The areas in **R2** that are sufficiently close to **R1** display arbitrarily high levels of security which in our approach can be interpreted as a higher degree of stateness. In other words, we may consider the safer parts of **R2**, **R1** and **R4** as the parameter combinations that yield statehood. But civilization requires more than security; it also requires the creation of surplus, which in our model amounts to growth ($v_2 - v_1 = i_1\rho$). Figure 2(c) shows how there is no growth in **R3** and **R4** (since there is no investment) and that there is growth in **R2** and **R1**. Growth increases in returns ρ and in **R1** it also increases in defense capability, as a higher defense capability lowers the costs of arming and releases resources for investment. In **R2** growth is unresponsive to defense capability because any increase in κ_1 is met with a similar increase in a_1, which keeps the resources devoted to defense $\frac{a_1}{\kappa_1}$ and investment i_1 constant.

We defined civilization as the joint attainment of growth and security. This would leave out parts of **R2** to the North-West, bordering **R3**, where growth can be high but security low. This is sensible if we consider that civilization requires to consolidate growth by defending production from attacks. A good proxy for civilization would then be the continuation value perceived by the incumbent in period 1, which reflects both growth and security. This is the expected future income of the incumbent resulting from investment and the probability that the incumbent prevails. This combination of the magnitudes in panels (b) and (c) yields the pattern in panel (d) of Figure 2. We observe that this "intersection" of growth and security increases with both defense and growth capabilities, and indicates areas in **R1** and **R2** near

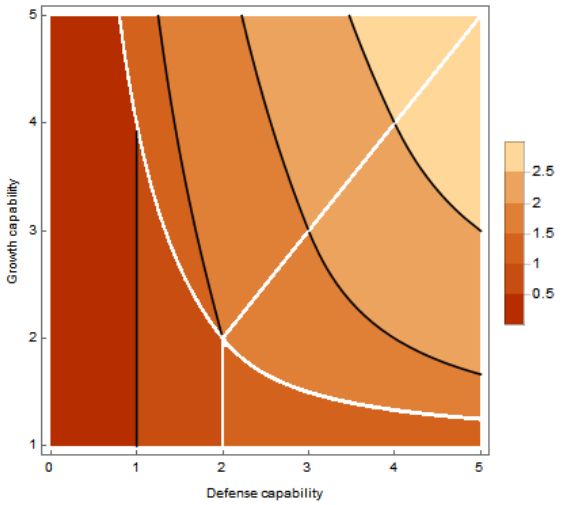


a) Army

b) Probability the incumbent prevails



c) Growth



d) Continuation value

Figure 2: Values of endogenous magnitudes in equilibrium with exogenous defense capability

the 45 degree line as those that best escape the security-prosperity paradox, and therefore as good parametric candidates for explaining the rise of civilizations.

3.3 Discussion

Internal vs external conflict and social structure of incumbent polity

The modern distinction between national an international conflict is irrelevant in our model. The process of civilization emergence precedes such distinctions. That process ended with the incorporation of the formerly hostile populations in some cases and their exclusion in others; but population integration is not the problem occupying us. If challengers are internal actors (ex ante or ex post) our definition of security is about internal order and the classic monopoly of violence. If challengers are external actors, our definision of security is best matched with the notion of sovereignty. Second, there is no distinction between ruler and subjects within the incumbent actor. The incumbent in our model can be taken to be either a representative agent in the civilized center, a perfectly benevolent ruler acting on behalf of that settled population, or a perfectly extractive ruler who is a residual claimant.

Asymmetries

We have kept as many aspects as possible symmetric between the incumbent and the challenger, and only introduced asymmetries that we deemed necessary to analyze the type of interaction of interest. One asymmetry is that the incumbent acts as a Stackelberg leader. This was convenient to generate deterrence. Another asymmetry is that while defense effort costs the incumbent resources, it does not deplete a budget for the challenger. This is for tractability. It would be possible to include a budget constraint for the challenger, and the advantage of a wealthier incumbent at being able to finance higher defense effort, and then attaining deterrence, would operate in similar fashion. However, the reaction function of the challenger would hit its constraint eventually and the analysis would become less elegant as kinks in the reaction function have to be taken into account.

4 Historical illustrations

4.1 Egypt and the birth of a state

Among the first civilizations, Egypt is the prototype of a "pristine territorial state," the undisputed pioneer in attaining both security and prosperity throughout a substantial territorial expanse. Although the Neolithic Revolution occurred in Egypt later than in

Mesopotamia, the ensuing process of social and political development in Egypt was extraordinarily fast.[12] In less than a thousand years, the outcome would be a state that not only presided over a wealthy economy but was also able to protect its territory and the surplus generated within it for long stretches of time. As Allen (1997: 135) put it, "*the Egyptian state lasted longer and was more stable than most Empires established elsewhere.*"

Although the specific conditions underlying Egypt's dual economic and political evolution are disputed, a strong consensus exists around the general idea that Egypt's geography played a key role. Our model can be used to identify three features of the Egyptian geography—the delta river, the potential for irrigation, and the surrounding desert—that collectively set it apart from other civilizations, and to assess their role in explaining the emergence of a successful state.

(1) The Nile River as a fundamental driver of the Egyptian economy. The Nile had at least two key properties: a yearly flood that fertilized the soil with rich stilt from Ethiopia, and a two-way navigability that facilitated exchange along the entire valley.[13] "*[T]he Nile was perfectly ordered—its current carried boats downstream, the wind blew them back upstream—and the Nile's regular flooding renewed the fields and made farming so easy that in the Delta men had 'only to throw out seeds to reap a crop.' "* (Bradford 2001: 9). Both properties, natural fertility and easy exchange, map into a high v_1 in our model, whereby initial income is high even before investments are made.

(2) The productivity of artificial irrigation. Egyptians could vastly increase their economic output by investing in water management, which in the Nile valley took the form of basin irrigation. Egypt developed a network of earthen banks in the agricultural fields that would form a grid of basins to trap the floodwater and hold it for much longer than it would naturally stay. The basins allowed the earth to become fully fertilized before planting, and a

[12]During the Neolithic revolution, gathering and hunting were gradually replaced by the domestication of plants and animals for food production. The process began in the ancient Near East (Southern Levant) about 10,000 years ago. The Neolithic in Egypt developed much later, around 5500 BC. According to Bard (1994, p. 267), "The beginning of the First Dynasty was only about 1000 years after the earliest farming villages appeared on the Nile, so the Predynastic period, during the 4th millennium B.C., was one of fairly rapid social and political evolution."

[13]Most of the flow originated from monsoons in the Ethiopian highlands, and a smaller part came from the upper watershed of the White Nile around Lake Victoria. With impressive precision, the river began to rise in the South in early July and the flood got to the northern end of the valley by mid September. The Tigris and Euphrates were not only less predictable in timing, but also more irregular and less benign in volume.

system of canals redirected the remaining floodwater to basins in need.[14] Economic sociologists agree that in Egypt irrigation agriculture "*could generate crop-to-seed yield of between 12:1 and 24:1 ... but only at the cost of high capital investments*" (Morris and Manning 2005: 141). In Michael Mann's genealogy of social power, artificial irrigation involved one of the earliest forms of economic investment, which in Egypt was even more productive than in Mesopotamia. Both in Egypt and Mesopotamia, irrigation agriculture could "*generate a surplus far greater than that known to populations on rain-watered soil*" (1986: 80). In Egypt, "*the process was as in Mesopotamia, but squared,*" and "*as productivity grew, so too did civilization, stratification, and the state*" (1986: 108). In our model, a high value of the parameter ρ reflects the existence of an environment in which investments yield large increases of economic surplus in the same way that the construction of irrigation systems resulted in major expansions of the food surplus in Egypt.

(3) Territorial isolation as natural protection. The Nile basin is surrounded by deserts, which made invasions much less likely than in other centers of civilization. According to Bradford (2001: 9), "*The sea to the north and the deserts west and east isolated the Egyptians from the rest of mankind, except for merchants, some infiltrators, and the occasional raid.*". The desert provided two kinds of protection. On the one hand, the desert's inhospitality to human settlement discouraged the emergence of hostile neighbors nearby. On the other, the desert was a formidable barrier against distant rivals. In terms of our model, Egypt's territorial isolation is a naturally high κ_1.

How do these conditions account for Egypt's twin achievements of security and prosperity in the context of our model? A high level of v_1 has no effects in the model with exogenous defense capability. The implication is that the Nile's extraordinary natural fertility was not a decisive factor *per se*. Highly productive soils are not unique to Egypt. But the model does highlight that a combination of a high κ_1 and a high ρ helped Egypt attain prosperity and security. Protected by the deserts, Egypt had virtually no challengers it could not deter. This would permit Egypt to fend off well against nomadic raiders interested in capturing Egypt's agricultural assets and surplus. Given the potential for productivity-enhancing irrigation infrastructure, Egyptian rulers had incentives to encourage investments that would increase future surplus.

[14] According to a long scholarly tradition (Weber [1909] 2013, Wittfogel 1957), water management and state formation were closely linked in ancient societies. The thesis of "hydraulic empires," which claims that irrigation was a public good with enormous fixed costs, and that pristine states formed precisely in order to provide them, has been discredited by evidence showing that irrigation was not preceded by the emergence of state administrations.

The resulting picture is one where Egypt is located in a favorable section of region **R2**, if not directly in **R1**. The reason to place Egypt during the state formation period (end of the Naqada II period, around 3200BC) in a good part of **R2** is that Egypt did face occasional attacks, and perhaps the total absence of challenges that characterizes **R1** is better reserved to the heights of Egyptian power under the new kingdom, when the Egyptian state was even more dominant than during its formative phase. A "good" part of **R2** is one near the frontier with **R1**, where κ_1 is so high that the cost of defense effort done by the incumbent is small but sufficient to guarantee victory with very high probability. Thus, a society in such "good" part of **R2** would grow and enjoy a relatively secure existence, because the probability of defeat is small, and the returns to investment are high. These conditions –high κ_1, high ρ— continued to prevail in Egypt for a long time, perhaps explaining the remarkable stability of the Egyptian policy alluded to earlier.

4.2 The end of the Bronze Age

Compared with the rise of the great civilizations of the Bronze Age, their demise was a much more sudden affair: a wave of political and economic collapse swept across the Eastern Mediterranean around 1200BC, with drastic consequences in the Near East, Anatolia, Greece, and Egypt. An irreversible legacy of the collapse was the extinction of dozens of cities that were at the very frontier of political and technological development. Few events in history provide such a large scale and definite proof against unilinear visions of social progress.

For a period of almost 400 years, the Eastern Mediterranean had seen the rise of multiple states that improved their productive capacity and were capable of defending their wealth against "barbarian" populations. Progress on the productive dimension included advanced irrigation and plowing techniques for expanding agricultural surplus, storage facilities especially conditioned for the preservation of cereals, and permanent bureaucracies for economic redistribution. Military power was sustained by chariots and fortified walls. In combination, these achievements backed a sophisticated division of labor that allowed for the emergence of specialized ceramic and metal craft, and the development of writing, religion and the arts. This set of thriving states included the city-ports of the Levant, the kingdoms of Anatolia, the Egyptian empire, and the city-states of Mesopotamia and Cyprus.

"*But then, the world as they had known it for more than three centuries collapsed and essentially vanished*" as Eric Cline puts it (2014: 241). According to the assessment of

Drews (1993: 3), "altogether the end of the Bronze Age was arguably the worst disaster in ancient history, even more calamitous than the collapse of the western Roman Empire." The proximate cause of the end of the Bronze Age was, in most areas, invasion by armed groups coming from beyond civilization. The "Sea Peoples," as the Egyptians called them, were actually a diverse array of intruders with different geographic and ethnic origins (Sandars 1987), including the Deniens (either Greeks or the Dan tribe among Israelites), the Sherden (possibly Sardinians), the Shekelesh (Sicilians), the Lukka (from the Anatolian Aegean), and the Teresh (possible ancestors of the Etruscans).

There has been a long debate on the fundamental causes behind the end of the Bronze Age. Since the beginning of the debate in the mid-1960s, archaeologists have hypothesized that the collapse was set in motion by earthquakes (Schaeffer 1948), droughts and famines (Carpenter 1968), internal rebellions (Zuckerman 2007 and Carpenter 1968), or innovations in military technology (Drews 1993).

Our model can help think about the end of the Bronze Age in two ways, one particular and one general. The particular application, as we will see, is to show that most of the fundamental explanations given for the end of the Bronze Age can be analyzed, using our model, as shocks affecting either the value to the challenger of the incumbent's asset (the parameter h), or the effectiveness of the challenger's military capability (the parameter κ_c). These interpretations, in addition, are consistent with the wealth of archaeological findings collected since the mid-1980s, which has tended to reinforce the notion that military struggle was involved in the process at least as a proximate cause for a high proportion of cases.

The more general point resulting from our use of the model is to relate changes in deep military and economic fundamentals to the arguments made by social theorists that the evolution of political complexity is not unilinear, but plagued by dead ends and reversals. According to our model, a few deep economic and military fundamentals shaped the equilibrium in the confrontation between the civilized centers and the "barbarian" periphery. A shock to any one of those fundamentals could shift societies from one combination of levels of security and prosperity to another combination. Importantly, those movements do not necessarily go in the direction of greater security and prosperity. The end of the Bronze Age involved state de-consolidation and a regression to lower income levels—a Dark Age–, as in the region of conflict and stagnation, **R3**, in our model.

Debate among archeologists around the end of the Bronze Age has made empirical and theoretical progress. Only two hypothesized causes are incompatible with the notion that the end of the Bronze Age involved a major military defeat of the civilized world: earthquakes

and internal rebellions. Both explanations face challenges. The hypothesis of earthquakes has been discredited in the face of new archaeological evidence showing that most urban destruction was caused not by natural forces but by an enemy attack, which in the case of the key city of Ugarit left numerous arrow-heads throughout the ruins (Yon 1992: 117; Singer 1999: 730; Kanievski et al 2011). Attacks not only occurred but were endemic, even if un-coordinated: from opposite ends of the Bronze Age world, Hittite and Egyptian rulers left unequivocal testimonies about the menace and calamities of the incursions by the Sea Peoples, both in pictorial and written form.

The hypothesis of internal rebellions relies on the least plausible premise given the geographical scope and speed of the collapse: an extraordinary level of simultaneity among the rebels across the different sites of Anatolia, Greece, Northern Mesopotamia and the Levant. One potential driver of the simultaneity could be a common climatic shock.[15] But a more serious challenge to a pure internal rebellion story is posed by the fact that there is evidence of large migration movements across Late Bronze Age civilizations. This evidence is more compatible with invasions and exoduses than with simultaneous infighting.

The theoretical problem with the invasions is, of course, what caused them in the first place. Two hypotheses consistent with available evidence are:

(1) A severe change in climate, which caused draught and famines, and compelled the populations living beyond the gates of civilization to invade in search for food. Cities that were storehouses of grain fell victim to "*a final resort to violence by a drought sicken people*" (Carpenter 1968: 69).

(2) A revolution in the means of war, including the introduction of the javelin, which tipped the military balance in favor of nomadic intruders from economically less developed regions. Thus, according to Drews (1998: 33), "*the Catastrophe was the result of a new style of warfare that appeared toward the end of the thirteen century BC, [which] opened up new and frightening possibilities for various uncivilized populations that until that time had been no cause of concern to the cities and kingdoms of the eastern Mediterranean*". What were the changes introduced by the "uncivilized populations"? Chrissantos (2008: 11) summarizes them: "*these tribes developed better and lighter body armor, [...] lighter and smaller round shields, [and] revolutionary longer, stronger swords [...] They also invented*

[15] A recent paleobotany study based on samples of pollen throughout the Bronze and Iron Ages confirms the existence of a substantial climate change around the time of the collapse, which caused a reduction in precipitation and resulted in the shrinkage of the Mediterranean forest (Langgut, Finkelstein, and Litt 2013). In the interpretation of these authors, climate change and the ensuing famine may have caused internal rebellions rather than foreign invasions.

a new weapon, the javelin, which could be used as a missile to hurl at an enemy. They [managed to] overcome the civilizations' chariot advantage [...] Once these tribes mastered sea travel, no shore was too far for an attack. The failure of the chariot in the face of this new warfare marks the beginning of the Bronze Age world's collapse".

At the theoretical level, of course, climate change and military innovation are not mutually incompatible causes and can be combined under the form of a "Perfect Storm" (Cline 2014: Chapter 5). Another recent theoretical development for combining ineliminable causes builds on the idea of "System Collapse." The point of departure is the fact that Late Bronze Age societies—kingdoms, villages, cities and empires—were all connected through frequent commercial, production, and diplomatic relations. The assumption is that such relations had reached such a level of intensity, specialization and complementarity that if the economy in one of them were to come to a halt, for whatever reason, the whole Eastern Mediterranean would collapse under "domino" and "multiplier" effects. This allows for the theoretical possibility that weather- and technology-induced invasions had devastated a critical number of nodes in the workings of the global Eastern Mediterranean, which eventually provoked its general collapse.

Our incumbent-challenger model is compatible with all surviving interpretations for the collapse of the Bronze Age, taken individually or in any of their combinations (general wave of invasions, invasions in critical sites, invasions prompted by climate-induced famines, invasions caused by changes in the art of war). More importantly, however, the definition of the challenger's valuation of the incumbent's asset hv, and the effectiveness of the challenger's military effort κ_c, helps distinguish between two separate forces at play in the invasions that put an end to the civilized Eastern Mediterranean: the motivation behind invasion versus the effectiveness of the means to invade. The historical debate has sometimes conflated both issues and other times considered motivation and effectiveness as factors driving rival explanations. While changes in h and κ_c capture substantively different forces, as discussed in Section 3 they are mathematically equivalent in that both affect the aggressiveness of the challenger in the margin. Therefore, studying the comparative statics of κ_c can illuminate the role of both changes in motivation and aggressiveness of barbarians.

The parameter κ_c was assumed equal to 1 in the baseline model. We now consider a move to $\kappa_c > 1$, under the maintained assumption that $\kappa_1 \geq \kappa_c$. This will allow us to compare the extent to which the incumbent can attain security and prosperity in a world where the challenger is tougher. In other words, we study how shifts in κ_c affect the partition of the parameter space derived in Proposition 1. The following proposition shows such changes.

Proposition 3 *Optimal behavior by the incumbent yields a division of the parameter space* $(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+$ *for a given* $\kappa_1 > \kappa_c$ *into four distinct regions:*

Region 1 *(R1):* $\left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | \rho < \kappa_1/\kappa_c,\ \rho > \kappa_1/(\kappa_1 - \kappa_c)\right\}$ *Security and prosperity*

In **R1** *the solution is:* $\left\{a_1 = v_1 \frac{\kappa_1 \kappa_c (1+\rho)}{(\kappa_1 + \kappa_c \rho)}, i_1 = v_1 \frac{(\kappa_1 - \kappa_c)}{(\kappa_1 + \kappa_c \rho)}, V_1 = v_1 \frac{\kappa_1 (1+\rho)}{(\kappa_1 + \kappa_c \rho)}\right\}$

Region 2 *(R2):*$\left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | \rho > \kappa_1/\kappa_c,\ \rho > 4\kappa_c/\kappa_1 \text{ and } \rho > 1\right\}$ *Prosperity without security*

In **R2** *the solution is:* $\left\{a_1 = \frac{\kappa_1 v_1}{2}\left(1 + \frac{1}{\rho}\right), i_1 = \frac{v_1}{2}\left(1 - \frac{1}{\rho}\right), V_1 = \frac{v_1}{2}\left(1 + \frac{1}{\rho}\right)\sqrt{\rho \kappa_1}\right\}$

Region 3 *(R3):*$\left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | 2\kappa_c > \kappa_1 \text{ and } \rho < 4\kappa_c/\kappa_1\right\}$ *Neither prosperity nor security*

In **R3** *the solution is:* $\left\{a_1 = v_1\left(\frac{\kappa_1}{2}\right)^2, i_1 = 0, V_1 = v_1\left(1 + \frac{\kappa_1}{4\kappa_c}\right)\right\}$

Region 4 *(R4):*$\left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | \kappa_1 > 2\kappa_c,\ \rho < \kappa_1/(\kappa_1 - \kappa_c)\right\}$ *Security without prosperity*

In **R4** *the solution is:* $\left\{a_1 = \kappa_c v_1, i_1 = 0, V_1 = v_1\left(2 - \frac{\kappa_c}{\kappa_1}\right)\right\}$.

Proof: See appendix.

The comparison of the solutions in Propositions 1 (with $\kappa_c = 1$) and 3 (with $\kappa_c > 1$) is best appreciated in Figure 3 where the dashed lines show the partition of the parameter space with $\kappa_c = 1$ and the solid lines show the partition with $\kappa_c > 1$, reflecting a more aggressive challenger.

The comparative statics of κ_c are intuitive; a higher military capacity by the challenger worsens outcomes in the following general sense. For any point in the (κ, ρ) space where either security or prosperity (or both) were enjoyed, a higher κ_c implies that security, prosperity or both may be lost. In other words, a higher κ_c expands the footprint of all regions against $R1$ where both security and prosperity obtain. In addition, $R3$ which combined insecurity and stagnation, grows at the expense of all others. This means a world with a more aggressive challenger is tougher on the incumbent.

Specifically, the historical victory of the "Sea Peoples" over the cities and kingdoms of the Ancient Near East involved a shift from the security-prosperity quadrant of our map ($R1$, or good parts of $R2$) to the conflict-stagnation quadrant ($R3$), as a result of positive shocks to the value of the economic output to the challengers (an increase in h, in turn an effect of a climatic change) or to the technology of attack (an increase in κ_c).

We can use the model to delve deeper into the diverging fates of the different regions that suffered the attacks of the Sea Peoples at the end of the Bronze Age. The extremes of that contrast are Egypt, which managed to repel the invasion, and cities near the Mediterranean coast of the Levant, like Ugarit, for which invasion resulted in irreversible destruction.

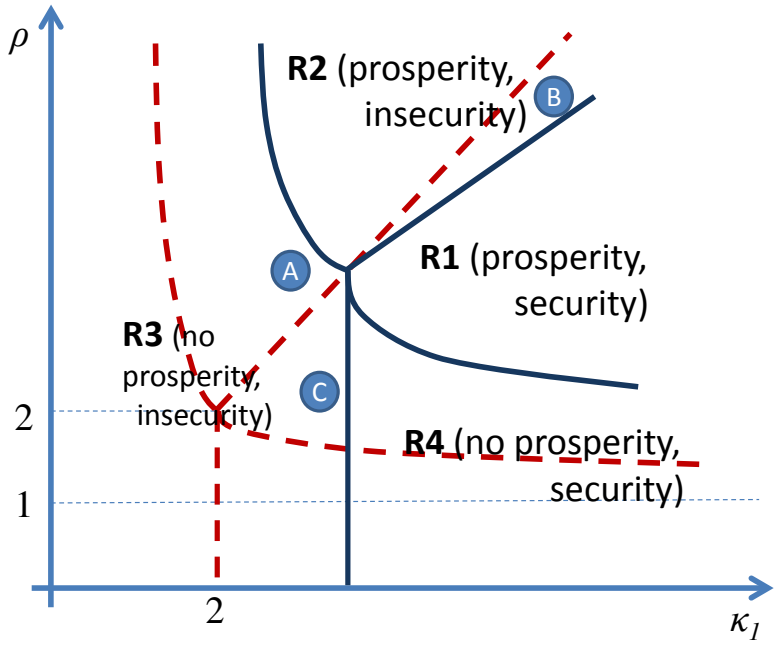


Figure 3: Equilibrium partition of the parameter space with more aggressive challenger

Ugarit is a model case of Bronze Age collapse because, in addition to its political complexity and economic prosperity before the invasion, the archeological excavation found clay tablets that survived the destruction to provide the most dramatic textual evidence on the threat of the Sea Peoples as well as on the efforts to prepare against them, which eventually proved futile. In the tablets, the king of Ugarit makes desperate requests to his Anatolian overlord, who was using Ugarit's maritime fleet to defend other sections of the Hittite empire. The tablets reveal that the attack was formidable and Ugarit was almost defenseless.[16] Hence, the eventual destruction. In terms of our model, Ugarit's vulnerability at the time of the invasions can be interpreted as either an initial location within $R1$ that was close to the vertex, and thereby not too far away from $R3$, or within a narrow strip along the $R2/R3$ border on the side of $R2$. When the shocks that prompted the invasion occurred—which,

[16] Ammurapi, the King of Ugarit, makes a desperate plea to the Hittite overloard, whom he addresses as his "father:" "My father, behold, the enemy's ships came (here); my cities(?) were burned, and they did evil things in my country. Does not my father know that all my troops and chariots(?) are in the Land of Hatti, and all my ships are in the Land of Lukka?...Thus, the country is abandoned to itself. May my father know it: the seven ships of the enemy that came here inflicted much damage upon us". Letter RS 18.147 in Jean Nougaryol et al. 1968. Ugaritica V, 24: 87–90.

as already seen, involved a redrawing of the borders of the regions—the effect was to push Ugarit deep into $R3$. In the new location, Ugarit faced the prospects of attacks that were too strong for the city to resist. In the model, a deep position in $R3$ (closer to the ρ axis) entails a lower probability that the incumbent will prevail. In fact, for Ugarit, foreign attack resulted in extinction.

Egypt was also a victim of barbarian attacks, and repeatedly so, but the outcome was very different as Egypt managed to resist and survive. Since the end of the Second Intermediate Period, Egypt had developed a military force of genuinely imperial strength, including a highly professional army and a formidable fleet of ships with the most advanced equipment at the time. Before the shock, Egypt was located deep in $R1$ so that the worst effect of the shock could have been to relocate Egypt in a relatively safe neighborhood of $R2$. Egypt became susceptible to challenges, but it could prevail in the battlefield with high probability. Pictorial inscriptions on the walls of the Karnak Temple attest to the threats posed by the Sea Peoples at roughly the same time they invaded the Levant. But, in contrast to Ugarit's tablets, the Egyptian inscriptions actually honor king Merneptah's success in subduing the invaders. The contrasting cases of Ugarit and Egypt correspond to the points A and B in Figure 3.

5 Endogenous defense capability and the transition to security and prosperity

5.1 Setup

We will now allow the incumbent to spend resources in one period to increase its defense effectiveness in the next period. Let us now introduce a period zero, before the periods 1 and 2 that we have analyzed so far. Since the challenger will never fight in period 2, the incumbent will never spend in expanding defense capability in period 1. Thus, the decision to augment defense capability will be relevant only in period 0. We postulate that in period 0 the incumbent has a defense capability κ_0, and can spend an amount m_0 that will take defense capability in the next period to $\kappa_1 = \kappa_0 + \gamma m_0$, where γ captures the purchasing power of income in terms of defense means. We assume $\gamma \in (0, 16/(4+\kappa_0))$ where the upper bound is a technical assumption to guarantee the possibility of partial regime transitions. To make things interesting, we assume κ_0, ρ are such that if things were left unchanged, in period 1 the incumbent would find himself in region **R3**, which means he can expect disorder

and stagnation. In particular, we impose the following,

Assumption 1 *$\rho\kappa_0 < 4$ and $\kappa_0 < 2$.*

All other aspects of the interaction between challenger and incumbent remain as before, and for simplicity we return to the case where $\kappa_c = 1$.

Timing

In period 0, the incumbent starts by selecting m_0. Then, in each period $t = 0, 1, 2$ the incumbent selects a_t and i_t.[17] After observing (a_t, i_t) the challenger selects b_t. If $b_t = 0$, the players retain their positions in the next period. If $b_t > 0$, then there is a war at the end of period t. The winner of the war becomes the incumbent in the next period, and faces a new challenger then.

Payoffs

The fact that there is a new type of expenditure changes the incumbent's budget constraint to $v_0 - m_0 - \frac{a_0}{\kappa_0} - i_0 \geq 0$. And the fact that there is an extra period now implies that a zero arming decision by the challenger in period 1 could open the challenger to vulnerability. So in this three-period model an additional assumption is that the challenger cannot be eliminated.[18] This matches the historical cases of settlers dealing with nomadic raiders, who have vast steppes on which to run away from the forces of the civilized, settled, center.

As before, we solve the model through backward induction. The solution for periods 1 and 2 is given by our analysis in the previous section. That analysis tells us the expected payoff for being an incumbent in period 1 is given by,

$$V_1(i_0, m_0) = (v_0 + \rho i_0) \times \begin{cases} \frac{(\kappa_0+\gamma m_0)(1+\rho)}{\kappa_0+\gamma m_0+\rho} & (\kappa_0 + \gamma m_0,\, \rho) \in \mathbf{R1} \\ \sqrt{\frac{(\kappa_0+\gamma m_0)}{\rho}}\frac{(1+\rho)}{2} & (\kappa_0 + \gamma m_0,\, \rho) \in \mathbf{R2} \\ \left(1 + \frac{\kappa_0+\gamma m_0}{4}\right) & (\kappa_0 + \gamma m_0,\, \rho) \in \mathbf{R3} \\ \left(2 - \frac{1}{\kappa_0+\gamma m_0}\right) & (\kappa_0 + \gamma m_0,\, \rho) \in \mathbf{R4} \end{cases} \equiv (v_0 + \rho i_0) S(m_0)$$

[17] The assumption that m_0 is decided before a_0 and i_0 is just to simplify the exposition. It is equivalent to assume that the incumbent selects all three variables simultaneously since the challenger does not move until the incumbent has selected all of his actions. What is of course important is that the incumbent makes his choices before the challenger.

[18] An alternative assumption with the same effect is that if eliminated, the challenger is replaced with another, identical, one. If the challenger can be definitively eliminated when selecting zero arming, then it would not be an equilibrium for the challenger to desist from arming itself and deterrence would be impossible.

Given this continuation value, we can solve for decisions in period 0. After the incumbent has selected m_0, a_0 and i_0, the challenger decides whether to arm himself. Using the same logic as in the previous section, we see that the challenger's best response function is given by,

$$b_0(a_0, m_0, i_0) = \begin{cases} \sqrt{a_0 V_1(i_0, m_0)} - a_0 & if\ a_0 < V_1(i_0, m_0) \\ 0 & if\ a_0 \geq V_1(i_0, m_0) \end{cases}$$

This notation embeds the four regions over which $V_1(i_0, m_0)$ is defined into the calculus of the challenger. Given this best response function, the incumbent has to choose a_0, i_0 after it chose m_0 such that she maximizes her expected utility.

The incumbent maximizes,

$$\max_{a_0, i_0 \geq 0} v_0 - m_0 - \frac{a_0}{\kappa_0} - i_0 + \frac{a_0}{a_0 + b_0(a_0, i_0, m_0)} V_1(i_0, m_0)$$

subject to

$$\begin{aligned} v_0 - m_0 - \frac{a_0}{\kappa_0} - i_0 &\geq 0\ (BC) \\ (v_0 + \rho i_0) S(m_0) - a_0 &\geq 0\ (DC) \\ a_0 &\geq 0 \\ i_0 &\geq 0 \end{aligned}$$

Notice this problem in period 1 is similar to the one with two periods in the previous section, except now the continuation value depends explicitly on m_0 (which is fixed at this stage, given the convention that it was selected before a_0 and i_0) through $S(m_0)$. The objective function is differentiable in a_0 and i_0. As before, the first order and complementary slackness conditions that characterize the optimum are given by

$$a_0 : \frac{1}{2} \sqrt{\frac{(v_0 + \rho i_0)\, S(m_0)}{a_0}} - \frac{1}{\kappa_0} - \frac{\lambda_{BC}}{\kappa_0} - \lambda_{ND} + \lambda_a = 0 \tag{14}$$

$$i_0 : \frac{\rho}{2} \sqrt{\frac{a_0}{(v_0 + \rho i_0)\, S(m_0)}} - 1 - \lambda_{BC} + \lambda_{ND} \rho S(m_0) + \lambda_i = 0 \tag{15}$$

$$\lambda_{BC}(v_0 - m_0 - \frac{a_0}{\kappa_0} - i_0) = 0,\ \lambda_{ND}\left((v_0 + \rho i_0) S(m_0) - a_0\right) = 0,\ \lambda_a a_0 = 0,\ \lambda_i i_0 = 0 \tag{16}$$

As before, $\lambda_{BC}, \lambda_{ND}$ are the Lagrange multipliers for the budget constraint and deterrence constraints, and λ_a, λ_i are the multipliers for the non-negativity constraints for the control variables. Again the infinite marginal utility of a_0 at zero implies $a_0 > 0$ and $\lambda_a = 0$, so there are in principle eight possible cases depending on whether the remaining three Lagrange multipliers are positive or zero. The following Lemma shows that, given our Assumption 1 there are only two feasible cases in period 0.

Lemma 1 *If Assumption 1 holds, then in period 0 the incumbent chooses:*
i) $i_0 = 0$ *and* $a_0 = \frac{\kappa_0^2}{4} v_0 S(m_0)$ *when* $\frac{v_0 S(m_0)}{(v_0 - m_0)} < \frac{4}{\kappa_0}$*; or*
ii) $i_0 = 0$ *and* $a_0 = \kappa_0(v_0 - m_0)$ *when* $\frac{v_0 S(m_0)}{(v_0 - m_0)} \geq \frac{4}{\kappa_0}$.

Proof: See Appendix.

Lemma 1 reveals that no productive investment occurs in period 0 and that the arming effort depends on the value of m_0 through its impact on the continuation payoff. With this result, we are now equipped to study the incentives of the incumbent to make changes in defense capability. We can trace how those changes will affect period 0 army decisions for both incumbent and challenger, victory probabilities, and subsequent future investment, army sizes, and security and prosperity outcomes.

5.2 Solution

Lemma 1 enables us to compute the incumbent's present expected utility for any value of m_0. The effect of m_0 on the incumbent's utility depends on the initial conditions in period 0. If the maximum utility is attained for extremely low m_0, then the incumbent will remain stuck with insecurity and no prosperity ($R3$) in period 1. On the contrary, if the optimal m_0 is sufficiently high, security and prosperity will obtain in period 1. Notice, however, that the path to prosperity (via investments in m_0) depends on ρ–the return on productive investment. In particular, when $\rho < 2$, the path to security and prosperity requires going from $R3$ to $R1$ through $R4$, and when $\rho \geq 2$, it requires going from $R3$ to $R1$ through $R2$ (see Figure 1). We analyze these cases in the following,

Proposition 4 *(a)Under Assumption 1 and provided that* $\rho < 2$*, there exist cutoffs* τ_L, τ_M *and* τ_H*,* $\tau_L < \tau_M < \tau_H$ *such that*

1. *If* $\gamma v_0 < \tau_L$*, the polity stays in* **R3** *(stagnation without security);*
2. *If* $\tau_H < \gamma v_0$*, the polity moves to* **R1** *(attains security and prosperity); and*
3. *If* $\tau_M < \gamma v_0 < \tau_H$*, the polity moves to* **R4** *(attains security without prosperity)*

(b) Under Assumption 1 and provided that $\rho \geq 2$, there exist cutoffs $\sigma_L, \sigma_{M1}, \sigma_{M2}$ and σ_H, $\sigma_L < \sigma_{M1} < \sigma_H$ and $\sigma_L < \sigma_{M2} < \sigma_H$ such that

1. *If $\gamma v_0 < \sigma_L$, the polity stays in* **R3** *(stagnation and conflict);*
2. *If $\sigma_H < \gamma v_0$, the polity moves to* **R1** *(attains security and prosperity); and*
3. *If $\sigma_{M1} < \gamma v_0 < \sigma_{M2}$, the polity moves to* **R2** *(attains prosperity without security)*

Proof: See Appendix.

The expression γv_0 captures the extent to which initial income can be used to improve defense capability. This proposition tells us that, given the initial military capacity κ_0 and the productivity of investment ρ, the transitions followed by the polity will be very different depending on the initial income γv_0 in terms of defense capability purchasing power. If γv_0 is very low, the polity will remain trapped without security or prosperity. If γv_0 lies in an intermediate region, the polity can move into a region of partial achievement. If $\rho < 2$, the transition is to $R4$ where it will attain peace but will not grow. The reason is that even though it attains a higher defense capability κ_1 in the next period, which gives the incumbent the ability to fend off attacks at a lower cost, the benefit from consumption will still be higher than the present value from investing. If $\rho > 2$, the "hybrid" transition is to $R2$ where it will grow without attaining full security (this transition requires some restrictions on (κ_0, ρ)). If γv_0 is very high, however, the subsequent military capacity κ_1 will allow the incumbent to free resources for both a deterrent army and a large-scale investment at $R1$.

To summarize, while large enough initial income (or cheap enough defense capability) guarantees security and prosperity through sufficient accumulation of defense capability, intermediate levels may only allow to attain either security *or* prosperity. The Hobbesian argument that security is a precondition for prosperity is qualified in this model. Baseline prosperity can buy defense capability, and only then can the ensuing security promote more prosperity.

6 Historical Illustration: Sumeria and the origin of civilization

The Fertile Crescent in Southwest Asia was the source of the first substantial economic surplus in human evolution, in turn the result of a major innovation: domestication of plants and animals for food production. The Fertile Crescent was a political pioneer as well. The

rudiments of large-scale political organization emerged in Southern Mesopotamia to form the pristine city-states of Sumer and ultimately shape the first major civilization.

Like in Egypt, in Mesopotamia it was a fertile riverine valley, exceptionally endowed for alluvial agriculture, that was the key for economic prosperity. The twin rivers Tigris and Euphrates flooded the land and replenished nutrients by spreading silt. Also like in Egypt, the natural advantages required systematic human effort to produce economic results. Southern Mesopotamians made massive investments in the creation of the proper irrigation infrastructure. The investments were made because of extraordinary returns. According to Mann (1986: 78), "*If [the alluvium] can be diverted onto a broad area of existing land, then much higher crop yields can be expected. This is the significance of irrigation in the ancient world: the spreading of water and silt over the land. Rain-watered soils gave lower yields*". Liverani (2005, p. 5) offers an idea of the increase in yields that could be obtained through judicious investments: "*The agricultural production of barley underwent a notable, possibly tenfold, increase thanks to the construction of water reservoirs and irrigation canals, of long fields adjacent to the canals watered by them, and thanks to the use of the plow, of animal power, of carts, of threshing sledges, of clay sickles, and of improved storage facilities*".

These high returns to productive investment help place Sumeria in the parameter space of our model as a case of high ρ. But what about the other parameter, the effectiveness of defense effort κ_1?

In contrast to Egypt, geography did not afford the Sumerians natural protection against attacks from outsiders. On the contrary, the natural landscape exposed Sumeria to numerous threats. As Bradford (1993: 4) puts it, "*Their neighbors to the west, the Amorites, nomads of the desert, infiltrated Mesopotamia... The neighbors to the east, who dwelled in the mountains, were the Gutians and the Elamites. The Gutians and, to a lesser extent, the Elamites considered Sumer and Akkad a treasurehouse to be raided*". Finer (1997: 102) also emphasized the porousness of the Sumerian frontiers.

In terms of our model, the vulnerability of Sumeria to invaders means that the effectiveness of defensive effort was low (low κ_1). Given a low κ_1, Sumeria's trajectory must have begun in the conflict-stagnation region, **R3**. But if output was so insecure, how could the first human civilization emerge at all? That is, how did Sumerians solve the problem of protecting surpluses from nomadic raiders and encouraging investments in productive infrastructure?

The extended model featuring endogenous defense capability provides an answer. Our proposition 4 states that a polity that is initially in **R3** due to a low military capability κ_0,

may invest in defense capability in order to attain sufficient security against the challenger. The key condition for this investment to be undertaken is for the polity to have enough initial income v_0. The archaeological record suggests Sumeria was well placed to meet that requirement. The availability of alluvial agriculture combined with an unparalleled initial endowment of plant and animal domesticates furnished the entire Fertile Crescent with exceptional advantages in food production. It is well known that due to altitude and climatic variation, the Fertile Crescent hosted a wide variety of plants with high potential for food production. The region had a wild flora with high yields of edible content, a high proportion of hermaphroditic plants that were more amenable to experimentation and selection based on yield, and a number of crops with high protein content. Diamond (1997: Ch. 8) highlights that all eight founder crops in the Neolithic were present in the area (the wild ancestors of einkorn, emmer wheat, flax, lentil, chick pea, pea, bitter vetch and barley). In addition, out of the five most important domesticated animals, four were available in the Fertile Crescent, namely pigs, cows, sheep and goats.[19] Given the natural advantages, Diamond (1997: 135) claims that "*any attempt to understand the origins of the modern world must come to grips with the question why the Fertile Crescent's domesticate plants and animals gave it such a potent head start*". In terms of our model, this combination of initial advantages can be captured by a high v_0.

What is delicate about the role of the Fertile Crescent's initial advantages is that a high v_0 can encourage predation by outside challengers. However, a high v_0 could also help finance the investments in defense that were needed to escape the conflict-stagnation trap characterizing **R3**. It is by no means obvious that the "defense-financing" force should dominate the predation force. This is a key tension investigated by our model, and the model makes an unambiguous prediction: for v_0 low enough, no escape from **R3** is possible. For v_0 high enough, the incentive to finance defensive capabilities dominates. Under such conditions, the investments in defense are made, and they bring enough security so as to incentivize productive investments and economic progress. Figure 4 shows the comparison of the cases of Egypt and Sumeria in terms of our parameter space (ρ, κ_1): Egypt started in a good region with relatively high levels of both parameters. Sumeria started with a low

[19] According to Trigger (2003: 281), the productive advantage of domesticated animals was important, and may help explain why Sumeria and Egypt were the first areas in the world to develop civilization: "*Egypt and Mesopotamia were the only early civilizations in our sample that supplemented human agricultural labour with that of domestic animals. Oxen[and donkey-drawn ploughs were present form an early period. Draft animals are estimated to have resulted in a 50 percent reduction in the human labour needed to grow grain, and this permitted small groups of men to work large, monocropped fields.*"

level of κ_0 and it was through investments that it raised its defense capability to a higher κ_1 that could deliver sufficient security.

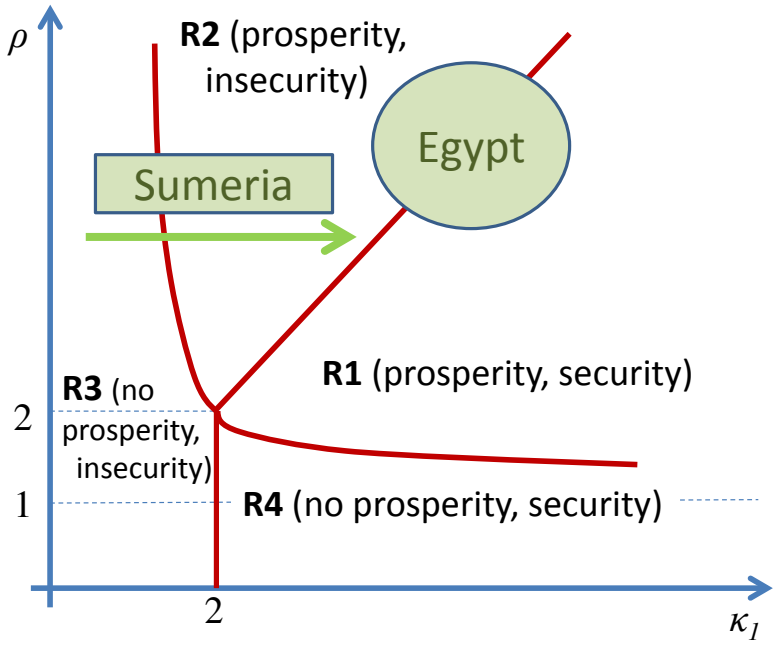


Figure 4: The different parameters of Egypt and Sumeria

What is the evidence of defense capabilities that grew endogenously in Sumeria? The archaeological record offers evidence of large and generalized investments to improve defense in the form of protective perimeter walls, which made Sumerian cities large-scale fortifications. Figure 5 includes illustrations of a number of Sumerian cities. All of them had walls. In fact, virtually every city in ancient history had walls. Walls were the endogenous, artificial substitute for the missing natural protection that was present in Egypt where cities did not have walls.

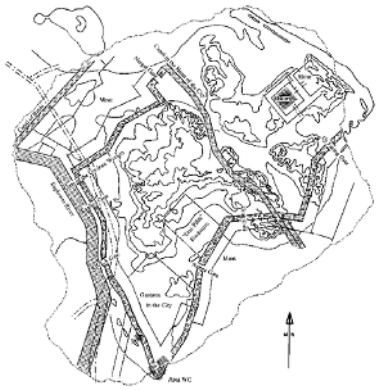
According to van de Mieroop's (1997) study of Mesopotamian cities, "*The inner cities were also clearly distinguished by their defensive walls. Perhaps the presence of walls was the main characteristic of a city in the eyes of an ancient Mesopotamian: all representations of cities prominently display walls, many kings boast of their building or repairing city walls, and even literary works sing their praise. A city without a wall might thus not have been conceivable.*"

The archaeological record substantiates not only the generalized presence of military

Ur

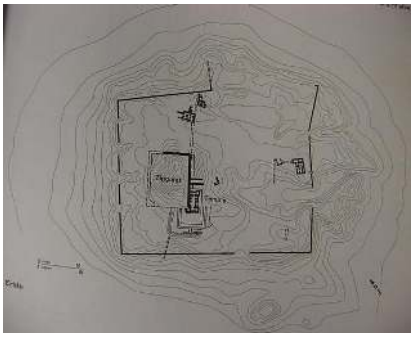
The new general plan of Ur (drawn up by F. Ghio): 1: City Wall. Source: Di Giacomo and Scardozzi (2012).

Nippur



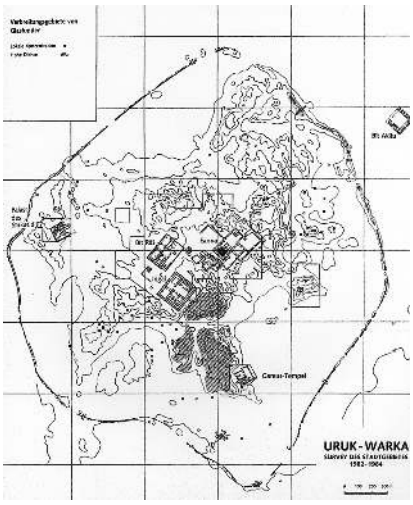
Source: Gibson, McGuire (1993).

Eridu



Source: Hawkes, J. (1974, p. 171).

Uruk



Source: J. Jordan (1931).

Figure 5: Plans of Sumerian cities

investments in rising city-states, but also their costliness, which would have been prohibitive to societies with low initial productive capacity. Both walls and the often complementary moats have been estimated to have required large investments of labor. The cost estimate for the moat in the Babylonian city of Dur-Jakin is ten thousand men working for three and a half months (Van de Mieroop, p. 76).

7 Conclusion

We present a model to investigate the dynamics of productive and defense capabilities in a society where an incumbent seeks to consolidate power and grow the economy. The components of the model are chosen by reference to the anthropological literature in order to capture both relevant environmental parameters as well as the minimalistic strategic environment facing proto-states at the point in in time where civilizations first arose. The central tension facing societies attempting the civilizational transition is captured in the notion of the civilizational paradox. A more prosperous society was worthy of attack, and the resulting insecurity would weaken incentives to create prosperity in the first place. In addition, states and civilizations arose together, and therefore stateness, defined as a high degree of security, had to emerge in association with the prosperity that tended to undermine it.

We construct a model to analyze the interplay between productive and defense capabilities to account for the rise of civilizations that combine prosperity and stateness. The model also helps evaluate claims about the relative role of security and prosperity that are central to classic theories of state formation. We show that all four combinations for the presence or absence of security and prosperity are possible, preventing simple characterizations of security or prosperity being necessary or sufficient conditions for one another. These regimes match various historical experiences of societies that attained neither, one, or both objectives. In that basic model, higher initial income does not affect the resulting regime the society is in, but tends to exacerbate conflict. In addition, the basic model helps consider how exogenous shocks to productive and defense capabilities might affect the ability of a society to attain security and prosperity. The key implication is that a balance between productive and defense capabilities is important in order to prevent a security breakdown. History offers examples of how naturally occurring high levels of productivity and defense, as in Egypt, enabled the emergence of civilization. But it also offers examples of negative shocks that destroyed civilizations, as in the end of the Bronze Age.

What the basic model cannot explain however is why Sumeria, together with Egypt the other early center of civilization, could develop when it did not enjoy a high level of natural defense. An answer is offered by the extension of the model to consider endogenous defense capability. A naturally occurring high initial productivity, which exacerbated conflict in the basic model, can now enable the transition into security with prosperity and resolve the civilizational paradox. The key is that initial productivity be high in terms of its purchasing power over improvements in defense capability. The possibility of accumulating means of defense helps create the conditions where productive investments can be made without triggering predatory challenges. This result may also help rationalize historical experiences where a temporary economic boom allows the state to consolidate its power and usher in a phase of more sustained growth. Isolating formally the pivotal role of defense capability to the civilizational process contributes to the demanding enterprise of discerning how economic shocks can hinder or help state formation and political stability more generally.

8 Appendix

Proof of Proposition 1: This is a particular case of the model with $\kappa_c \neq 1$, which is studied in Proposition 3.

Proof of Proposition 2: That a_1 increases in v_1 in all regimes follows directly from inspection of the solution for a_1 in each regime in Proposition 1. To see that whenever positive b_1 also increases in v_1, take the value of b_1 from the best response expression (4), and substitute in the values of i_1, a_1 in regions **R2** and **R3**. This yields respectively,

$$\begin{aligned} b_{1,\mathbf{R2}} &= v_1 \left\{ \sqrt{\frac{\kappa_1}{2}\left(1+\frac{1}{\rho}\right)\left(1+\rho\frac{\kappa_1-1}{\kappa_1+\rho}\right)} - \frac{\kappa_1}{2}\left(1+\frac{1}{\rho}\right) \right\} \\ b_{1,\mathbf{R3}} &= v_1 \left(\frac{\kappa_1}{2}\right)\left\{1-\frac{\kappa_1}{2}\right\} \end{aligned}$$

which are both positive and increasing in v_1.■

Proof of Proposition 3: The problem is to maximize,

$$\begin{aligned} = \; & v_1 - \frac{a_1}{\kappa_1} - i_1 + \frac{a_1}{a_1+b_1}(v_1+\rho i_1) \\ & +\lambda_{BC}(v_1 - \frac{a_1}{\kappa_1} - i_1) + \lambda_{DC}(\kappa_c v_1 - a_1 + \kappa_c \rho i_1) + \lambda_a a_1 + \lambda_i i_1. \end{aligned} \tag{17}$$

We will characterize the solution $(a_1, i_1, \lambda_{BC}, \lambda_{DC}, \lambda_a, \lambda_i)$ to this problem for each parameter combination $(\rho, \kappa_1, \kappa_c, v_1)$ given $\kappa_1 > \kappa_c$.

The first order and complementary slackness conditions that characterize the optimum are given by,

$$\frac{\partial}{\partial a_1} = \frac{1}{2\sqrt{\kappa_c}}\sqrt{\frac{v_1+\rho i_1}{a_1}} - \frac{1}{\kappa_1} - \frac{\lambda_{BC}}{\kappa_1} - \lambda_{DC} + \lambda_a = 0; a_1 \geq 0, \lambda_a \geq 0, \lambda_a a_1 = 0 \text{ c.s.} \quad (18)$$

$$\frac{\partial}{\partial i_1} = \frac{\rho}{2\sqrt{\kappa_c}}\sqrt{\frac{a_1}{v_1+\rho i_1}} - 1 - \lambda_{BC} + \lambda_{DC}\kappa_c\rho + \lambda_i = 0; i_1 \geq 0, \lambda_i \geq 0, \lambda_i i_1 = 0 \text{ c.s.} \quad (19)$$

$$\lambda_{BC}(v_1 - \frac{a_1}{\kappa_1} - i_1) = 0 \text{ c.s.}, \ \lambda_{DC}\left(\kappa_c v_1 - a_1 + \kappa_c\rho i_1\right) = 0 \text{ c.s.} \quad (20)$$

Given that $\lambda_a = 0$, we have eight possible cases given by whether the three remaining Lagrange multipliers $\lambda_{BC}, \lambda_{DC}$, and λ_i are zero or positive. We analyze each one of them. The general approach will be to assume in each case that the conditions defining it hold, and then determine which part if any of the parameter space $\{(\kappa_1, \kappa_c, \rho, v_1)|\kappa_1, \kappa_c, \rho, v_1 \geq 0\}$ can support a solution given the case's conditions. When the case implies conditions for the parameters that are mutually exclusive, the case will be deemed infeasible. When the case implies that the solution can be supported for combinations of the parameter values with measure zero we consider the case to be non-generic and also drop it from further consideration. The four regions detailed in the proposition hold for parametric areas with positive measure.

1. Case $\lambda_{BC} > 0$ (BC binds), $\lambda_{DC} > 0$ (DC binds, consolidation), and $\lambda_i = 0$ ($i_1 > 0$)

Since a_1 is always positive, and in this case i_1 is also positive, the FOCs in (18), (12) must hold with equality. Because this case involves binding BC and DC constraints, they also hold with equality. This is,

$$\begin{aligned}
a_1 &: \frac{1}{2\sqrt{\kappa_c}}\sqrt{\frac{1}{\kappa_c}} - \frac{1}{\kappa_1} - \frac{\lambda_{BC}}{\kappa_1} - \lambda_{DC} = 0 \\
i_1 &: \frac{\rho}{2\sqrt{\kappa_c}}\sqrt{\kappa_c} - 1 - \lambda_{BC} + \lambda_{DC}\rho\kappa_c = 0 \\
DC &: \kappa_c v_1 - a_1 + \kappa_c\rho i_1 = 0 \\
BC &: v_1 - \frac{a_1}{\kappa_1} - i_1 = 0,
\end{aligned}$$

implying that investment and army are,

$$i_1 = v_1 \frac{(\kappa_1 - \kappa_c)}{(\kappa_1 + \kappa_c \rho)} \tag{21}$$

$$a_1 = v_1 \frac{\kappa_1 \kappa_c (1+\rho)}{(\kappa_1 + \kappa_c \rho)} \tag{22}$$

As a result $\lambda_i = 0$ (or $i_1 > 0$) is supported by $\kappa_1 > \kappa_c$. After some algebra (using the FOCs) we find that $\lambda_{DC} > 0 \Leftrightarrow \kappa_1 > \rho\kappa_c$ and $\lambda_{BC} > 0 \Leftrightarrow \rho > \kappa_1/(\kappa_1 - \kappa_c)$. Therefore the parameter set supporting this solution to the incumbent's problem in period 1 is given by,

$$\mathbf{R1} = \left\{ (\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | \rho < \kappa_1/\kappa_c,\ \rho > \kappa_1/(\kappa_1 - \kappa_c),\ \kappa_1 > \kappa_c \right\},$$

and in this area there is investment and deterrence. The expected utility in period 1 is computed by substituting the solutions into the Lagrangian. In this first case expected utility is,

$$V_1 = v_1 \frac{\kappa_1 (1+\rho)}{(\kappa_1 + \kappa_c \rho)}.$$

2. Case $\lambda_{BC} > 0$ (BC binds), $\lambda_{DC} = 0$ (DC does not bind, conflict),and $\lambda_i = 0$ $(i > 0)$

Again we check the first order and complementary slackness conditions to see for which parameter set this case contains the solution. The relevant conditions are,

$$\begin{aligned}
a_1 &: \frac{1}{2\sqrt{\kappa_c}} \sqrt{\frac{v_1 + \rho i_1}{a_1}} - \frac{1}{\kappa_1} - \frac{\lambda_{BC}}{\kappa_1} = 0 \\
i_1 &: \frac{\rho}{2\sqrt{\kappa_c}} \sqrt{\frac{a_1}{v_1 + \rho i_1}} - 1 - \lambda_{BC} = 0 \\
BC &: v_1 - \frac{a_1}{\kappa_1} - i_1 = 0
\end{aligned}$$

Investment and army solutions are respectively given by,

$$\begin{aligned}
i_1 &= \frac{v_1}{2}\left(1 - \frac{1}{\rho}\right) \\
a_1 &= \frac{\kappa_1 v_1}{2}\left(1 + \frac{1}{\rho}\right).
\end{aligned}$$

This solution is respectively consistent with $\lambda_{DC} = 0$ (DC holds with strict inequality) and $\lambda_i = 0 \Leftrightarrow \rho > \frac{\kappa_1}{\kappa_c}$ and $\rho > 1$. The solution is consistent with $\lambda_{BC} > 0 \Leftrightarrow \rho \geq \frac{4\kappa_c}{\kappa_1}$ (this

comes from checking the conditions such that $\lambda_{BC} > 0$ in the two FOCs). As a result, the parameter set supporting this second case is,

$$\mathbf{R2} = \left\{ (\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | \rho > \kappa_1/\kappa_c,\ \rho > 4\kappa_c/\kappa_1 \right\}.$$

Expected utility for the incumbent in this case is,

$$V_1 = \frac{v_1}{2}\left(1 + \frac{1}{\rho}\right)\sqrt{\rho\kappa_1}$$

3. Case $\lambda_{BC} = 0$ (BC does not bind), $\lambda_{DC} = 0$ (DC does not bind, conflict), and $\lambda_i = 0$ ($i_1 > 0$)

Non-generic, since it is consistent for subset of the space (ρ, κ_1, v_1) that has measure zero. This follows from (18) and (12), so when λ_{BC}, $\lambda_{DC}, \lambda_i = 0$

$$\begin{aligned} a_1 &: \frac{1}{2\sqrt{\kappa_c}}\sqrt{\frac{v_1 + \rho i_1}{a_1}} - \frac{1}{\kappa_1} = 0 \\ i_1 &: \frac{\rho}{2\sqrt{\kappa_c}}\sqrt{\frac{a_1}{v_1 + \rho i_1}} - 1 = 0. \end{aligned}$$

The first FOC implies $\sqrt{\frac{v_1+\rho i_1}{a_1}} = \frac{2\sqrt{\kappa_c}}{\kappa_1}$ and substituting into the second FOC, we get $\frac{\rho}{2\sqrt{\kappa_c}}\frac{\kappa_1}{2\sqrt{\kappa_c}} = 1$ or $\frac{\rho\kappa_1}{4\kappa_c} = 1$, which implies this holds for a non-generic parameter set.

4. Case $\lambda_{BC} = 0$ (BC does not bind), $\lambda_{DC} > 0$ (DC binds, consolidation), and $\lambda_i = 0$ ($i_1 > 0$)

The FOCs are,

$$\begin{aligned} a_1 &: \frac{1}{2\sqrt{\kappa_c}}\sqrt{\frac{v_1 + \rho i_1}{a_1}} - \frac{1}{\kappa_1} - \lambda_{DC} = 0 \\ i_1 &: \frac{\rho}{2\sqrt{\kappa_c}}\sqrt{\frac{a_1}{v_1 + \rho i_1}} - 1 + \kappa_c\rho\lambda_{DC} = 0, \end{aligned}$$

where $a_1 = \kappa_c(v_1 + \rho i_1)$ indicating that λ_{DC} must simultaneously equal $\frac{1}{2\kappa_c} - \frac{1}{\kappa_1}$ and $\frac{\left(\frac{\rho}{2}-1\right)}{\kappa_c\rho}$, which forces the equality $\rho = \frac{\kappa_1}{\kappa_c}$, which is non-generic.

5. Case $\lambda_{BC} > 0$ (BC binds), $\lambda_{DC} > 0$ (DC binds, deterrence), and $\lambda_i > 0$ ($i_1 = 0$)

Because $i_1 = 0$, BC binding implies that $a_1 = \kappa_1 v_1$, but DC binding implies that $a_1 = \kappa_c v_1$, so $\kappa_1 = \kappa_c$ which is non-generic given the assumption $\kappa_1 \geq \kappa_c$.

6. Case $\lambda_{BC} > 0$ (BC binds), $\lambda_{DC} = 0$ (DC does not bind, conflict), and $\lambda_i > 0$ ($i_1 = 0$)

The BC binding and $i_1 = 0$ yield $a_1 = v_1\kappa_1$. The DC not binding implies $v_1\kappa_c - v_1\kappa_1 > 0 \Leftrightarrow \kappa_c > \kappa_1$ which violates the assumption $\kappa_1 > \kappa_c$, making this case infeasible.

7. Case $\lambda_{BC} = 0$ (BC does not bind), $\lambda_{DC} = 0$ (DC does not bind, conflict), and $\lambda_i > 0$ ($i_1 = 0$)

In this case $a_1 = v_1\kappa_1^2/(4\kappa_c)$ and $i_1 = 0$. This solution is consistent with $\lambda_{BC} = 0$ and $\lambda_{DC} = 0 \Leftrightarrow \kappa_1 < 2\kappa_c$. Also for $\lambda_i > 0$ we need $1 - \rho\kappa_1/(4\kappa_c) > 0$ (from the FOC of i_1). Thus, this holds for any triple $(\rho, \kappa_1, v_1) \in \mathbb{R}^3_+$ such that $\kappa_1 < 2\kappa_c$ and $\rho < 4\kappa_c/\kappa_1$. In other words, the parameter set for which this region contains the solution to the incumbent's problem is

$$\mathbf{R3} = \left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | 2\kappa_c \geq \kappa_1, \rho < 4\kappa_c/\kappa_1\right\}.$$

Expected utility in this case is given by,

$$V_1 = v_1\left(1 + \frac{\kappa_1}{4\kappa_c}\right).$$

8. Case $\lambda_{BC} = 0$ (BC does not bind), $\lambda_{DC} > 0$ (DC binds, consolidation),and $\lambda_i > 0$ ($i_1 = 0$)

In this case the system of conditions is given by,

$$\begin{aligned} a_1 &: \frac{1}{2\kappa_c} - \frac{1}{\kappa_1} - \lambda_{DC} = 0 \\ i_1 &: \frac{\rho}{2} - 1 + \lambda_{DC}\rho\kappa_c + \lambda_i = 0 \\ BC &: \kappa_c v_1 - a_1 = 0. \end{aligned}$$

Since $i_1 = 0$, the DC yields $a_1 = \kappa_c v_1$. For this to be consistent with $\lambda_{DC} > 0$, we must have from the first equation that $\kappa_1 > 2\kappa_c$, and to be consistent with $\lambda_i > 0$ we need $\rho < \kappa_1/(\kappa_1 - \kappa_c)$, yielding,

$$\mathbf{R4} = \left\{(\kappa_1, \rho, v_1) \in \mathbb{R}^3_+ | \kappa_1 > 2\kappa_c,\ \rho < \kappa_1/(\kappa_1 - \kappa_c)\right\}.$$

The expected utility in this case is,

$$V_1 = v_1\left(2 - \frac{\kappa_c}{\kappa_1}\right).$$

■

Proof of Lemma 1:

There are two cases that constitute a solution out of eight possible ones. We will show that the assumption 1 implies that the first case holds when $\frac{v_0 S(m_0)}{v_0 - m_0} < \frac{4}{\kappa_0}$ and the second case holds when $\frac{v_0 S(m_0)}{v_0 - m_0} > \frac{4}{\kappa_0}$. The remaining six cases can be shown to be either inconsistent with any set of parameter values or consistent only with a non-generic set.

1. Case $\lambda_{BC} = 0$ (BC does not bind), $\lambda_{ND} = 0$ (DC does not bind, conflict), and $\lambda_i > 0$ ($i_0 = 0$)

The FOCs are,

$$a_0 : \frac{1}{2}\sqrt{\frac{v_0 S(m_0)}{a_0}} - \frac{1}{\kappa_0} = 0 \tag{23}$$

$$i_0 : \frac{\rho}{2}\sqrt{\frac{a_0}{v_0 S(m_0)}} - 1 + \lambda_i = 0 \tag{24}$$

This implies,

$$a_0 = \frac{\kappa_0^2}{4} v_0 S(m_0)$$

$$\lambda_i = 1 - \frac{\rho \kappa_0}{4}$$

The necessary and sufficient conditions for this case to hold are,

$$\begin{aligned}
\lambda_{BC} &= 0 \Leftrightarrow \frac{v_0 S(m_0)}{v_0 - m_0} < \frac{4}{\kappa_0} \text{ (from the BC not binding)} \\
\lambda_{DC} &= 0 \Leftrightarrow 1 > \frac{\kappa_0^2}{4} \Leftrightarrow \kappa_0 < 2 \text{ (from the DC not binding)} \\
\lambda_i &> 0 \Leftrightarrow 1 - \frac{\rho \kappa_0}{4} > 0 \Leftrightarrow \rho < \frac{4}{\kappa_0} \text{ (from the FOC for } i_0)
\end{aligned}$$

The first inequality holds for values of m_0 low enough given $\kappa_0 < 2$, and the second and third hold by assumption 1.

2. Case $\lambda_{BC} > 0$ (BC binds), $\lambda_{DC} = 0$ (DC does not bind, conflict), and $\lambda_i > 0$ ($i_0 = 0$)

The FOCs are,

$$a_0 : \frac{1}{2}\sqrt{\frac{v_0 S(m_0)}{a_0}} - \frac{1}{\kappa_0} - \frac{\lambda_{BC}}{\kappa_0} = 0 \tag{25}$$

$$i_0 = \frac{\rho}{2}\sqrt{\frac{a_0}{v_0 S(m_0)}} - 1 - \lambda_{BC} + \lambda_i = 0 \tag{26}$$

$\lambda_{BC} > 0$ implies $\kappa_0 (v_0 - m_0) = a_0$.

From the FOCs, we obtain,

$$\lambda_{BC} = \frac{\kappa_0}{2}\sqrt{\frac{v_0 S(m_0)}{a_0}} - 1$$

$$\lambda_i = \frac{\kappa_0}{2}\sqrt{\frac{v_0 S(m_0)}{a_0}} - \frac{\rho}{2}\sqrt{\frac{a_0}{v_0 S(m_0)}}.$$

The parameter conditions for this to be a solution are,

$$\begin{aligned}
\lambda_{BC} &= \frac{\kappa_0}{2}\sqrt{\frac{v_0 S(m_0)}{a_0}} - 1 > 0 \Leftrightarrow \frac{v_0 S(m_0)}{v_0 - m_0} > \frac{4}{\kappa_0} \\
\lambda_{DC} &= 0 \Leftrightarrow \frac{v_0 S(m_0)}{\kappa_0 (v_0 - m_0)} > 1 \\
\lambda_i &= \frac{\kappa_0}{2}\sqrt{\frac{v_0 S(m_0)}{\kappa_0 (v_0 - m_0)}} - \frac{\rho}{2}\sqrt{\frac{\kappa_0 (v_0 - m_0)}{v_0 S(m_0)}} > 0 \Leftrightarrow \frac{v_0 S(m_0)}{\rho (v_0 - m_0)} > 1.
\end{aligned}$$

The inequalities $\frac{v_0 S(m_0)}{\kappa_0 (v_0 - m_0)} > 1$ and $\frac{v_0 S(m_0)}{\rho (v_0 - m_0)} > 1$ are both implied by the condition $\frac{v_0 S(m_0)}{v_0 - m_0} > \frac{4}{\kappa_0}$.

We now cover the cases that are inconsistent with any set of parameter values or consistent with only a non-generic set.

3. Case $\lambda_{BC} > 0$ (BC binds), $\lambda_{DC} > 0$ (DC binds), and $\lambda_i = 0$

If $\lambda_{DC} > 0$ then

$$(v_0 + \rho i_0) S(m_0) = a_0,$$

FOC

$$\frac{\kappa_0}{2} - 1 - \lambda_{BC} - \lambda_{DC}\kappa_0 = 0 \tag{27}$$

$$\frac{\rho}{2} - 1 - \lambda_{BC} + \lambda_{DC}\rho S(m_0) = 0 \tag{28}$$

If $\lambda_{BC} > 0$ then

$$v_0 - m_0 - \frac{a_0}{\kappa_0} - i_0 = 0$$

We have four equations and four unknowns. Therefore

$$\lambda_{DC} = \frac{1}{2}\left(\frac{\kappa_0 - \rho}{\kappa_0 + \rho S(m_0)}\right)$$

$$\lambda_{BC} = \frac{\rho}{2}\left(\frac{\kappa_0 + \kappa_0 S(m_0)}{\kappa_0 + \rho S(m_0)}\right) - 1$$

$$i_0 = \frac{v_0\kappa_0 - m_0\kappa_0 - v_0 S(m_0)}{\kappa_0 + \rho S(m_0)}$$

$$a_0 = \left(\frac{v_0\kappa_0\,(1+\rho) - \rho m_0\kappa_0}{\kappa_0 + \rho S(m_0)}\right) S(m_0)$$

We need to establish the conditions on the parameters and m_0 such that those parameters support this equilibrium.

$$\lambda_{DC} = \frac{1}{2}\left(\frac{\kappa_0 - \rho}{\kappa_0 + \rho S(m_0)}\right) > 0 \iff \kappa_0 > \rho$$

$$\lambda_{BC} = \frac{\rho\kappa_0}{2}\left(\frac{1 + S(m_0)}{\kappa_0 + \rho S(m_0)}\right) - 1 > 0 \iff \rho > \frac{2\kappa_0}{\kappa_0\,(1 + S(m_0)) - 2S(m_0)}$$

$$\iff S(m_0)\left(\frac{\kappa_0}{2} - 1\right) > \frac{\kappa_0}{\rho} - \frac{\kappa_0}{2}.$$

The LHS of this inequality is positive since from above $\lambda_{DC} > 0 \iff \kappa_0 > \rho$ and by assumption $\kappa_0 < 2$. This last assumption also implies the RHS is negative, so the inequality can never hold and this case can never occur.

4. Case $\lambda_{BC} > 0$ (BC binds), $\lambda_{DC} = 0$ (DC does not bind, conflict), and $\lambda_i = 0$ ($i_0 > 0$)

FOC

$$\frac{\kappa_0}{2}\sqrt{\frac{(v_0 + \rho i_0)\,S(m_0)}{a_0}} - 1 - \lambda_{BC} = 0$$

$$\frac{\rho}{2}\sqrt{\frac{a_0}{(v_0 + \rho i_0)\,S(m_0)}} - 1 - \lambda_{BC} = 0$$

If $\lambda_{BC} > 0$ then

$$v_0 - m_0 - \frac{a_0}{\kappa_0} = i_0$$

Equating the two FOCs after solving for λ_{BC} yields $\frac{\kappa_0}{\rho}\left(v_0+\rho i_0\right)S(m_0)=a_0$, and using this into the investment equation above we get $i_0=\frac{v_0\left(1-\frac{S(m_0)}{\rho}\right)-m_0}{1+S(m_0)}$ which then yields $a_0=\frac{\kappa_0}{\rho}\left(\frac{v_0+\rho(v_0-m_0)}{1+S(m_0)}\right)S(m_0)>0$.Using these expressions for a_0 and i_0 we can write,

$$\lambda_{BC}=\frac{\kappa_0}{2}\sqrt{\frac{\left(v_0+\rho\frac{v_0\left(1-\frac{S(m_0)}{\rho}\right)-m_0}{1+S(m_0)}\right)S(m_0)}{\frac{\kappa_0}{\rho}\left(\frac{v_0+\rho(v_0-m_0)}{1+S(m_0)}\right)S(m_0)}}-1=\frac{1}{2}\sqrt{\kappa_0\rho}-1.$$

Since $\lambda_{BC}>0$, it follows that $\sqrt{\kappa_0\rho}>2$, or $\kappa_0\rho>4$, which violates the assumption placing the polity in R3.

5. Case $\lambda_{BC}=0$ (BC does not bind), $\lambda_{DC}=0$ (DC does not bind), and $\lambda_i=0$ ($i_0>0$)

The FOCs are,

$$a_0:\frac{\kappa_0}{2}\sqrt{\frac{\left(v_0+\rho i_0\right)S(m_0)}{a_0}}-1=0 \tag{29}$$

$$i_0:\frac{\rho}{2}\sqrt{\frac{a_0}{\left(v_0+\rho i_0\right)S(m_0)}}-1=0. \tag{30}$$

The first FOC yields $\sqrt{\frac{(v_0+\rho i_0)S(m_0)}{a_0}}=\frac{2}{\kappa_0}$, and substituting into the second FOC we get

$$\rho\kappa_0=4$$

which violates the assumption $\rho\kappa_0<4$, making this an infeasible case.

6. Case $\lambda_{BC}=0$ (BC does not bind, $\lambda_{DC}>0$ (DC binds, deterrence), and $\lambda_i=0$ ($i_0>0$)

The FOCs are,

$$\begin{aligned} a_0 &: \frac{1}{2}-\frac{1}{\kappa_0}-\lambda_{DC}=0 \\ i_0 &: \frac{\rho}{2}-1+\lambda_{DC}\rho S(m_0)=0. \end{aligned}$$

From the first FOC, $\lambda_{DC}=\frac{1}{2}-\frac{1}{\kappa_0}$ and substituting into the second FOC we get,

$$S(m_0)=\frac{\frac{1}{\rho}-\frac{1}{2}}{\frac{1}{2}-\frac{1}{\kappa_0}}.$$

Now we show this cannot hold. $S(m_0)$ is increasing in m_0. The lowest it can be is when staying in R3, yielding $S(m_0)=1+\frac{\kappa_0}{4}$. We show that $\frac{\frac{1}{\rho}-\frac{1}{2}}{\frac{1}{2}-\frac{1}{\kappa_0}}<1+\frac{\kappa_0}{4}$, implying the

equality $S(m_0) = \frac{\frac{1}{\rho}-\frac{1}{2}}{\frac{1}{2}-\frac{1}{\kappa_0}}$ can never hold. This will require $\frac{\frac{1}{\rho}-\frac{1}{2}}{\frac{1}{2}-\frac{1}{\kappa_0}} = \kappa_0\left(1-\frac{1}{\rho}\right) < 1+\frac{\kappa_0}{4}$, or $\frac{3\kappa_0}{4}-1 < \frac{\kappa_0}{\rho}$. If $\frac{3\kappa_0}{4}-1<0$, then the inequality $\frac{3\kappa_0}{4}-1<\frac{\kappa_0}{\rho}$ must always hold, making the equality $S(m_0) = \frac{\frac{1}{\rho}-\frac{1}{2}}{\frac{1}{2}-\frac{1}{\kappa_0}}$ impossible. If $\frac{3\kappa_0}{4}-1>0$, then we need $\rho < \frac{\kappa_0}{\frac{3\kappa_0}{4}-1}$. The RHS of this last inequality is decreasing in κ_0, hence it attains its lowest value at the highest permissible value of κ_0 keeping $\frac{3\kappa_0}{4}-1>0$. This value is 2. Substituting that value into the RHS of $\rho < \frac{\kappa_0}{\frac{3\kappa_0}{4}-1}$ we get,$\rho < \frac{2}{\frac{3}{2}-1} = 4$. Now because in this case $\kappa_0 = 2$ and by assumption $\rho\kappa_0 < 4$, then $\rho < 2$, guaranteeing that $\rho < 4$ and the equality $S(m_0) = \frac{\frac{1}{\rho}-\frac{1}{2}}{\frac{1}{2}-\frac{1}{\kappa_0}}$ is impossible, randering this case infeasible.

7. Case $\lambda_{BC} > 0$ (BC binds), $\lambda_{DC} > 0$ (DC binds, deterrence), and $\lambda_i > 0$ ($i_0 = 0$)

The FOCs are,

$$a_0 : \frac{1}{2} - \frac{1}{\kappa_0} - \frac{\lambda_{BC}}{\kappa_0} - \lambda_{DC} = 0 \tag{31}$$

$$i_0 : \frac{\rho}{2} - 1 - \lambda_{BC} + \lambda_{DC}\rho S(m_0) = 0 \tag{32}$$

$$v_0 - m_0 - \frac{a_0}{\kappa_0} = 0, v_0 S(m_0) - a_0 = 0. \tag{33}$$

Using $\lambda_{DC} = \frac{1}{2} - \frac{1}{\kappa_0} - \frac{\lambda_{BC}}{\kappa_0}$ into $\lambda_{BC} = \frac{\rho}{2} - 1 + \lambda_{DC}\rho S(m_0)$, we obtain $\lambda_{BC} = \frac{\frac{\rho}{2}-1+\left(\frac{1}{2}-\frac{1}{\kappa_0}\right)\rho S(m_0)}{1+\frac{1}{\kappa_0}\rho S(m_0)}$. For $\lambda_{BC} > 0$ we need $\frac{\rho}{2} - 1 + \left(\frac{1}{2}-\frac{1}{\kappa_0}\right)\rho S(m_0) > 0$ or,

$$\frac{\frac{1}{2}-\frac{1}{\rho}}{\frac{1}{\kappa_0}-\frac{1}{2}} > S(m_0),$$

which can never happen. We established that $S(m_0) > \frac{\frac{1}{\rho}-\frac{1}{2}}{\frac{1}{2}-\frac{1}{\kappa_0}}$ when analyzing the previous case, and since $\frac{\frac{1}{\rho}-\frac{1}{2}}{\frac{1}{2}-\frac{1}{\kappa_0}} = \frac{\frac{1}{2}-\frac{1}{\rho}}{\frac{1}{\kappa_0}-\frac{1}{2}}$, the inequality can never hold.

8. Case $\lambda_{BC} = 0$ (BC does not bind), $\lambda_{DC} > 0$ (DC binds, deterrence), and $\lambda_i > 0$ ($i_0 = 0$)

$$a_0 : \frac{1}{2} - \frac{1}{\kappa_0} - \lambda_{DC} = 0 \tag{34}$$

$$i_0 : \frac{\rho}{2} - 1 + \lambda_{DC}\rho S(m_0) + \lambda_i = 0 \tag{35}$$

$$v_0 - m_0 - \frac{a_0}{\kappa_0} > 0,\ v_0 S(m_0) = a_0 \tag{36}$$

From the first FOC,

$$\lambda_{DC} = \frac{1}{2} - \frac{1}{\kappa_0} > 0$$

which cannot happen because it requires $\kappa_0 > 2$, which violates the assumption $\kappa_0 < 2$.

■

Proof of Proposition 4: The way to analyze whether the incumbent is interested in raising m_0 to exit R3 is to analyze the expected utility of doing so. This requires utilizing the function $S(m_0)$ that corresponds to the parametric region where the polity will land in period 1. However, matters are complicated by the presence of two potential cases in period 0 as per Lemma 1, depending on whether $\frac{v_0 S(m_0)}{v_0 - m_0} < \frac{4}{\kappa_0}$, or $\frac{v_0 S(m_0)}{v_0 - m_0} > \frac{4}{\kappa_0}$. These inequalities show that which case should be considered to be in play in period 0–which will affect incentives to raise m_0–depends on $S(m_0)$, which depends on what parametric region the polity will land on in period 1, which in turn depends on whether the incentives are present to make the necessary investments in the first place. Therefore this proof proceeds by checking which combinations of cases in period 0 can obtain for the different possible plans to expand military capacity and land in each of the possible alternative parametric regions in period 1.

The first step to this analysis is to compute the expected utility in period $t = 0$ for each m_0 fixing all the other parameters for the two cases highlighted in Lemma 1:

1. **Case $\lambda_{BC} = 0$ (BC not binding), $\lambda_{ND} = 0$ (DC not binding, conflict), and $\lambda_i > 0$ ($i_0 = 0$)**

The proof to Lemma 1 showed that in this case the Lagrange multiplier conditions defining the case respectively imply $\frac{v_0 S(m_0)}{v_0 - m_0} < \frac{4}{\kappa_0}$, $\kappa_0 < 2$, and $\rho < \frac{4}{\kappa_0}$, and expected utility is,

$$EU = v_0 - m_0 + \frac{\kappa_0}{4} v_0 S(m_0).$$

2. Case $\lambda_{BC} > 0$ (BC binds), $\lambda_{ND} = 0$ (DC not binding, conflict) and $\lambda_i > 0$ ($i_0 = 0$)

The Lagrange multiplier conditions imply, $\frac{v_0 S(m_0)}{v_0 - m_0} > \frac{4}{\kappa_0}$, $\frac{v_0 S(m_0)}{v_0 - m_0} > \kappa_0$, $\frac{v_0 S(m_0)}{v_0 - m_0} > \rho$, and expected utility is,

$$EU = \sqrt{\kappa_0 (v_0 - m_0) v_0 S(m_0)}.$$

The second step is to note that there are critical values of investment in military capacity that shift the regimes the polity is in both in period 0 and period 1. Denote with $\bar{m}$ the

value of m_0 that satisfies $\frac{v_0 S(\bar{m})}{v_0-\bar{m}} = \frac{4}{\kappa_0}$ and which makes the polity switch from case 1 to case 2 in Lemma 1 in period 0.

Part (a) ($\rho < 2$):

Denote with $m_{\mathbf{R3|R4}}$ and $m_{\mathbf{R4|R1}}$ the values of m_0 such that regimes change in period 1 from R3 to R4 and from R4 to R1 respectively: $m_{\mathbf{R3|R4}} = \frac{2-\kappa_0}{\gamma}$ and $m_{\mathbf{R4|R1}} = \frac{1}{\gamma}\left(\frac{\rho}{\rho-1} - \kappa_0\right)$, $m_{\mathbf{R3|R4}} < m_{\mathbf{R4|R1}}$. Because $\bar{m}$ is an implicit function of $S(.)$, we need to compute the conditions on the parameters when $\bar{m}$ lies below and above $m_{\mathbf{R3|R4}}$ and above $m_{\mathbf{R4|R1}}$. The reason it is important to know where $\bar{m}$ lies relative to $m_{\mathbf{R3|R4}}$ and $m_{\mathbf{R4|R1}}$ is that it will indicate which expected utility expression to use to evaluate choices of m_0. If, for example, $\bar{m} > m_{\mathbf{R4|R1}}$ then we know the payoff from choosing an m_0 that keeps the polity in **R3**, moves it to **R4** or an early part of **R1** can be evaluated with a single expected utility expression, namely that in case 1 from Lemma 1.

Before proving part (a) of Proposition **4** we need a technical result. The following lemma establishes the conditions of the parameters that determine the value of $\bar{m}$ relative to $m_{\mathbf{R3|R4}}$ and $m_{\mathbf{R4|R1}}$.

Lemma 2 *Under assumption 1,*

i) If $0 < \gamma v_0 < \frac{8(2-\kappa_0)}{8-3\kappa_0}$, *then* $\bar{m} < m_{\mathbf{R3|R4}}$.

ii) If $\frac{8(2-\kappa_0)}{8-3\kappa_0} < \gamma v_0 < \frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1}-\kappa_0\right)}{\frac{4}{\kappa_0}-\frac{(1+\rho)}{\rho}}$, *then* $m_{\mathbf{R3|R4}} < \bar{m} < m_{\mathbf{R4|R1}}$

ii) If $\frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1}-\kappa_0\right)}{\frac{4}{\kappa_0}-\frac{(1+\rho)}{\rho}} < \gamma v_0$, *then* $m_{\mathbf{R4|R1}} < \bar{m}$

Proof. To determine whether $\bar{m}$ lies within $[0, m_{\mathbf{R3|R4}}]$, $[m_{\mathbf{R3|R4}}, m_{\mathbf{R4|R1}}]$ or $[m_{\mathbf{R4|R1}}, \infty]$ first notice that $\frac{v_0 S(m_0)}{v_0-m_0}$ is increasing in m_0. Therefore, the conditions on the parameters for each of these cases to hold are:

For $\bar{m} < m_{\mathbf{R3|R4}}$ This is the case if $\frac{v_0 S(m_{\mathbf{R3|R4}})}{v_0-m_{\mathbf{R3|R4}}} > \frac{4}{\kappa_0} \iff v_0\gamma < \frac{8(2-\kappa_0)}{8-3\kappa_0}$.

For $m_{\mathbf{R3|R4}} < \bar{m} < m_{\mathbf{R4|R1}}$ From above $m_{\mathbf{R3|R4}} < \bar{m} \iff \frac{8(2-\kappa_0)}{8-3\kappa_0} < v_0\gamma$. Now we need to find the condition for $\bar{m} < m_{\mathbf{R4|R1}}$. This requires $\frac{v_0 S(m_{\mathbf{R4|R1}})}{v_0-m_{\mathbf{R4|R1}}} > \frac{4}{\kappa_0}$, and this follows iff $v_0\gamma < \frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1}-\kappa_0\right)}{\frac{4}{\kappa_0}-\frac{(1+\rho)}{\rho}}$. Therefore, this case occurs when

$$\frac{8(2-\kappa_0)}{8-3\kappa_0} < v_0\gamma < \frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1}-\kappa_0\right)}{\frac{4}{\kappa_0}-\frac{(1+\rho)}{\rho}}$$

For $m_{\mathbf{R4|R1}} < \bar{m}$ It follows directly from before

$$\frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1}-\kappa_0\right)}{\frac{4}{\kappa_0}-\frac{(1+\rho)}{\rho}} < v_0\gamma$$

■

Due to this lemma, we know how to write expected utility depending on the value of γv_0, given all other parameters.

Part (a)1: We only need to find a cutoff in the space of possible values for γv_0 such that the incumbent prefers to stay in **R3**. We propose $\tau_L \equiv \frac{8(2-\kappa_0)}{8-3\kappa_0}$.

In this case, $v_0\gamma < \frac{8(2-\kappa_0)}{8-3\kappa_0}$ is equivalent to a regime described by $\bar{m} < m_{\mathbf{R3|R4}}$, which means whe have to use two different EU expressions depending on whether $m_0 < \bar{m}$, or $m_0 > \bar{m}$. Let us analyze the expected utility in each of these situations. A useful fact will be that $\frac{8(2-\kappa_0)}{8-3\kappa_0}$ is strictly decreasing in κ_0 so its maximum value is at $\kappa_0 = 1$ (since $\kappa_0 > \kappa_c = 1$). In this case $\frac{8(2-1)}{8-3\times 1} = \frac{8}{5} < 2$.

Segment $[0, \bar{m}]$ Expected utility in period $t = 0$ is

$EU = v_0 - m_0 + \frac{\kappa_0}{4}v_0 S(m_0) = v_0 - m_0 + \frac{\kappa_0}{4}v_0\left(1+\frac{\kappa_0+\gamma m_0}{4}\right) = v_0(1+\frac{\kappa_0}{4}+\left(\frac{\kappa_0}{4}\right)^2) + m_0\left(\frac{\kappa_0 v_0\gamma}{16}-1\right)$

Since $\bar{m} < m_{\mathbf{R3|R4}} \iff v_0\gamma < \frac{8(2-\kappa_0)}{8-3\kappa_0}$ then $\frac{\kappa_0 v_0\gamma}{16} - 1 < 0$. To see why, replace $v_0\gamma = \frac{8(2-\kappa_0)}{8-3\kappa_0}$ in $\frac{\kappa_0 v_0\gamma}{16}$ so $\frac{\kappa_0 v_0\gamma}{16} = \frac{\kappa_0\left(\frac{8(2-\kappa_0)}{8-3\kappa_0}\right)}{16} \leq \frac{\kappa_0\frac{8}{5}}{16} < 1$. This implies EU is decreasing in m_0 and the optimal choice is $m_0 = 0$.

Segment $[\bar{m}, m_{\mathbf{R3|R4}}]$ Expected utility in period $t = 0$ is

$EU = \sqrt{\kappa_0(v_0-m_0)v_0 S(m_0)} = \sqrt{\kappa_0(v_0-m_0)v_0\left(1+\frac{\kappa_0+\gamma m_0}{4}\right)}$, and we now show this to decrease in m_0. Note,

$\frac{dEU}{dm} = \frac{1}{2}\left[\kappa_0(v_0-m_0)v_0\left(1+\frac{\kappa_0+\gamma m_0}{4}\right)\right]^{-\frac{1}{2}}\left(-\kappa_0 v_0\left(1+\frac{\kappa_0+\gamma m_0}{4}\right)+\frac{\gamma}{4}\kappa_0(v_0-m_0)v_0\right)$ and,

$\frac{dEU}{dm} < 0 \iff \frac{\gamma}{4}\kappa_0(v_0-m_0)v_0 < \kappa_0 v_0\left(1+\frac{\kappa_0+\gamma m_0}{4}\right)$, or, iff $\gamma v_0 < 4+\kappa_0+2\gamma m_0$.

If $4+\kappa_0+2\gamma m_0$ is higher than $\frac{8(2-\kappa_0)}{8-3\kappa_0}$, our condition to be in this scenario $\bar{m} < m_{\mathbf{R3|R4}}\left(\iff v_0\gamma < \frac{8(2-\kappa_0)}{8-3\kappa_0}\right)$ is a suficient condition for EU in this segment to be decreasing. So, it is sufficient to show that $4+\kappa_0+2\gamma m_0 > \frac{8(2-\kappa_0)}{8-3\kappa_0}$. Because the right hand side is decreasing in κ_0, it attains a maximum at $\kappa_0 = 1$ and it is equal to 8/5 which is smaller than any feasible value of the expression in the left hand side, which is at least 4. Therefore, in this segment utility is maximized at $m_0 = \bar{m}$, and equals $EU = \sqrt{\kappa_0(v_0-m_0)v_0 S(m_0)} = \sqrt{4(v_0-\bar{m})^2} = 2(v_0-\bar{m})$.

Segment $[m_{\mathbf{R3|R4}}, m_{\mathbf{R4|R1}}]$ Since now m_0 can only be larger than $\bar{m}$, we know expected utility is $\sqrt{\kappa_0 (v_0 - m_0) v_0 S(m_0)}$. In **R4**, we have $EU = \sqrt{\kappa_0 (v_0 - m_0) v_0 \left(2 - \frac{1}{\kappa_0 + \gamma m_0}\right)}$. Computing the first derivative with respect to m_0, we get,

$$\frac{dEU}{dm_0} = \left(\kappa_0 (v_0 - m_0) v_0 \left(2 - \frac{1}{\kappa_0 + \gamma m_0}\right)\right)^{-\frac{1}{2}} \left[-\kappa_0 v_0 \left(2 - \frac{1}{\kappa_0 + \gamma m_0}\right) + \frac{\kappa_0 (v_0 - m_0) v_0 \gamma}{(\kappa_0 + \gamma m_0)^2}\right]$$

which is negative whenever $2(\kappa_0 + \gamma m_0) - 1 > \frac{(v_0 - m_0)\gamma}{\kappa_0 + \gamma m_0}$, or, $2(\kappa_0 + \gamma m_0)^2 - \kappa_0 > v_0 \gamma$. Note $2\left(\kappa_0 + \gamma m_{R3|R4}\right)^2 - \kappa_0 = 8 - \kappa_0$. Now note $8 - \kappa_0 > \frac{8(2-\kappa_0)}{8-3\kappa_0} \equiv \tau_L$, since the LHS is at least 6 and the RHS is at most $\frac{8}{5}$. Thus, $\frac{dEU}{dm_0} < 0$ and utility would be maximized at $m_{R3|R4}$ in this segment.

Segment $[m_{\mathbf{R4|R1}}, \infty]$ Here, $EU = \sqrt{\kappa_0 (v_0 - m_0) v_0 \frac{(\kappa_0 + \gamma m_0)(1+\rho)}{\kappa_0 + \gamma m_0 + \rho}}$ and

$$\frac{dEU}{dm_0} = \left(\kappa_0 (v_0 - m_0) v_0 \frac{(\kappa_0 + \gamma m_0)(1+\rho)}{\kappa_0 + \gamma m_0 + \rho}\right)^{-\frac{1}{2}} \begin{pmatrix} -\kappa_0 v_0 \frac{(\kappa_0 + \gamma m_0)(1+\rho)}{\kappa_0 + \gamma m_0 + \rho} \\ +\kappa_0 (v_0 - m_0) v_0 \frac{\gamma(1+\rho)(\kappa_0 + \gamma m_0 + \rho) - (\kappa_0 + \gamma m_0)(1+\rho)\gamma}{(\kappa_0 + \gamma m_0 + \rho)^2} \end{pmatrix}.$$

Note $\frac{dEU}{dm_0} < 0$ whenever $\gamma v_0 < \frac{(\kappa_0 + \gamma m_0 + \rho)(\kappa_0 + \gamma m_0)}{\rho} + \gamma m_0$. The right hand side of this expression is increasing in m_0, so the minimum is attained at $m_0 = m_{\mathbf{R4|R1}}$ and it equals $\frac{\rho}{(\rho-1)^2} + 2\frac{\rho}{(\rho-1)} - \kappa_0$. The highest possible value of γv_0, $\tau_L = \frac{8(2-\kappa_0)}{8-3\kappa_0}$ is smaller than $\frac{8}{5}$ which, in turn, is always smaller than $\frac{\rho}{(\rho-1)^2} + 2\frac{\rho}{(\rho-1)} - \kappa_0$ given that $\rho < 2$. Therefore the maximum of EU in this segment is attained at $m_0 = m_{\mathbf{R4|R1}}$.

Considering all of the segments together, we now show that **the global maximum in this case is $m_0 = 0$.** This follows from the just demonstrated fact that the maximum within each segment of the support is at the minimum value, and from the fact that EU is continuous. $S(.)$ is continuous for all m_0 and EU in period $t = 0$ is also continuous at $\bar{m}$. In $t = 0$, in segment $[0, \bar{m}]$ EU evaluated at $\bar{m}$ is $2(v_0 - \bar{m})$ which is equal to the EU in segment $[\bar{m}, m_{\mathbf{R3|R4}}]$ evaluated at $\bar{m}$. This can be shown noticing that $\frac{\kappa_0 v_0 S(\bar{m})}{4} = v_0 - \bar{m}$, and reeplacing in EU in segment $[0, \bar{m}]$. Thus, the polity will stay at **R3** in period 1.

Part (a)(2-3) For these parts of Proposition **4** (the existence of cutoffs such that the polity will move away from **R3** in period 1), we consider the case in which $\frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1} - \kappa_0\right)}{\frac{4}{\kappa_0} - \frac{(1+\rho)}{\rho}} < v_0 \gamma$ (as we only need to focus on a sufficient condition for the polity to exit **R3** and move respectively into **R4** or **R1**). In this case $m_{\mathbf{R4|R1}} < \bar{m}$ by Lemma 2. We proceed by analyzing the optimal decision of m_0 under the different segments of the domain of m_0:

Segment $[0, m_{\mathbf{R3|R4}}]$ Expected utility in period $t = 0$ is

$EU = v_0 - m_0 + \frac{\kappa_0}{4} v_0 S(m_0) = v_0 - m_0 + \frac{\kappa_0}{4} v_0 \left(1 + \frac{\kappa_0 + \gamma m_0}{4}\right) = v_0(1 + \frac{\kappa_0}{4} + \left(\frac{\kappa_0}{4}\right)^2) + m_0 \left(\frac{\kappa_0 v_0 \gamma}{16} - 1\right)$. In this case, $\frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1} - \kappa_0\right)}{\frac{4}{\kappa_0} - \frac{(1+\rho)}{\rho}} > \frac{16}{\kappa_0}$ does not always hold under Assumption 1, so marginal utility is not necessarily positive. Therefore, a sufficient condition for the polity to exit **R3** is that γv_0 be higher than $\tau_M \equiv \max\left\{\frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1} - \kappa_0\right)}{\frac{4}{\kappa_0} - \frac{(1+\rho)}{\rho}}, \frac{16}{\kappa_0}\right\}$. Next, we show that it may either stay in **R4** or move to **R1**.

Segment $[m_{\mathbf{R3}|\mathbf{R4}}, m_{\mathbf{R4}|\mathbf{R1}}]$ Expected utility in period $t = 0$ is $EU = v_0 - m_0 + \frac{\kappa_0}{4} v_0 S(m_0) = v_0 - m_0 + \frac{\kappa_0}{4} v_0 \left(2 - \frac{1}{\kappa_0 + \gamma m_0}\right)$ (recall that since $\frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1} - \kappa_0\right)}{\frac{4}{\kappa_0} - \frac{(1+\rho)}{\rho}} < v_0\gamma$, $\bar{m} > m_{\mathbf{R4}|\mathbf{R1}}$, so in the interval $[m_{\mathbf{R3}|\mathbf{R4}}, m_{\mathbf{R4}|\mathbf{R1}}]$ we are considering $m_0 < \bar{m}$).

The marginal utility of m_0 is: $-1 + \frac{\kappa_0 v_0}{4} \frac{\gamma}{(\kappa_0 + \gamma m_0)^2}$. The value of m_0 that maximizes EU is $m_0 = \frac{1}{\gamma}\left(\sqrt{\frac{\kappa_0 v_0 \gamma}{4}} - \kappa_0\right)$. For the optimum to fall in the segment $[m_{\mathbf{R3}|\mathbf{R4}}, m_{\mathbf{R4}|\mathbf{R1}}]$ we need, in addition to $\gamma v_0 > \tau_M$, that,

$$m_{\mathbf{R3}|\mathbf{R4}} < \frac{1}{\gamma}\left(\sqrt{\frac{\kappa_0 v_0 \gamma}{4}} - \kappa_0\right) < m_{\mathbf{R4}|\mathbf{R1}},$$

The first inequality requires $4\frac{(2-\kappa_0)^2}{\kappa_0} < v_0\gamma$. Since $\tau_M \equiv \max\left\{\frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1} - \kappa_0\right)}{\frac{4}{\kappa_0} - \frac{(1+\rho)}{\rho}}, \frac{16}{\kappa_0}\right\}$ and $\frac{16}{\kappa_0} > 4\frac{(2-\kappa_0)^2}{\kappa_0}$, a sufficient condition for the first inequality is that $v_0\gamma > \tau_M$. The second inequality requires $v_0\gamma < \frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1}\right)^2$. Note that $\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1}\right)^2 > \tau_M$ (because simple algebra shows that $\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1}\right)^2 > \frac{16}{\kappa_0}$, and $\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1}\right)^2 > \frac{\frac{4}{\kappa_0}\left(\frac{\rho}{\rho-1} - \kappa_0\right)}{\frac{4}{\kappa_0} - \frac{(1+\rho)}{\rho}}$). Defining $\tau_H \equiv \left(\frac{\rho}{\rho-1}\right)^2 \frac{4}{\kappa_0}$, we conclude that whenever $\tau_M < \gamma v_0 < \tau_H$ the polity reaches **R4**. If $\tau_H < \gamma v_0$ then the polity must reach **R1**. Now we explore if it is possible to move into the interior of **R1**.

Segment $[m_{\mathbf{R4}|\mathbf{R1}}, \bar{m}]$ Expected utility is $EU = v_0 - m_0 + \frac{\kappa_0}{4} v_0 S(m_0) = v_0 - m_0 + \frac{\kappa_0}{4} v_0 \left(\frac{(\kappa_0 + \gamma m_0)(1+\rho)}{\kappa_0 + \gamma m_0 + \rho}\right)$. This is a concave function with a maximum at $m_0 = \frac{1}{\gamma}\begin{pmatrix} \sqrt{\frac{\kappa_0 v_0 \gamma}{4}\rho(1+\rho)} \\ -\kappa_0 - \rho \end{pmatrix}$. Note if $\gamma v_0 > \tau_H = \left(\frac{\rho}{\rho-1}\right)^2 \frac{4}{\kappa_0}$ then that maximum is an interior optimum and larger than $\mathbf{R4}|\mathbf{R1}$. To see this, rewrite the inequality $m_0 > \mathbf{R4}|\mathbf{R1}$ as $\sqrt{\frac{\kappa_0 v_0 \gamma}{4}\rho(1+\rho) - \kappa_0 - \rho} > \frac{\rho}{\rho-1} - \kappa_0$ then plug τ_H into the LHSto obtain $1 + \rho > \rho$, which always holds. Therefore for $\gamma v_0 > \tau_H$ the polity will be in the interior of **R1** in period 1.

Part (b) ($\rho \geq 2$): As in the proof of part (a), $\bar{m}$ satisfies the equation: $\frac{v_0 S(\bar{m})}{v_0 - \bar{m}} = \frac{4}{\kappa_0}$, and represents the value of m_0 at which the polity switches from case 1 to case 2 from Lemma 1 in period 0.

Let us call $m_{\mathbf{R3|R2}} = \frac{1}{\gamma}\left(\frac{4}{\rho} - \kappa_0\right)$ and $m_{\mathbf{R2|R1}} = \frac{1}{\gamma}(\rho - \kappa_0)$ the values in which regimes change in period 1. The following lemma shows the conditions on the parameters such that for any given m_0 we can fully describe the *EU* in period 0.

Lemma 3 *Under assumption 1,*

i) If $0 < \gamma v_0 < \frac{16-4\kappa_0\rho}{4\rho-(1+\rho)\kappa_0}$ *then* $\bar{m} < m_{\mathbf{R3|R2}}$

ii) If $\frac{16-4\kappa_0\rho}{4\rho-(1+\rho)\kappa_0} < \gamma v_0 < \frac{8(\rho-\kappa_0)}{8-(1+\rho)\kappa_0}$ *then* $m_{\mathbf{R3|R2}} < \bar{m} < m_{\mathbf{R2|R1}}$

ii) If $\frac{8(\rho-\kappa_0)}{8-(1+\rho)\kappa_0} < \gamma v_0$ *then* $m_{\mathbf{R2|R1}} < \bar{m}$

Proof. It follows from replacing the definitions of $\bar{m}, m_{\mathbf{R3|R2}}, m_{\mathbf{R2|R1}}$ and following steps analogous to Lemma 2, as follows. To determine whether $\bar{m}$ lies within $[0, m_{\mathbf{R3|R2}}]$, $[m_{\mathbf{R3|R2}}, m_{\mathbf{R2|R1}}]$ or $[m_{\mathbf{R2|R1}}, \infty]$ recall that $\frac{v_0 S(m_0)}{v_0 - m_0}$ is increasing in m_0. Therefore, the conditions on the parameters for each of these cases to hold are:

For $\bar{m} < m_{\mathbf{R3|R2}}$ This is the case when $\frac{v_0 S(m_{\mathbf{R3|R2}})}{v_0 - m_{\mathbf{R3|R2}}} > \frac{4}{\kappa_0} \iff v_0\gamma < \frac{16-4\kappa_0\rho}{4\rho-(1+\rho)\kappa_0}$.

For $m_{\mathbf{R3|R2}} < \bar{m} < m_{\mathbf{R2|R1}}$ From above $m_{\mathbf{R3|R2}} < \bar{m} \iff \frac{16-4\kappa_0\rho}{4\rho-(1+\rho)\kappa_0} < v_0\gamma$. Now we need to find the condition for $\bar{m} < m_{\mathbf{R2|R1}}$. This requires $\frac{v_0 S(m_{\mathbf{R2|R1}})}{v_0 - m_{\mathbf{R2|R1}}} > \frac{4}{\kappa_0}$, or equivalently $v_0\gamma < \frac{8(\rho-\kappa_0)}{8-(1+\rho)\kappa_0}$. Therefore, this case occurs when

$$\frac{16 - 4\kappa_0\rho}{4\rho - (1+\rho)\kappa_0} < v_0\gamma < \frac{8(\rho - \kappa_0)}{8 - (1+\rho)\kappa_0} \tag{37}$$

For $m_{\mathbf{R2|R1}} < \bar{m}$ It follows directly from before

$$\frac{8(\rho - \kappa_0)}{8 - (1+\rho)\kappa_0} < v_0\gamma.$$

■

Due to this lemma, we know how to write expected utility depending on the value of γv_0, given all other parameters.

Part (b)1 We need to find a cutoff such that the polity stays in **R3**. We propose $\sigma_L \equiv \frac{16-4\kappa_0\rho}{4\rho-(1+\rho)\kappa_0}$. In this case, $v_0\gamma < \sigma_L$ is equivalent to a regime in which $\bar{m} < m_{\mathbf{R3|R2}}$ which means we have to use two different EU expressions depending on whether $m_0 < \bar{m}$ or $m_0 > \bar{m}$. Let us analyze the expected utility in each of these situations. A useful fact will be that $\sigma_L = \frac{16-4\kappa_0\rho}{4\rho-(1+\rho)\kappa_0}$ is decreasing in both ρ and κ_0. Thus, its maximum value is at $\rho = 2$ and $\kappa_0 = 1$. Then $\sigma_L = \frac{16-4\times1\times2}{4\times2-(1+2)\times1} = \frac{8}{5}$.

Segment $[0, \bar{m}]$ Expected utility in period 0 is

$EU = v_0 - m_0 + \frac{\kappa_0}{4} v_0 S(m_0) = v_0 - m_0 + \frac{\kappa_0}{4} v_0 \left(1 + \frac{\kappa_0 + \gamma m_0}{4}\right) = v_0(1 + \frac{\kappa_0}{4} + \left(\frac{\kappa_0}{4}\right)^2) + m_0 \left(\frac{\kappa_0 v_0 \gamma}{16} - 1\right)$.

Since $\bar{m} < m_{\mathbf{R3|R2}} \iff v_0\gamma < \sigma_L$ then it follows that $\frac{\kappa_0 \sigma_L}{16} > \frac{\kappa_0 v_0 \gamma}{16}$. Therefore, if $\frac{\kappa_0 \sigma_L}{16} < 1$, it must follow that $\frac{\kappa_0 v_0 \gamma}{16} - 1 < 0$, implying the optimal choice is $m_0 = 0$. To see that $\frac{\kappa_0 \sigma_L}{16} < 1$, note that σ_L is at at most $\frac{8}{5}$ and $\kappa_0 < 2$.

Segment $[\bar{m}, m_{\mathbf{R3|R2}}]$ Expected utility in period 0 is

$EU = \sqrt{\kappa_0 v_0 (v_0 - m_0) \left(1 + \frac{\kappa_0 + \gamma m_0}{4}\right)}$. To show this payoff is decreasing in m_0 note that,

$\frac{dEU}{dm_0} = \left(\kappa_0 v_0 (v_0 - m_0) \left(1 + \frac{\kappa_0 + \gamma m_0}{4}\right)\right)^{-\frac{1}{2}} \left(-\kappa_0 v_0 \left(1 + \frac{\kappa_0 + \gamma m_0}{4}\right) + \frac{\gamma}{4} \kappa_0 v_0 (v_0 - m_0)\right)$, which must be negative because $\frac{\gamma}{4}(v_0 - m_0) < 1 + \frac{\kappa_0 + \gamma m_0}{4} \Leftrightarrow \gamma v_0 < 4 + \kappa_0 + 2\gamma m_0$ which must hold since $\gamma v_0 < \sigma_L < 4$. As a result, the maximum is attained at $m_0 = 0$ in the interval $[0, m_{\mathbf{R3|R2}}]$.

Segment $[m_{\mathbf{R3|R2}}, m_{\mathbf{R2|R1}}]$ Expected utility in period 0 is,

$EU = \sqrt{\kappa_0 v_0 (v_0 - m_0) \left(\sqrt{\frac{\kappa_0 + \gamma m_0}{\rho}} \frac{(1+\rho)}{2}\right)}$. Marginal expected utility is,

$\frac{dEU}{dm_0} = \frac{\sqrt{\kappa_0 v_0}}{2} \left((v_0 - m_0) \left(\sqrt{\frac{\kappa_0 + \gamma m_0}{\rho}} \frac{(1+\rho)}{2}\right)\right)^{-\frac{1}{2}} \begin{pmatrix} -\left(\sqrt{\frac{\kappa_0 + \gamma m_0}{\rho}} \frac{(1+\rho)}{2}\right) \\ +(v_0 - m_0) \frac{(1+\rho)}{4} \left(\frac{\kappa_0 + \gamma m_0}{\rho}\right)^{-\frac{1}{2}} \frac{\gamma}{\rho} \end{pmatrix}$ and this is negative whenever $(v_0 - m_0) \frac{1}{2} \left(\frac{\kappa_0 + \gamma m_0}{\rho}\right)^{-\frac{1}{2}} \frac{\gamma}{\rho} < \sqrt{\frac{\kappa_0 + \gamma m_0}{\rho}}$, or, $\gamma v_0 < 2\kappa_0 + 3\gamma m_0$. The RHS of this expression is higher than $\frac{8}{5}$, and since $\gamma v_0 < \sigma_L \leq \frac{8}{5}$, we must conclude that $\frac{dEU}{dm_0} < 0$. In this segment EU would be maximized at $m_0 = m_{\mathbf{R3|R2}}$.

Segment $[m_{\mathbf{R2|R1}}, \infty]$ Expected utility in period 0 is,

$EU = \sqrt{\kappa_0 v_0 (v_0 - m_0) \frac{(\kappa_0 + \gamma m_0)(1+\rho)}{\kappa_0 + \gamma m_0 + \rho}}$. Marginal expected utility is,

$\frac{dEU}{dm_0} = \frac{\sqrt{\kappa_0 v_0}}{2} \left((v_0 - m_0) \frac{(\kappa_0 + \gamma m_0)(1+\rho)}{\kappa_0 + \gamma m_0 + \rho}\right)^{-\frac{1}{2}} \begin{pmatrix} -\frac{(\kappa_0 + \gamma m_0)(1+\rho)}{\kappa_0 + \gamma m_0 + \rho} \\ +(v_0 - m_0)(1+\rho) \frac{\rho}{(\kappa_0 + \gamma m_0 + \rho)^2} \end{pmatrix}$. This is negative whenever $(v_0 - m_0) \frac{\rho}{\kappa_0 + \gamma m_0 + \rho} < (\kappa_0 + \gamma m_0)$, or whenever,

$$v_0 \gamma < \frac{1}{\rho} (\kappa_0 + \gamma m_0)^2 + (\kappa_0 + \gamma m_0) + \gamma m_0.$$

The right hand side of this expression is increasing in m_0, so the minimum of this expression is attained at $m_0 = m_{\mathbf{R2|R1}}$ and it equals $3\rho - \kappa_0$. Since $\gamma v_0 < \frac{8}{5} < 3\rho - \kappa_0$ the result follows.

In sum, the global maximum when $\bar{m} < m_{\mathbf{R3|R2}} \left(\iff v_0\gamma < \frac{16 - 4\kappa_0\rho}{4\rho - (1+\rho)\kappa_0} = \sigma_L \right)$ **is** $m_0 = 0$**.** This follows from the just demonstrated fact that the maximum within each

segment of the support is at the minimum point, and from the fact that EU is continuous. $S(.)$ is continuous for all m_0 and hence EU in period 0 is also continuous at $\bar{m}$. In period 0 in segment $[0, \bar{m}]$, EU evaluated at $\bar{m}$ is $2(v_0 - \bar{m})$ which is equal to EU in segment $[\bar{m}, m_{\mathbf{R3|R2}}]$ evaluated at $\bar{m}$. This can be shown noticing that $\frac{\kappa_0 v_0 S(\bar{m})}{4} = v_0 - \bar{m}$, and substituting into the expression for EU in segment $[0, \bar{m}]$. Thus, the polity will stay at **R3** in period 1 whenever $\gamma v_0 < \sigma_L$. Next, we show that there exist $\sigma_H = 16/\kappa_0 > \sigma_{M2}$, such that the polity moves to **R1**.

Part (b)(2) To prove the existence of values for γv_0 high enough that the polity will move away from **R3** into **R1**, consider the case in which $\frac{16}{\kappa_0} < \gamma v_0$. Note that $\frac{8(\rho - \kappa_0)}{8 - (1+\rho)\kappa_0} < \frac{16}{\kappa_0}$, so $m_{\mathbf{R2|R1}} < \bar{m}$. We proceed by analyzing the optimal m_0 in each segment in what follows.

Segment $[0, m_{\mathbf{R3|R2}}]$ Expected utility in period $t = 0$ is,

$EU = v_0 - m_0 + \frac{\kappa_0}{4} v_0 S(m_0) = v_0 - m_0 + \frac{\kappa_0}{4} v_0 \left(1 + \frac{\kappa_0 + \gamma m_0}{4}\right) = v_0(1 + \frac{\kappa_0}{4} + \left(\frac{\kappa_0}{4}\right)^2) + m_0 \left(\frac{\kappa_0 v_0 \gamma}{16} - 1\right)$. Marginal utility is positive iff $\gamma v_0 > \frac{16}{\kappa_0}$, so the polity moves to **R2**, or advances to **R1.**

Segment $[m_{\mathbf{R3|R2}}, m_{\mathbf{R2|R1}}]$ The (concave) expected utility is, $EU = v_0 - m_0 + \frac{\kappa_0 v_0 (1+\rho)}{8} \sqrt{\frac{\kappa_0 + \gamma m_0}{\rho}}$.

The marginal utility of m_0 is: $-1 + \frac{\kappa_0 v_0 (1+\rho)}{8} \frac{1}{2} \sqrt{\frac{\rho}{\kappa_0 + \gamma m_0}} \frac{\gamma}{\rho}$. This marginal utility may be either negative, zero or positive in $[m_{\mathbf{R3|R2}}, m_{\mathbf{R2|R1}}]$. Given concavity, it is negative iff the value at which the marginal utility is zero, $m_0 = \frac{\left(\frac{\kappa_0 v_0 \gamma}{16} \frac{1+\rho}{\sqrt{\rho}}\right)^2 - \kappa_0}{\gamma}$, is smaller than $m_{\mathbf{R3|R2}}$:

$$m_0 = \frac{\left(\frac{\kappa_0 v_0 \gamma}{16} \frac{1+\rho}{\sqrt{\rho}}\right)^2 - \kappa_0}{\gamma} < m_{\mathbf{R3|R2}} = \frac{(\frac{4}{\rho} - \kappa_0)}{\gamma}$$

$\iff v_0 \gamma < \left(\frac{16}{\kappa_0}\right) \frac{2}{1+\rho}$. In this case, the maximum EU in this segment is at $m_0 = m_{\mathbf{R3|R2}}$.

The solution is interior in **R2** if $m_0 = \frac{\left(\frac{\kappa_0 v_0 \gamma}{16} \frac{(1+\rho)}{\sqrt{\rho}}\right)^2 - \kappa_0}{\gamma}$ is in between $m_{\mathbf{R3|R2}}$ and $m_{\mathbf{R2|R1}}$:

$$\frac{(\frac{4}{\rho} - \kappa_0)}{\gamma} = m_{\mathbf{R3|R2}} < \frac{\left(\frac{\kappa_0 v_0 \gamma}{16} \frac{(1+\rho)}{\sqrt{\rho}}\right)^2 - \kappa_0}{\gamma} < m_{\mathbf{R2|R1}} = \frac{(\rho - \kappa_0)}{\gamma}$$

$\iff \frac{16}{\kappa_0} \frac{2}{1+\rho} < v_0 \gamma < \frac{16}{\kappa_0} \frac{\rho}{1+\rho}$. But since $\frac{16}{\kappa_0} \frac{\rho}{1+\rho} < \frac{16}{\kappa_0}$, then $v_0 \gamma > \frac{16}{\kappa_0} \frac{\rho}{1+\rho}$ and the incumbent moves to **R1.** The following analysis shows the incumbent moves to the interior of **R1** in this case.

Segment $[m_{\mathbf{R2}|\mathbf{R1}}, \bar{m}]$ Expected utility is $EU = v_0 - m_0 + \frac{\kappa_0}{4} v_0 S(m_0) = v_0 - m_0 + \frac{\kappa_0 v_0}{4}\left(\frac{(\kappa_0+\gamma m_0)(1+\rho)}{\kappa_0+\gamma m_0+\rho}\right)$, and marginal utility is $\frac{dEU}{dm_0} = \frac{\kappa_0 v_0}{4}\frac{(\kappa_0+\gamma m_0+\rho)\gamma(1+\rho)-(\kappa_0+\gamma m_0)\gamma(1+\rho)}{(\kappa_0+\gamma m_0+\rho)^2} = \frac{\kappa_0 v_0}{4}\frac{\gamma\rho(1+\rho)}{(\kappa_0+\gamma m_0+\rho)^2} - 1$, so the interior optimum is

$m_0 = \frac{1}{\gamma}\left(\sqrt{\frac{\kappa_0 v_0 \gamma}{4}\rho(1+\rho)} - \kappa_0 - \rho\right)$. It is straightforward (using arguments analogous to the case in segment $[m_{\mathbf{R3}|\mathbf{R2}}, m_{\mathbf{R2}|\mathbf{R1}}]$) to show that if $\frac{16}{\kappa_0}\frac{\rho}{(1+\rho)} < \gamma v_0$ then we also have $m_0 = \frac{1}{\gamma}\left(\sqrt{\frac{\kappa_0 v_0 \gamma}{4}\rho(1+\rho)} - \kappa_0 - \rho\right) > m_{\mathbf{R2}|\mathbf{R1}}$. This implies that for $\gamma v_0 > \frac{16}{\kappa_0} = \sigma_H$ the polity must move to the interior of **R1** in period 1.

In sum, whenever $\sigma_H < \gamma v_0$ the polity moves to **R1.** In what follows, we show that there exist

$$\sigma_{M1} = \max\left\{\frac{16-4\kappa_0\rho}{4\rho-(1+\rho)\kappa_0}, \frac{\frac{4}{\kappa_0} - \sqrt{\left(\frac{4}{\kappa_0}\right)^2 - 3\times\frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{4}\right)}}{2\times\frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{16}\times\frac{3}{4}\right)}\right\}$$

$$\sigma_{M2} = \min\left\{\frac{8(\rho-\kappa_0)}{8-(1+\rho)\kappa_0}, \frac{\frac{4}{\kappa_0} + \sqrt{\left(\frac{4}{\kappa_0}\right)^2 - 3\times\frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{4}\right)}}{2\times\frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{16}\times\frac{3}{4}\right)}, 3\rho-\kappa_0\right\},$$

with $\sigma_L < \sigma_{M1} < \sigma_H$ and $\sigma_L < \sigma_{M2} < \sigma_H$, such that the polity moves to **R2** whenever $\sigma_{M1} < \gamma v_0 < \sigma_{M2}$.

Part (b)(3) To prove the existence of values of γv_0 such that the polity will move into **R2**, consider the case in which (37) holds, so $m_{\mathbf{R3}|\mathbf{R2}} < \bar{m} < m_{\mathbf{R2}|\mathbf{R1}}$. We characterize the set of parameters (κ_0, ρ) such that the optimal point lies in **R2**.

Segment $[0, m_{\mathbf{R3}|\mathbf{R2}}]$ Expected utility in period $t = 0$ is, $EU = v_0 - m_0 + \frac{\kappa_0}{4} v_0 S(m_0) = v_0 - m_0 + \frac{\kappa_0}{4} v_0\left(1 + \frac{\kappa_0+\gamma m_0}{4}\right) = v_0(1 + \frac{\kappa_0}{4} + \left(\frac{\kappa_0}{4}\right)^2) + m_0\left(\frac{\kappa_0 v_0 \gamma}{16} - 1\right)$. Marginal utility is positive iff $\gamma v_0 > \frac{16}{\kappa_0}$ and negative iff $\gamma v_0 < \frac{16}{\kappa_0}$. Algebra shows that under Assumption 1 $\frac{16-4\kappa_0\rho}{4\rho-(1+\rho)\kappa_0} < \frac{8(\rho-\kappa_0)}{8-(1+\rho)\kappa_0} < \frac{16}{\kappa_0}$. Thus, $m_0 = 0$ is optimal in $[0, m_{\mathbf{R3}|\mathbf{R2}}]$ and the polity will stay in **R3**. EU evaluated at $m_0 = 0$ is $v_0(1 + \frac{\kappa_0}{4} + \left(\frac{\kappa_0}{4}\right)^2)$, and we later show that this value is lower than EU at the optimum in **R2** whenever $\sigma_{M1} < \gamma v_0 < \sigma_{M2}$.

Segment $[m_{\mathbf{R3}|\mathbf{R2}}, \bar{m}]$ There are two possibilities for an optimum in **R2**. It could lie in $[m_{\mathbf{R3}|\mathbf{R2}}, \bar{m}]$ or in $[\bar{m}, m_{\mathbf{R2}|\mathbf{R1}}]$. We only need to show the result for one of the two cases so we focus on the first. The (concave) expected utility is,

$EU = v_0 - m_0 + \frac{\kappa_0 v_0 (1+\rho)}{8}\sqrt{\frac{\kappa_0+\gamma m_0}{\rho}}$. The marginal utility of m_0 is: $-1+\frac{\kappa_0 v_0(1+\rho)}{8}\frac{1}{2}\sqrt{\frac{\rho}{\kappa_0+\gamma m_0}}\frac{\gamma}{\rho}$. Hence, the optimum point is given by $m_{IntR2} = \frac{\left(\frac{\kappa_0 v_0 \gamma}{16}\frac{1+\rho}{\sqrt{\rho}}\right)^2 - \kappa_0}{\gamma}$. This point is less than $\bar{m}$ if and only if $\frac{v_0 S(m_{IntR2})}{v_0 - m_{IntR2}} < \frac{4}{\kappa_0}$, which is equivalent to,

$$\frac{\frac{4}{\kappa_0} - \sqrt{\left(\frac{4}{\kappa_0}\right)^2 - 3 \times \frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{4}\right)}}{2 \times \frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{16} \times \frac{3}{4}\right)} < \gamma v_0 < \frac{\frac{4}{\kappa_0} + \sqrt{\left(\frac{4}{\kappa_0}\right)^2 - 3 \times \frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{4}\right)}}{2 \times \frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{16} \times \frac{3}{4}\right)}. \tag{38}$$

Then $EU\left(m_{IntR2}\right)$ is $\frac{1}{\gamma}\left(\gamma v_0\left(1 + v_0\gamma\left(\frac{\kappa_0}{16}\frac{(1+\rho)}{\sqrt{\rho}}\right)^2\right) + \kappa_0\right)$. Hence, the optimal value lies in $[m_{\mathbf{R3}|\mathbf{R2}}, \bar{m}]$ if $EU\left(m_{IntR2}\right) > EU\left(m_0 = 0\right)$. This inequality implies (after some algebra) that $\left(\frac{\kappa_0}{4}\left(1+\frac{\kappa_0}{4}\right) - \frac{\kappa_0}{\gamma}\right)\left(\frac{16}{\kappa_0}\frac{\sqrt{\rho}}{(1+\rho)}\right)^2 < v_0\gamma$. Note the LHS of this expression is negative as long as $\gamma < 16/(4+\kappa_0)$, the technical assumption introduced in the text.

Our strategy for this part of the proof is to find the combinations of (κ_0, ρ) for which $EU\left(m_{IntR2}\right) > EU\left(m_0 = 0\right)$. This is a sufficient condition for the optimal point to lie in **R2** as long as EU is decreasing in m_0 over **R1.**

Segment $[m_{\mathbf{R2}|\mathbf{R1}}, \infty]$ Expected utility is

$EU = \sqrt{\kappa_0 v_0 \left(v_0 - m_0\right) S(m_0)} = \sqrt{\kappa_0 v_0 \left(v_0 - m_0\right)\frac{(\kappa_0+\gamma m_0)(1+\rho)}{\kappa_0+\gamma m_0+\rho}}$. Marginal utility is, $\frac{dEU}{dm_0} = \sqrt{\kappa_0 v_0}\frac{1}{2}\left[-\left(v_0 - m_0\right)^{-\frac{1}{2}}\sqrt{\frac{(\kappa_0+\gamma m_0)(1+\rho)}{\kappa_0+\gamma m_0+\rho}} + \sqrt{\left(v_0 - m_0\right)}\left(\frac{(\kappa_0+\gamma m_0)(1+\rho)}{\kappa_0+\gamma m_0+\rho}\right)^{-\frac{1}{2}}\frac{\gamma\rho(1+\rho)}{(\kappa_0+\gamma m_0+\rho)^2}\right]$. Note $\frac{dEU}{dm_0} < 0$ iff $-(\kappa_0+\gamma m_0)(\kappa_0+\gamma m_0+\rho) + \gamma\rho\left(v_0 - m_0\right) < 0$. The LHS of this expression is decreasing in m_0 so if evaluated at $m_{\mathbf{R2}|\mathbf{R1}} = \frac{\rho-\kappa_0}{\gamma}$ such LHS is negative, then EU is decreasing in m_0 within **R1.** Evaluating at $m_{\mathbf{R2}|\mathbf{R1}} = \frac{\rho-\kappa_0}{\gamma}$ yields,

$$\gamma v_0 < 3\rho - \kappa_0. \tag{39}$$

Combining (37), (38), and (39), we define $\sigma_{M1} = \max\left\{\frac{16-4\kappa_0\rho}{4\rho-(1+\rho)\kappa_0}, \frac{\frac{4}{\kappa_0} - \sqrt{\left(\frac{4}{\kappa_0}\right)^2 - 3\times\frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{4}\right)}}{2\times\frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{16}\times\frac{3}{4}\right)}\right\}$ and $\sigma_{M2} = \min\left\{\frac{8(\rho-\kappa_0)}{8-(1+\rho)\kappa_0}, \frac{\frac{4}{\kappa_0} + \sqrt{\left(\frac{4}{\kappa_0}\right)^2 - 3\times\frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{4}\right)}}{2\times\frac{(1+\rho)^2}{\rho}\left(\frac{\kappa_0}{16}\times\frac{3}{4}\right)}, 3\rho - \kappa_0\right\}$. As a result, if $\sigma_{M1} < \gamma v_0 < \sigma_{M2}$ then the polity moves to **R2**. Note the set over the real line bounded by σ_{M1} and σ_{M2} is non-empty for a measurable set of (κ_0, ρ). For example, fix $\rho = 4$ $\kappa_0 = 1$. In this case, the conditions (37), (38), and (39) become respectively $0 < \gamma v_0 < 8$, $\frac{\frac{5}{16} - \sqrt{\left(\frac{5}{16}\right)^2 - 4\left(\frac{5}{16\times 2}\right)^2}}{2\left(\frac{5}{32}\right)^2} < 0 <$

$\gamma v_0 < \frac{32}{5}$ and $\gamma v_0 < 11$.Hence, $\sigma_{M1} = 0$ and $\sigma_{M2} = 32/5$ and for $0 < \gamma v_0 < 32/5$ the polity moves to **R2.** As these inequalities are strict, pairs (ρ, κ_0) in a neighborhood of $(4, 1)$ also yield a transition to **R2** for a measurable set of γv_0.

In sum, there exist $\sigma_L, \sigma_{M1}, \sigma_{M2}$ and σ_H defined above, $\sigma_L < \sigma_{M1} < \sigma_H$ and $\sigma_L < \sigma_{M2} < \sigma_H$, (simple algebra shows that $\sigma_{M1} < \sigma_H$ and $\sigma_{M2} < \sigma_H$) such that if $\gamma v_0 < \sigma_L$ the polity stays in **R3;** if $\sigma_{M1} < \gamma v_0 < \sigma_{M2}$ the polity moves from **R3** to **R2;** and if $\gamma v_0 > \sigma_H$ the polity moves from **R3** to **R1.**■

References

Acemoglu, D., S. Johnson, and J. Robinson. Institutions as a fundamental cause of long-run growth. Handbook of economic growth 1 (2005): 385-472.

Allen, R. C. (1997), Agriculture and the Origins of the State in Ancient Egypt. Explorations in Economic History 34, 135-54.

Bard, K. (1994), "The Egyptian Predynastic: A Review of the Evidence," *Journal of Field Archaeology* 21, 265–288.

Bates, R. (2001), Prosperity and violence: the political economy of development. New York: WW Norton.

Besley, T. and T. Persson (2011), Pillars of Prosperity: The Political Economics of Development Clusters: The Political Economics of Development Clusters. Princeton University Press.

Boix, C. (2015), Political order and inequality: Their foundations and their consequences for human welfare. Cambridge University Press.

Bradford, A. (2001), With Arrow, Sword, and Spear. A History of Warfare in the Ancient World. Westport, Connecticut and London: Praeger.

Carneiro, R. (1970). A Theory of the Origin of the State. *Science* 169 (3947): 733–738.

Carpenter, R. (1968), Discontinuity in Greek Civilization. New York: W. W. Norton & Co.

Childe, V. G. (1936), Man Makes Himself. London: Watts Publishers.

Chrissanthos, S. (2008), Warfare un the Ancient World. From the Bronze Age to the Fall of Rome. Westport, Connecticut and London: Praeger.

Claessen, H. and P. Skalnik (1978), The Early State. The Hague: Mouton Publishers.

Cline, E. (2014), 1177 BC - The Year Civilization Collapsed. Princeton, NJ: Princeton University Press.

Collier P. and A. Hoeffler (1998), On Economic Causes of Civil War, *Oxford Economic Papers* 50, 563-73.

Diamond, J. (1999), Guns, Germs, and Steel: The Fates of Human Societies. WW Norton & Company.

Di Giacomo, G. and G. Scardozzi (2012), Multitemporal High-Resolution Satellite Images for the Study and Monitoring of an Ancient Mesopotamian City and its Surrounding Landscape: The Case of Ur. *International Journal of Geophysics*,, Vol. 2012, Article ID 716296.

Drews, R. (1993), The End of the Bronze Age. Changes in Warfare and the Catastrophe ca. 1200BC. Princeton, NJ: Princeton University Press.

Dube, O. and J.F. Vargas (2006), Are All Resources Cursed? Coffee, Oil and Armed Conflict in Colombia, mimeo Harvard University.

Fried, M. (1960), On the evolution of social stratificiation and the state. Indianapolis: Bobbs-Merrill.

Fritz, V. (1997), Cities: Cities of the Bronze and Iron Age, in Eric M. Meyers (ed.), The Oxford Encyclopedia of Archeology in the Near East, Vol. II, Oxford and New York, Oxford University Press, 19-25.

Garfinkle, S. (2013), "Ancient Near Eastern City-States". In Peter F. Bang and Walter Scheidel (eds.), The Oxford Handbook of the State in the Ancient Near East and Mediterranean. New York, Oxford University Press: 94-119.

Gennaioli, N. and H-J. Voth (2015), State capacity and military conflict. *The Review of Economic Studies*

Gibson, McG. (1993), Nippur – Sacred City of Enlil, *Al-Rafidan: Journal of Western Asiatic Studies* vol. XIV.

Hawkes, J. (1974), Eridu, Iraq. Sumerian city circa 5000 B.C. Atlas of Ancient Archaeology. New York. McGraw-Hill Book Company.

Hirshleifer, J. (1991), The Paradox of Power. *Economics & Politics* 3(3), 177-200.

Hirshleifer, J. (2001). The dark side of the force: economic foundations of conflict theory. Cambridge: Cambridge University Press.

Johnson, A. and T. Earle (2000),. The evolution of human societies: from foraging group to agrarian state. Stanford University Press.

Jordan, J. (1931), Abhandlungen der Preussischen Akademie der Wissenschaften, Philosophisch-historische Klasse, Jahrgang 1930, Nr. 4.

Kaniewski D,, Van Campo E, Van Lerberghe K, Boiy T, Vansteenhuyse K, et al. (2011), The Sea Peoples, from Cuneiform Tablets to Carbon Dating. *PLoS ONE* 6(6): e20232. doi:10.1371/journal.pone.0020232.

Keeley, L. (1996), War before civilization: The myth of the peaceful savage. Oxford University Press.

Langutt, D., I. Finkelstein, and T. Litt. 2013. Climate and the Late Bronze Collapse: New Evidence from the Southern Levant, Tel Aviv, Vol. 40, 2013, 149–175.

Mann, M. (1986), The Sources of Social Power Vol. I: A history of power from the beginning to A.D. 1760. Cambridge University Press.

Mayshar, J., O. Moav, and Z. Neeman (2013). Geography, transparency and institutions. Mimeo University of Warwick.

Mayshar, J., O. Moav, Z. Neeman and L. Pascali (2015). Cereals, Appropriability and Hierarchy. CEPR Discussion Paper 10742.

McNeill, W. (1982), The Pursuit of Power: Technology, Armed Force, and Society since A.D. 1000. University of Chicago Press.

Morris, I. and J. Manning (2005), "The economic sociology of the ancient world," The Handbook of Economic Sociology, 2d. ed. Eds. Neil Smelser & Richard Swedberg., Princeton: Princeton University Press, 131-159.

North, D., and B. Weingast (1989). Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England. *The Journal of Economic History* 49(4), 803-832.

Olson, M. (2000). Power And Prosperity: Outgrowing Communist And Capitalist Dictatorships. Basic Books.

Powell, R. (2012), Persistent Fighting and Shifting Power, *American Journal of Political Science* (2012) 56(3), 620-37.

Powell, R. (2013), Monopolizing Violence and Consolidating Power, *Quarterly Journal of Economics* 128 (2), 807-859

Ross, M. (2003), The Natural Resource Curse: How Wealth Can Make You Poor, in Collier, P. and I. Bannon (eds.) *Natural Resources and Violent Conflict*. The World Bank.

Sánchez de la Sierra, R. (2014), On the Origin of States: Stationary Bandits and Taxation in Eastern Congo. Mimeo UC Berkeley.

Sandars, N.K. (1987), The Sea Peoples: Warriors of the ancient Mediterranean, Revised Edition. London: Thames and Hudson.

Schaeffer, C.F.A. (1948), Stratigraphie comparée et chronologie de l'Asie occidentale. London: Oxford University Press.

Singer, I. (1999), A Political History of Ugarit. In Handbook of Ugaritic Studies, W.G.E. Watson and N. Wyatt (eds.), 603–733. Leiden: Brill.

Spruyt, H. (1996). The sovereign state and its competitors: an analysis of systems change. Princeton University Press.

Tilly, Ch. (1975), The formation of national states in Western Europe. Vol. 8. Princeton Univ Press.

Tilly, Ch. (1992), Coercion, capital, and European states, AD 990-1992. Oxford: Blackwell.

Trigger, B. (2003), Understanding early civilizations. Cambridge University Press.

Van de Mieroop, M. (1997), The ancient Mesopotamian city. Oxford University Press.

Weber, M. (1978), Economy and society. Berkeley, CA: California University Press.

Weber, M. (2013) [1909], The agrarian sociology of ancient civilizations. London: Verso Books.

Wittfogel, K. A. (1957). Oriental Despotism. New Haven: Yale Univ. Press.

Yon, M. (1992), The End of the Kingdom of Ugarit, in Ward, W., and M. Joukowsky (eds.), The Crisis years: the 12th century BC: from beyond the Danube to the Tigris. Kendall Hunt Pub Co.

Zuckerman, S. (2007), Anatomy of a Destruction: Crisis Architecture, Termination Rituals and the Fall of Canaanite Hazor. *Journal of Mediterranean Archaeology* 20, 1: 3–32.