The Paradox of Civilization
Pre-Institutional Sources of Security and Prosperity

Ernesto Dal Bó Pablo Hernández-Lagos Sebastián Mazzuca

June 2019

Abstract

The rise of civilization involved economic surplus production (“prosperity”) and states that could provide surplus protection (“security”). But the security-prosperity combination poses a paradox: prosperity attracts predation, which discourages the investments that create prosperity. Drawing from the anthropological and historical literatures, we model the trade-offs facing a society on its path to civilization. We emphasize pre-institutional forces, especially the geographical environment, that shape growth and defense capabilities, and specify the conditions under which these capabilities help escape the civilizational paradox. We provide narrative and quantitative illustration of the model by analyzing the rise and fall of Old World Bronze Age civilizations, with special focus on Egypt and Sumer.

1 Introduction

The first civilizations emerged over 5,000 years ago as a cluster of fundamental transformations: urbanization, public infrastructure, writing, surplus food, and military organization. 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, civilization was driven by favorable agro-nomic endowments and the development of techniques for food production. For V. Gordon Childe (1936), the key precursors of civilizational take-off in Lower Mesopotamia (the “cradle of civilization”) were the fertile alluvial soil, abundant domesticable animals, and irrigation technology. Identical factors were emphasized for the rise of Egypt, the other pristine civilization to emerge shortly after Sumer. Both in Sumer 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 surplus above subsistence needs, it was not possible to fund the other components of the civilization cluster. However, surplus production was only a necessary condition for civilization, not a sufficient one. In fact, prosperity attracted trouble. The first food producers were often surrounded by pastoralist tribes ready to loot agricultural surpluses. The clash between sedentary, agricultural food producers and nomadic pastoralist raiders is a primordial conflict shaping the civilizational process. According to McNeill, “[s]oon after cities first arose ... the relatively enormous wealth that resulted from irrigation and plowing made such cities worthwhile objects of attack by armed outsiders” (1992: 85).

For many anthropologists, intergroup violence is part of the human condition and predated civilization (Keeley 1996), but civilization intensified conflict. As Mann puts it, “the greater the surplus generated, the more desirable it was to preying outsiders” (1986: 48). Since civilization involved the joint achievement of prosperity and security, its emergence is paradoxical to the extent that under normal circumstances prosperity and security were incompatible. A primitive society that pursued prosperity undermined its own security.

The conditions that helped solve the civilizational paradox are rare, as evidenced by the fact that, out of thousands of primitive societies, only a handful could develop independent civilizations, starting with Sumer in Lower Mesopotamia and Egypt. We develop a model to identify conditions for a viable civilization, seen as the joint emergence of prosperity and
security. To build the model, we select elements highlighted by archaeologists and historians. The strategic forces at play generate analytic restrictions with empirical content. We show that a high growth potential in food production is a necessary but not sufficient condition for civilization (a high income level does not guarantee civilization either). A civilization-friendly environment must provide at least one out of two additional advantages, and alleviate the security tension. The natural levels of defense must be high and keep in check the threats that come when growth is attempted, or, if natural defenses are low, the level of initial income must be high enough to finance artificial defenses.

Our model connects the Hobbesian normative political theory tradition viewing security as a precondition for prosperity with positive political theory. Hobbesian theory admits a weak form positing that some security is needed for prosperity, while a strong form posits that full security is needed. Our model rejects the strong form, by showing that partial rather than full security is sufficient for civilization. We characterize the exact conditions under which this finding is logically true. At the same time, our model supports the weak form of the claim, which has not been the focus of positive political theory linking early state formation to defense and growth fundamentals. Our theory also fills a gap in dominant anthropological visions of early civilizations, which have largely focused on productive capabilities at the expense of defensive ones, and, as a consequence, have failed to account for large blocs of historical data.

The civilizational paradox can be compared to the paradox present in the rise of the modern state in the post-Westphalia context. European rulers striving for a monopoly of violence reached unprecedented levels of power. However, in the process they undermined their own ability to commit to respecting private rights. Unstable property rights in turn hindered growth, and ultimately threatened to damage 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 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).

We argue that the paradox of the early state, as opposed to the modern one, was resolved not by formal political rules but by tangible assets, and the assets were relevant not only for production but also for defense. Pristine civilizations emerged in areas with favorable
conditions for growth in food production, and the salient human contributions to civilization were productive and defense investments. Two engineering accomplishments are the birthmark of pristine civilizations, water management and perimetric walls. They were pillars of civilization in Mesopotamia and the Levant, the Yellow River, as well as the Indus Valley. Each public good had an unambiguous mission: surplus production and surplus protection. The prominence of the two types of public work reflects the centrality of production, but also protection, to civilization.

We contrast the predictions of our model with historical narratives and a quantification exercise focused on Egypt and Sumer. In addition to their historical prominence as the earliest civilizations, Egypt and Sumer provide evidence that the potential for surplus emphasized by archaeologists was only half of the civilization story. Blessed with high potential for growth, both locations needed to resolve the problem of surplus protection, which occurred in two contrasting ways: defense could be natural as in Egypt, or man-made as in Sumer.

Our pre-institutional theory on the joint achievement of security and prosperity sheds light on the problem of state formation more generally. A broader goal is to understand developments where a potentially prosperous region, being surrounded by predatory threats, flourishes, or alternatively falls in the traps of security-preserving stagnation or self-defeating prosperity. These dangers, as well as the role of defense that we emphasize throughout the paper, likely operated in many historical contexts. Thucydides, the first systematic Western historian, characterized inhabitants in pre-Hellenic Greece as “cultivating no more of their territory than the exigencies of life required, destitute of capital, never planting their land, [f]or they could not tell when an invader might not come and take it all away, and when he did come they had no walls to stop him ... consequently it neither built large cities nor attained any other form of greatness.” (Thucydides 431BC Book 1 [Warner, 1971: 35-36]).

Many societies have suffered the tension between prosperity and security. Examples include medieval Eurasian proto-cities facing barbarian invasions from the northern steppes; 19th century city-ports in the Americas challenged by hostile native tribes, bandits and rural warlords; and contemporary failed states of Sub-Saharan Africa and the Middle East, where international economic aid, if not coupled with military buildup, can increase the voracity of predatory actors. Echoing concerns in history and anthropology about the reversibility of gains in social complexity, our theory also provides an account for civilization collapse,
seen as economic and military reversal in societies that had achieved prosperity and security. For illustration, we use the model to provide new insight into the end of the Bronze Age, a much-debated process in which East Mediterranean civilization centers collapsed one after the other within a relative short period, ushering in the first “dark ages” in the historical record.

**Overview of the model and historical illustrations** In our model, an “incumbent” may invest and grow future income, which would lead to “prosperity,” but faces potential attacks by a “challenger.” The possibility of attacks may induce the incumbent to spend resources in consumption and defense instead. Three key parameters govern tradeoffs: initial income, and the respective rates at which income can be turned into defense (defense capability) and future income (growth capability). Productive investment increases future income, but it also attracts stronger predation. The result is a tradeoff between investment-led growth and security. If sufficient defense can be financed, the challenger is deterred (“security” is attained), investments are safe, and growth follows. Defense, however, costs current consumption, and the price of safe investments may be too high for growth to be desirable.

The key question is whether some combination of parameter values allows for both security and investment to occur, yielding a civilizational breakthrough. Previous work, discussed in the next section, has analyzed related problems; but the choice between consumption, productive investment, and defense over multiple periods is a natural problem that, to the best of our knowledge, has not been addressed.

In the first part of our analysis the incumbent’s defense capability is exogenous, and in the second part defense capability can be improved. Both parts help account for varieties in the rise of civilizations. They also provide a framework to think about short-run scenarios (when defense capabilities are fixed), and long-run scenarios (when defense capabilities are variable).

The analysis in the first part shows that when both defense and growth capabilities are low, neither prosperity nor security are possible, and societies remain economically stagnant and mired in conflict, a situation documented by Keeley (1996), which echoes the Hobbesian “state of nature.”

---

1The 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
If growth capability is high relative to defense capability, prosperity becomes possible even in the face of attacks. Such prosperity may be short-lived for attacks may provoke the collapse of the incumbent. While anti-Hobbesian, the possibility of growth despite predation is consistent with a common pattern in history, like the Chinese and Saxon settlements under the threat of Mongolian and Viking looting in the 10th century. Lastly, when both defense and growth capabilities are high and balanced, the incumbent can grow and also deter predators. The latter two cases explain the emergence of civilizations. Civilizations emerge when returns to productive investment are high enough to allow for growth, and when challengers are deterred or defeated with high probability.

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

The rise of civilization in Southern Mesopotamia poses a challenge to the baseline model, however, because in contrast to Egypt, the Sumerian settlements lacked natural protection and faced challenges from pastoral tribes. How could the Sumerian city-states ever emerge? We show that when initial income is high enough, the incumbent can fund its way out of the parametric region without security or prosperity, and move into a region with high levels of both, a transition that is easier when growth capabilities are also high. The result is not obvious. Investments both productive and defensive incite immediate predation, suggesting that society might prefer to avoid both. Investment in defense occurs in equilibrium due to an intertemporal complementarity with productive investment. If enough defense capability is built today, enhanced security will increase effective returns to (and thereby induce) productive investments tomorrow. The larger intertemporal pie of tomorrow justifies today’s defense investments.

The broad theoretical conclusion is that favorable conditions such as high initial income and growth capabilities cannot individually produce civilization. In a world where security is a concern, growth capabilities need to interact with high natural defenses, or high initial income to fund artificial defenses. We take these predictions to the historical record through a 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.
quantitative exercise. We divide the Old World in grid cells and use historical and geographic information to proxy the income and capabilities in each location. We project grid cells to empirically render the theoretical parameter space, and then we locate in it the ancient Egyptian and Sumerian cities in Models’ki’s (1999, 2003) dataset.

The quantitative exercise matches the message of the theory and the historical narratives. Egyptian cities belong in a statistically rare area boasting high levels of both growth and natural defense capabilities while Sumerian cities display lower defense capabilities. However, Sumerian cities were disproportionally located in areas that satisfy the joint requirement of high initial income and high growth capabilities. Both Egyptian and Sumerian locations lie in a statistically sparse part of the parameter space, showing that not many other locations were similarly equipped for success (i.e., there are virtually no “false positives”). One concern is that other civilizations may have arisen in areas that the model predicts to be failures (“false negatives”). But only two other pristine Bronze Age Civilizations appeared in the Old World, in China and the Indus Valley, and they lie in a similar part of the parametric space as Sumer.

2 Related Literature

The problem of order and prosperity is a foundational one in the social sciences. To motivate and develop our model, we build on the vast literature on state formation in Political Science and Anthropology, and on theories of conflict and state capacity in Economics.

2.1 Ancient States and Modern State Formation

Founding fathers of archaeology like Childe (1936) emphasized developments in agriculture and cities in the civilizational process, focusing on the means and relations of production while abstracting from the accompanying innovations in military protection. More recent archaeological research has 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 have explicitly connected food production with protection needs (Michael
Mann 1986: 48). 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.

Our approach builds on, but departs from, historical accounts that emphasize the geographic sources of economic prosperity. Some authors emphasize the availability of domesticable plants and animals to explain why some regions generated surpluses while others did not (e.g., Diamond 1997). These approaches contribute a necessary building block for understanding civilization. However, a purely geographic account is incomplete, for it misses the role of incentives and strategic action. Our approach incorporates both strategic actors and geographic factors such as the potential for food production or protective terrain.

We abstract from some aspects present in anthropological theories of the state, such as social stratification and political domination (Fried 1960, 728; Carneiro 1970). We do so not because we think stratification is unimportant, but because it helps to focus attention on the incumbent-challenger interaction. Yet, our results provide new insights into key facts of state formation as stylized by classical work. According to the “circumscription theory” (Carneiro 1970), states originated as growing populations preferred to submit to the authority of inchoate rulers in fertile areas over migrating to less productive land.

The Nile valley, surrounded by deserts, is a good candidate location for circumscription theory. Our model generates a similar empirical implication. However, the implication is not driven by exploitation but by the fact that Egypt’s extremely low-quality surrounding land was a protective buffer against challengers. In contrast to Carneiro’s theory, our model does not appeal to population pressure, an assumption that has been challenged by other scholars (Allen 1997).

It is customary in the social sciences to view the state as the monopoly of violence. Adapting from Max Weber, we define the state not in binary terms but as a matter of degrees (Weber 1978: ch. I, s. 16), so that state formation involves higher degrees of protection from attacks. We focus on the state as “sovereignty,” defense from threats, and abstract from “rulership,” the creation of a political hierarchy that divides rulers from ruled within a society. The exclusion of rulership from our model produces a minimalist view of early civilizations, which usefully redefines them as the combination of surplus producing

---

2 For a review of anthropological theories of early states see for example Claessen and Skalnik (1978).
3 See Schönholzer (2017) for a recent investigation of Carneiro’s hypothesis.
economies and effective surplus protection through statehood.

Our work is related both to the long lineage of theories of state formation (Tilly 1975, 1992, Spruyt 1996) and to the vibrant field of analysis of the political sources of prosperity (North and Weingast 1989; Olson 1993, 2000, Bates 2001; Acemoglu et al. 2005, 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-state forces.

We share with Boix (2015) an interest in mechanisms of state formation extending back into prehistoric times, as well as in “hard” causes related to the physical environment. Although Boix finds sources of pre-institutional cooperation under conditions of anarchy (absence of state), he focuses on state formation as the selection of either republican or monarchic institutional settings. By contrast, we focus on the conditions for state formation that allow for investment and security before political institutions become central.

Our work has important complementarities with that by Mayshar, Moav and Neeman (2017), 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 (2017) use a principal-agent model to show how monitoring capabilities shape the extent of political centralization, and account for contrasting trajectories in Sumer 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 emphasize crop appropriability. They equate the state with the political hierarchy that results from appropriability and assume it results in the full prevention of conflict. We abstract from appropriability and internal hierarchy, and investigate whether it is true that conflict can be eliminated.

2.2 Conflict and state capacity

Our model can be viewed as a development within the vast literature on conflict, in particular the research on the tradeoff between “guns” and “butter” (Garfinkel and Skaperdas 2007
provide the review we omit here for the sake of space). The closest example is the model of Grossman and Kim (1995), in which agents choose between contestation and production; the agents, like our incumbent, are concerned with deterrence. Like Grossman and Kim, our model considers a tradeoff between consumption and security. Yet, a key new element in our model is a separate choice to invest. Therefore, in contrast to Grossman and Kim, our model provides the framework to analyze other tradeoffs, pitting investment against both consumption and security. Finally, their model assumes only interior solutions, while we investigate all interior and corner solutions.

A more recent literature studies investments in state capacity by a ruler who may lose control of the polity to a competing faction (Besley and Persson 2011), or to a foreign power (Gennaioli and Voth 2015). A difference with Besley and Persson’s model is that our incumbent controls defense and the economy directly, while their incumbent controls different forms of state capacity. As importantly, investments in our model augment the virulence of challenges. Gennaioli and Voth (2015) formalize and test Tilly’s (1992) argument that modern European states were formed as a byproduct of the competitive pressures of military conflict, which created the need to centralize fiscal control. Gennaioli and Voth (2015) model the problem of a ruler choosing the level of fiscal centralization. In their setup, centralization contains elements of both defense and investment: higher centralization today generates resources that, if conflict were to follow, become the funding for the war effort tomorrow, and the disputable loot the day after tomorrow. Crucially, their ruler makes a single choice, while our incumbent selects investments in production and defense separately. This separation is important for our goal of characterizing a tension between growth and security in a robust manner. If we tied together the defense and investment decisions, we might observe investment in equilibrium due to the implicit defense component of a higher future income. But such an effect would not arise if the incumbent could make separate decisions. Our paper is also related to models of state consolidation (Powell 2012, 2013); the key difference is that in our model consolidation is studied in relation to investment and growth.
3 Theory

3.1 Basic Model

3.1.1 Setup

Players An “incumbent” controls a productive asset that yields a nonstorable income flow $v_t > 0$ in each of two periods, $t = 1, 2$. The asset can be any bundle of productive resources, including land, a port, and people. The initial level $v_1$ tracks properties of the environment (e.g., climate, quality of the soil) that affect productivity. The incumbent faces a “challenger” in each period, who is interested in gaining control of the asset. The incumbent and challenger in our model can be taken to be representative agents of their respective groups, perfectly benevolent rulers acting on their behalf, or perfectly extractive rulers who are residual claimants. This and many other matters of interpretation, modeling choices, robustness, and extensions are developed in the Online Appendix. We relegate to the Online Appendix all proofs, which are labor-intensive but not of substantive importance.

Actions, resources, technology In each period $t$ the incumbent can spend $v_t$ on consumption, productive investment, or mobilizing resources for defense. One unit of productive investment $i_t$ costs one unit of income and it adds $\rho > 1$ units to the future yield of the asset. That is, income evolves according to the relation $v_{t+1} = v_t + \rho i_t$; we abstract from depreciation and discounting for simplicity. $\rho$ captures anything that affects the returns to investments in the asset; like $v_1$, $\rho$ could reflect conditions of the physical environment, and economic factors such as the price of goods sold.

The effectiveness of the incumbent’s defense (or “army”) is denoted $a_t$ and such an army costs the incumbent 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 expense $\frac{a_t}{\kappa_t}$. The parameter $\kappa_t$ captures anything that affects the costs for the incumbent of producing defense, such as a surrounding desert, better military

\footnote{Given that (with linear preferences) investment is never worthwhile if $\rho < 1$, failure to obtain it in equilibrium is inevitable and uninteresting. Hence our assumption $\rho > 1$ which makes investment possible (though not inevitable, due to insecurity).}

\footnote{If the value of what the incumbent produces follows a standard price $\times$ quantity formulation we can write $v_1 = pq$, and $v_2 = pq + \rho'pi = v_1 + \rho i$, where $q$ and $i$ are physical units. Then, changes in $p$ cause changes in both $v_1$ and $\rho$. Changes in the baseline physical capacity of production $q$ are captured through changes in $v_1$, and changes in the physical returns to investment as changes in $\rho'$ and therefore $\rho$.}
technology, or expertise. We fix $\kappa_t = \kappa_1$ in our baseline analysis and later endogenize $\kappa_t$ in subsection 3.2.

**Timing** In each period the incumbent acts as a Stackelberg leader, moving first to choose $a_t$ and $i_t$. After observing $(a_t, i_t)$ the challenger selects its own conflict effort $b_t$. If $b_t > 0$, then there is conflict at the end of period $t$. The winner appropriates the asset, and hence the income it generates in the next period.\(^6\) 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). If the incumbent does not attack or is defeated, it obtains a payoff of zero. If the challenger selects $b_t = 0$ we say the incumbent has successfully deterred the challenger, and this lack of challenge results in full security.

**Payoffs** Both challenger and incumbent are risk neutral and care linearly about consumption, which equals income net of costs (of defense for both players, and of investment for the incumbent). The incumbent chooses $a_t$ and $i_t$ to maximize the value of his expected intertemporal consumption $V_t = v_t - \frac{a_t}{\kappa_t} - i_t + \frac{a_t}{a_t + b_t} V_{t+1}$, while observing the budget constraint (or non-negative consumption condition), $v_t - \frac{a_t}{\kappa_t} - i_t \geq 0$. The challenger chooses $b_t$ to maximize the value of her own expected intertemporal consumption $\frac{b_t}{a_t + b_t} V_{t+1} - b_t$.

**Solution concept** We solve for a Subgame Perfect Nash Equilibrium by backward induction.

### 3.1.2 Solution

**Second period** The rewards from success in conflict accrue one period later, so the challenger does not fight in the second and last period, leaving $b^*_2 = 0$ (asterisks denote equilibrium choices). Anticipating this, the incumbent chooses $a^*_2 = 0$. Since the proceeds from productive investment only materialize in the next period, the incumbent selects $i^*_2 = 0$ yielding $V_2 = v_2 = v_1 + \rho i_1$.

**First period** The challenger observes the pair $(a_1, i_1)$ and chooses $b_1$ to maximize $\frac{b_1}{a_1 + b_1} v_2 - b_1$. Since the first order condition is $\frac{a_1}{(a_1 + b_1)^2} v_2 = 1$, the best response function of the challenger is, $b_1(a_1, v_2) = \sqrt{a_1 v_2} - a_1$ if $a_1 < v_2$ and zero otherwise.

Since $v_2 = v_1 + \rho i_1$, the best response $b_1(a_1, v_2)$ exhibits a key trade-off of the model.

---

\(^6\)In the two period model it is immaterial whether expropriation involves the income flow or the asset itself. Both cases were observed historically: intermittent raids, and invasion with “replacement,” such as when Sargon of Akkad took over the Sumerian city-states, or the Mongols invaded China.
Productive investments $i_1$ raise the value of the productive asset since $\rho > 1$. Thus, conditional on maintaining control of the asset, investment is a good idea for the incumbent. However, investment raises the incentives of the challenger to arm itself since it makes it more attractive to become the incumbent (formally, $\frac{dn}{di_1} > 0$ if $a_1 < v_2$). In sum, while productive investments increase the value of the asset, they 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.\(^7\) The resulting question is: are there any parameter values $v_1$, $\kappa_1$, and $\rho$ that map into security and prosperity?

The incumbent maximizes his initial expected intertemporal consumption $V_1$ subject to the budget constraint (BC) $v_1 - i_1 - \frac{a_1}{\kappa_1} \geq 0$, and a deterrence constraint (DC) $a_1 \leq v_2$ stemming from the facts that the challenger will not fight if $a_1 \geq v_2$, and the incumbent does not arm beyond the point that attains deterrence. The incumbent’s problem in period 1 can then be written as,

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

subject to,

$$v_1 - \frac{a_1}{\kappa_1} - i_1 \geq 0 \quad (BC)$$

$$v_1 + \rho i_1 - a_1 \geq 0 \quad (DC)$$

$$a_1 \geq 0 \quad (4)$$

$$i_1 \geq 0 \quad (5)$$

The Lagrangian, which expresses the expected utility $V_1$ of the incumbent, is:

\(^7\)The civilizational paradox, involving as it does the incentives of a challenger, is related to, but differs from, Hirshleifer’s (1991) paradox of power. Hirshleifer’s paradox consists of the fact that the poorer contender can end up better off. We instead use the term “paradox” to denote a tension: investments leading to prosperity reduce security and therefore the motivation to bring about that prosperity.
There is a unique equilibrium, which yields a partition of the parameter space. The following proposition summarizes the solution.

**Proposition 1** There is a unique equilibrium, which yields a partition of the parameter space \((\kappa_1, \rho, v_1)\) into four regions:

- **Stagnant Insecurity** (Region SI): \(\{(\kappa_1, \rho, v_1) | 2 > \kappa_1, \rho < 4/\kappa_1\}\)
  
  Solution: \(a_1^* = v_1 \left(\frac{1}{\kappa_1}\right)^2, i_1^* = 0, V_1 = v_1 \left(1 + \frac{2}{\kappa_1}\right)\)

- **Stagnant Security** (Region SS): \(\{(\kappa_1, \rho, v_1) | \kappa_1 > 2, \rho < \kappa_1/(\kappa_1 - 1)\}\)
  
  Solution: \(a_1^* = v_1, i_1^* = 0, V_1 = v_1 \left(2 - \frac{1}{\kappa_1}\right)\).

- **Prosperity Despite Insecurity** (Region PDI): \(\{(\kappa_1, \rho, v_1) | \rho > \kappa_1, \rho > 4/\kappa_1\}\)
  
  Solution: \(a_1^* = \frac{\kappa_1 v_1}{2} \left(1 + \frac{1}{\rho}\right), i_1^* = \frac{v_1}{2} \left(1 - \frac{1}{\rho}\right), V_1 = \frac{v_1}{2} \left(1 + \frac{1}{\rho}\right) \sqrt{\rho \kappa_1}\)

- **Prosperous Security** (Region PS): \(\{(\kappa_1, \rho, v_1) | \kappa_1 > 2, \rho > \kappa_1/(\kappa_1 - 1), \kappa_1 > 1\}\)
  
  Solution: \(a_1^* = v_1 \frac{\kappa_1 (1 + \rho)}{\kappa_1 + \rho}, i_1^* = v_1 \frac{\kappa_1 (1 - \rho)}{(\kappa_1 + \rho)}, V_1 = v_1 \frac{\kappa_1 (1 + \rho)}{(\kappa_1 + \rho)}\)

Panel (a) in Figure 1 contains a graphical representation of the solution. We restrict attention to the bidimensional space \((\kappa_1, \rho)\) because the shape of the four regions is invariant in \(v_1\).

One implication of this invariance is that income advantage does not ensure full

---

8Initial income does not affect the four regions because (in equilibrium) it does not affect the incumbent’s incentives at the margin. These incentives depend on a ratio involving the prize under dispute (future income)
security. Only a technological advantage at defense does.

A salient feature of the solution is that all four combinations of security and prosperity can be observed depending on the values of defense and growth capabilities ($\kappa_1, \rho$). For low values of both parameters, the incumbent will be stuck in a situation of stagnation and conflict in SI. In this region the prospect of conflict lowers the net return to investment, preventing growth. If defense capability $\kappa_1$ is higher but growth capability $\rho$ is still low, the incumbent will be in SS, where the challenger is deterred but there is no investment. In this region growth is foreclosed not by existing but by potential conflict: investment would foster challenge and raise the costs of maintaining deterrence. If growth capability is relatively high and defense capability relatively low—i.e., in PDI—growth occurs despite the fact that full security is not attained. The high growth capability makes the incumbent willing to invest even if the prospect of a larger income makes the challenger more aggressive.\footnote{It may surprise that growth increases with $\rho$ in PDI. When $\rho$ goes to infinity, the challenger grows infinitely aggressive, against an incumbent who is resource-constrained and cannot keep up, which suggests investments are almost sure to be lost and should go to zero. However, the relative speed at which gross returns grow relative to the challenger’s threat justify positive investments even at the limit. In PDI growth is unresponsive to defense capability because any increase in $\kappa_1$ 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.} If, starting from PDI or SS, defense capability $\kappa_1$ were to become sufficiently higher, the incumbent would enter PS, featuring both investment and deterrence. Note that strict military superiority ($\kappa_1 > 2$) by the incumbent is needed for complete security (the challenger has a military capacity of 1).

Figure 1(a) highlights “extensive margin” variation, as there is either no growth or some growth, and no insecurity or some insecurity. But the equilibrium magnitudes of growth and security vary continuously. The other panels in Figure 1 display contour plots of relevant equilibrium magnitudes. The continuous lines within each region represent level curves, and the lighter shades of color represent higher values of the respective magnitude. Figure 1(b) shows that since the army $a_1$ of the incumbent increases as defense capability is higher, so does security, proxied by the probability that the incumbent will prevail. This probability is 1 in SS and PS, and it decreases in PDI as defense capability goes down or growth capability goes up (as this fires up the challenger). The areas in PDI that are sufficiently close to PS display very high levels of security (approaching full security where PDI meets PS). But civilization requires more than security; it also requires the creation of surplus,

and arming expense (as seen in conditions 20 and 21 in the Online Appendix). Both the prize $v_2$ and arming $a_1$ are proportional to initial income in equilibrium, and income changes become irrelevant.
Figure 1: Equilibrium with exogenous defense capability. Assumption: $v_1 = 1$. 

- **a)** Regions
- **b)** Probability the incumbent prevails
- **c)** Growth
- **d)** Expected continuation value
which in our model amounts to growth \((v_2 - v_1 = i_1 \rho)\). Figure 1(c) shows how there is no growth in SI and SS (since there is no investment) and that there is growth in PDI and PS. Growth increases with returns \(\rho\) and in PS it also increases with defense capability, as a higher defense capability lowers the costs of arming and releases resources for investment.

The panels (b) and (c) in Figure 1 show that, all else equal, increases in defense capability increase both security and prosperity, while increases in growth capability help growth but undermine security. Panel (d) of Figure 1 shows the combination of growth and security, as given by the expected continuation value perceived by the incumbent in period 1. This is the value \(EV_2 = \frac{a_t^2}{a_t + b_t} v_2^*\), which involves investment-led future income \(v_2^*\) and the probability \(\frac{a_t^2}{a_t + b_t}\) that the incumbent prevails in equilibrium.

### 3.2 Endogenous defense capability and the transition to prosperity and security

In our basic model, an incumbent with low defense capabilities cannot attain prosperity and security. But some societies succeeded despite an initial situation of relative insecurity. In fact, the mark of many civilizations was the erection of defensive structures to enhance the effectiveness of defense efforts. We now extend the model to allow for the endogenous expansion of defense capabilities.

#### 3.2.1 Setup

We consider the arrival of a different challenger each period who aims to replace the incumbent and inherit its productive and defense capabilities. Like in our baseline model, in each period the incumbent selects productive investment \(i_t\) and defense \(a_t\); but the incumbent can now spend resources \(m_t\) in one period to increase its defense capability in the next, and move horizontally in the \((\kappa_t + 1, \rho)\) space. Defense capabilities evolve according to the relation \(\kappa_{t+1} = \kappa_t + m_t\), and the incumbent’s budget constraint in period \(t\) becomes \(v_t - m_t - \frac{a_t}{\kappa_t} - i_t \geq 0\).

Now we need to consider three periods, 0, 1 and 2. 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 is only relevant in period 0. All other aspects of the interaction between challenger and incumbent remain as before. After observing
the challenger selects \( b_t \). If \( b_t = 0 \), the incumbent retains his position in the next period. If \( b_t > 0 \), then there is war at the end of period \( t \). The winner becomes the incumbent in the next period, and faces a new challenger then.

To make things as stark as possible, consider an incumbent that, barring investments in defense capability, will find itself in region \( \text{SI} \) in period 1, by imposing the following,

Assumption 1 \( \kappa_0 \rho (1 + \rho) < 4 \) and \( \kappa_0 < 2 \).

We ask whether a society with high enough initial productivity \( v_0 \) can transition into prosperity and security despite a low initial defense capability \( \kappa_0 \). More formally: can investments in defense capabilities land the incumbent in \( \text{PS} \) or \( \text{PDI} \) in \( t = 1 \)? The answer is not obvious because investments in defense capability become useful only in the future, and since they can be appropriated by the challenger, they incite predation now.

3.2.2 Solution

As before, we solve the model through backward induction. Actions in periods 1 and 2 follow the logic in our baseline model. Given the initial parameters \( (v_0, \kappa_0, \rho) \), the choices \( (i_0, m_0) \) of the incumbent in period 0 generate a continuation value of incumbency \( V_1(i_0, m_0) = (v_0 + \rho i_0) \times S(m_0) \), where \( S(m_0) \) is a function (detailed in the Online Appendix C) that captures changes in payoffs in period 1 depending on what region \( \text{PS}, \text{PDI}, \text{SI} \) or \( \text{SS} \) of the parametric space \( (\kappa_1, \rho) \) the incumbent lands in. Given the continuation value \( V_1(i_0, m_0) \), 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 fight. Using the same logic as in the baseline model, the challenger’s best response function is \( b_0(a_0, i_0, m_0) = \sqrt{a_0 V_1(i_0, m_0)} - a_0 \) if \( a_0 < V_1(i_0, m_0) \) and zero otherwise. Since the value of incumbency \( V_1(i_0, m_0) \) in period 1 is increasing in investments both productive \( i_0 \) and defensive \( m_0 \), investments of both kinds incentivize challenges and represent non-trivial decisions.

Given the challenger’s best response function, the incumbent chooses \( a_0, i_0 \) and \( m_0 \) to maximize his expected utility,

\[
\max_{a_0, i_0, m_0 \geq 0} v_0 - m_0 - \frac{a_0}{\kappa_0} - i_0 + \frac{a_0}{a_0 + b_0(a_0, V_1(i_0, m_0))} V_1(i_0, m_0)
\]
subject to the nonnegativity constraints $a_0 \geq 0, i_0 \geq 0, m_0 \geq 0$, the budget constraint $v_0 - m_0 - \frac{a_0}{\kappa_0} - i_0 \geq 0$ and the deterrence constraint $(v_0 + \rho i_0) S(m_0) - a_0 \geq 0$.

We now establish,

**Proposition 2** Suppose that $\rho > 1$ and Assumption 1 holds. Then,

1. If $v_0$ is low enough, the incumbent is trapped in $\text{SI}$. More formally, there exists $v(\kappa_0, \rho)$ such that if $v_0 < v(\kappa_0, \rho)$ investments in defense capability are zero ($m_0^* = 0$) and in $t = 1$ the incumbent remains in $\text{SI}$;

2. (i) If $v_0$ is high enough, then in $t = 1$ the incumbent will land somewhere in $\text{PS} \cup \text{PDI}$ and enjoy increased security and growth. More formally, there exists $\bar{v}(\kappa_0, \rho)$ such that if $v_0 > \bar{v}(\kappa_0, \rho)$ the incumbent makes positive investments in defense capability ($m_0^* > 0$) in $t = 0$ to land somewhere in $\text{PS} \cup \text{PDI}$ in $t = 1$; (ii) If $v_0 > \bar{v}(\kappa_0, \rho)$, higher $v_0$ yields (weakly) higher levels of growth and security, and a strictly higher continuation value $EV_2$.

3. Consider any point in $\text{PS} \cup \text{PDI}$ that can be reached in $t = 1$ by making an investment $m_0^*$ in defense capability. The initial income $v_0$ that makes such investment optimal is decreasing in $\rho$, implying that, given a distribution of initial incomes $v_0$, more income levels allow the incumbent a transition into $\text{PS} \cup \text{PDI}$ when $\rho$ is high than when it is low.

The intuition for the result is as follows. Investments both productive and defensive augment predation, and become discouraged for $v_0$ low enough. It is preferable to consume in the present rather than risk any investment. This holds even if $v_0$ is sufficient to finance defenses that would allow the incumbent to exit $\text{SI}$ – the key to the result is not just financial feasibility. When $v_0$ is high enough, a complementarity arises between defense and productive investments. A large enough investment in defense at $t = 0$ increases security in $t = 1$ so much that the effective rate of return makes productive investment in $t = 1$ incentive-compatible. In other words, a large enough investment in defense grows the intertemporal pie, and makes the risk of additional immediate predation worth taking.

### 3.3 Predictions

We use the theoretical propositions to derive predictions about the connection between parameter values and civilizational success. Consider first the world with exogenous defense capabilities. If civilization is the joint attainment of growth and a substantial degree of
security, SI and SS, which feature no growth, are incompatible with civilization. Civilization requires that defense and growth capabilities \((\kappa_1, \rho)\) be high enough so that the polity can be somewhere in PS or PDI and attain growth with some degree of security.

However, areas panels (b) and (c) of Figure 1 show that the areas PS or PDI contain locations with high growth and low security and vice versa. The right proxy for the likelihood of civilization is one that combines security and growth outcomes in a measure of “expected growth,” given by the expected continuation value \(EV_2\) in panel (d) of Figure 1. In areas where there is growth, \(EV_2\) increases with both defense and growth capabilities (in PS, \(EV_2 = v_1 \frac{\kappa_1 (1 + \rho)}{\kappa_1 + \rho}\), and in PDI, \(EV_2 = \frac{v_1 (1 + \rho)}{2} \sqrt{\frac{\kappa_1}{\rho}}\). Thus, higher capabilities \((\kappa_1, \rho)\) make civilization more likely, both because it is more likely that the polity will be in regions PS or PDI and because conditional on being in those regions, higher values of \((\kappa_1, \rho)\) increase prosperity and security. Initial income \(v_1\) also raises \(EV_2\), but notably, this is conditional on growth and defense capabilities being high enough that the polity finds itself in areas where growth and security are possible in the first place. Summarizing,

**Remark 1** In a world with exogenous defense capabilities, higher values of both capabilities \((\kappa_1, \rho)\) make civilization more likely. Conditional on relatively high values of \((\kappa_1, \rho)\), civilization is more likely for higher values of initial income \(v_1\); initial income is otherwise irrelevant.

When the incumbent can strengthen defense capabilities, high natural defenses are not necessary for civilization. But such investments are costly, so if initial income is low, stagnation and insecurity are unavoidable (point 1 in proposition 2). In contrast, if initial income is high enough, stronger defense capabilities will help improve security and attain prosperity (point 2). This effect should arise mainly in areas where growth capabilities are relatively high (point 3). These observations immediately yield the following,

**Remark 2** If (exogenous) defense capabilities are low, locations associated with civilizational success should have relatively high initial income and growth capabilities, while at least one of these parameters should be low among civilizational failures.

These remarks lay out the patterns that we seek to illustrate in the next section. The predictions differ from the blunter observation that it is good if every parameter is high. The key point is that high initial income and growth capabilities are not individually sufficient.
for civilization. High growth capabilities must be paired either with high natural defenses or high income.

4 Historical illustrations

We illustrate the model through qualitative case studies on the earliest two civilizations, Sumer and Egypt. These cases embody the historical narratives that motivate the parameters of our model, defense capabilities \( \kappa_1 \), growth capabilities \( \rho \), and initial income \( v_1 \). The historical cases are further characterized through a quantitative exercise linking the theoretical parameters to measurable environmental conditions. We then check if civilization, represented by ancient cities in Egypt and Sumer, tended to arise in parts of the parametric space that the model highlights as conducive to civilization.

Rather than select cities on an ad hoc basis, we rely on Modelski’s (1999) data on ancient cities, which was digitized by Reba et al. (2016). This dataset contains cities that are believed to have surpassed 10,000 inhabitants. We consider Sumerian and Egyptian cities meeting that criterion at any time until 1,200BC, the estimated date for the end of the Bronze Age in the Eastern Mediterranean.\(^{10}\) Modelski raises the population cutoff to 100,000 after 1000BC, and 1 million after 1000AD which precludes consistent attention to the New World. Table 1 lists the cities in our sample.

We locate cities in a spatial grid with cells of 1/5th of a degree side length, roughly 22km at the Equator. The construction of the empirical proxies relies on various datasets. We describe below how we compute each empirical measure. Due to space constraints, we restrict attention to essential aspects, and offer numerous details about sources, robustness, and alternative choices in the Online Appendix.

4.1 Egypt and Sumer

Southern Mesopotamia gave rise to the first major civilization, based on a cluster of city-states. The Egyptian civilization emerged slightly later, but its development after the adoption of agriculture was faster: in only about 1,000 years after the adoption of farming, a

\(^{10}\)These criteria have pros and cons. The high population cutoff restricts attention to cities that were undoubtedly marks of civilization, but causes us to miss others that did not grow as much. The timeline balances a focus on earliest cities against more observations.
Table 1: Ancient cities

<table>
<thead>
<tr>
<th>City</th>
<th>1st Observation</th>
<th>City</th>
<th>1st Observation</th>
<th>City</th>
<th>1st Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eridu</td>
<td>3700 BC</td>
<td>Memphis</td>
<td>2500 BC</td>
<td>Harappa</td>
<td>2500 BC</td>
</tr>
<tr>
<td>Larak</td>
<td>3500 BC</td>
<td>Heliopolis</td>
<td>2400 BC</td>
<td>Mohenjodaro</td>
<td>2500 BC</td>
</tr>
<tr>
<td>Uruk</td>
<td>3500 BC</td>
<td>Elephantine</td>
<td>1900 BC</td>
<td>Rakigarhi</td>
<td>2300 BC</td>
</tr>
<tr>
<td>Nippur</td>
<td>3000 BC</td>
<td>Thebes</td>
<td>1800 BC</td>
<td>Dholavira</td>
<td>2300 BC</td>
</tr>
<tr>
<td>Zabalam</td>
<td>2800 BC</td>
<td>Hermopolis</td>
<td>1300 BC</td>
<td>Erlitou</td>
<td>1700 BC</td>
</tr>
<tr>
<td>Adab</td>
<td>2800 BC</td>
<td></td>
<td></td>
<td>Bo (Yanshi)</td>
<td>1600 BC</td>
</tr>
<tr>
<td>Ur</td>
<td>2800 BC</td>
<td></td>
<td></td>
<td>Ao (Zhengzho)</td>
<td>1400 BC</td>
</tr>
<tr>
<td>Larsa</td>
<td>2800 BC</td>
<td></td>
<td></td>
<td>Yin</td>
<td>1300 BC</td>
</tr>
<tr>
<td>Umma</td>
<td>2800 BC</td>
<td></td>
<td></td>
<td>Sanxingdui</td>
<td>1200 BC</td>
</tr>
<tr>
<td>Kish</td>
<td>2800 BC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shuruppak</td>
<td>2800 BC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lagash</td>
<td>2500 BC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girsu</td>
<td>2400 BC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isin</td>
<td>2000 BC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eshmunna</td>
<td>1800 BC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Badtibira</td>
<td>1800 BC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excluded</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akshak</td>
<td>2500 BC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akkad</td>
<td>2200 BC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Reba et al. render spatially the data in Chandler (1987) and Modelski (1999, 2003). Chandler considers the largest cities with population over 20,000 inhabitants from AD 800 to AD 1850. This high population threshold yields an under-representation of Bronze Age cities, which are better covered by Modelski, whose data we use. Modelski uses a lower threshold of 10,000 inhabitants for urban settlements from 3700 BC to 1000 BC. Most of Modelski’s population values come from archeological site reports (Modelski 1999: 384). Sometimes Modelski uses additional historical data (as in Adams 1981). We eliminated 2 Sumerian cities because their locations are controversial (Akkad and Akshak; see Foster 2013). For each century-city, Modelski’s dataset features a missing value or an estimate of the population (in multiples of 1,000). We consider cities with at least one non-missing observation between 3700 BC and 1200 BC.
state emerged that managed a relatively wealthy economy and was also able to protect it for long stretches of time (Bard 1994, Allen 1997). It is accepted that geography shaped the development of both civilizations. Sumer was located in a riverine valley, along the Tigris and Euphrates, exceptionally endowed for alluvial agriculture. And the Nile river and surrounding deserts are credited with shaping outcomes in Ancient Egypt. Such consensus highlights connections between properties of the natural environment and the conditions for prosperity and security. In what follows we detail these connections.

(1) Growth capabilities - The potential of infrastructure for water management. Egyptians could vastly increase their economic output by investing in water management, which in the Nile valley took the form of basin irrigation. Egyptians used a grid of basins to trap the floodwater and increase soil fertility.\textsuperscript{11} Scholars agree that in Egypt irrigation agriculture “could generate crop-to-seed yields of between 12:1 and 24:1 \ldots but only at the cost of high capital investments” (Morris and Manning 2005: 141). For Mann, artificial irrigation was one of the earliest forms of substantial economic investment with high returns. Both in Egypt and Mesopotamia, irrigation agriculture could “generate a surplus far greater than that known to populations on rain-watered soil” (1986: 80).

Like Egyptians, Sumerians made massive investments in irrigation infrastructure, securing 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... Rain-watered soils gave lower yields.” Liverani (2008: 5) gives an idea of the increase in yields: “The agricultural production of barley underwent a notable, possibly tenfold, increase thanks to the construction of water reservoirs and irrigation canals...”

In our model, a high value of the parameter $\rho$ reflects an environment in which investments yield large increases in productivity, reflecting how the construction of irrigation and flood control systems resulted in major expansions of production in Egypt and Sumer. Following this narrative, when we proxy $\rho$ empirically for a given cell in the grid, we use the difference in an income index (the baseline income index is explained below). That difference captures how much extra income, in terms of the caloric potential of agricultural yields, can be generated when moving from rain-fed agriculture to irrigation agriculture. We do this in

\textsuperscript{11}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” claims that irrigation was a public good with large fixed costs, and that pristine states formed in order to provide it. However, there is evidence that irrigation was not always preceded by the emergence of state administrations.
a way that keeps track of the availability of riverine water, to ensure that irrigation was feasible. The availability of riverine water is proxied through a standard measure of river flow accumulation. See Online Appendix F.1 for details.\textsuperscript{12}

(2) \textit{Defense capabilities - Territorial isolation as natural protection.} The Nile basin is surrounded by deserts, which made invasions much less likely than in other food-producing centers. According to Bradford (2001: 9), \textquote{\textit{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 sea (as we discuss below) eventually became a threat rather than a protection, but the desert is considered to have provided two durable kinds of protection. It discouraged the emergence and settlement of hostile neighbors nearby, and acted as a barrier against distant rivals. In terms of our model, Egypt’s territorial isolation maps into a naturally high $\kappa_1$. Following this narrative, when we proxy $\kappa_1$ empirically for a given cell in the map, we use the percentage of the territory in surrounding cells that is covered by desert.\textsuperscript{13} Note that the desert is not an economically beneficial feature of the geography, but the historical narrative suggests it provided security.

In contrast to Egypt, Sumer was exposed to numerous threats. As Bradford (2001: 4) puts it, \textquote{\textit{Their neighbors to the west... 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: Book I, 105) located Sumer in a plain \textquote{\textit{ringed to north and east by mountains, the millenial home of barbarous highlanders, always ready and eager to descend on the wealthy cities below.}} In terms of our model, the vulnerability of Sumerian settlements to invaders suggests that defense capability was lower than in Egypt.

The narratives above suggest that Egypt had high growth capabilities $\rho$ and high exogenous defense capabilities $\kappa_1$. Thus, Egypt lied in a favorable section of $\textbf{PDI}$ or in $\textbf{PS}$, in the North-East of Figure 1(d), which Remark 1 highlights as favorable to civilization. This pat-

\textsuperscript{12}Our quantitative measure of growth capabilities captures something central about the mode of production of these societies, which can be quantified with available data. Ancient populations, however, also made investments in things other than water management, such as storage facilities, kilns, or domesticated animals that consume fodder but raise human productivity.

\textsuperscript{13}We can augment our empirical measure of defense capabilities to include ruggedness without altering the empirical patterns we obtain (results available upon request). A priori, we restricted attention to the desert measure because its effect on security is relatively uncontroversial among historians. Ruggedness, instead, is also thought to prevent states from consolidating power (Fearon and Laitin 2003, Scott 2009).
tern is confirmed quantitatively by the location of Egyptian cities in our empirically-proxied parametric space – see Figure 2 a). Out of five Egyptian cities, four display high values of both parameters $\rho$ and $\kappa_1$ ($\kappa_1$ is extremely high for all five).

Figure 2: Egypt and Sumeria: natural defense capabilities, growth capabilities, and initial income. Darker dots represent higher density of world cells.

In contrast, the case for civilization in Sumer is weaker than Egypt’s. Sumerian locations were described as having high growth capabilities like Egypt, but low exogenous defense capabilities. Given their vulnerability, the trajectories of most Sumerian cities must have begun in relatively insecure parts of $\text{PDI}$, or directly in the stagnant insecurity region, $\text{SI}$. This is borne out in the quantitative approximation in Figure 2 a), where it is shown that natural defense capabilities in Sumer were lower.\(^{14}\) But if output was insecure, how could the first human civilization emerge?

The answer is that Sumerians had a natural endowment that allowed them to invest in their defense capabilities. The archaeological record offers evidence of large investments to improve defense, such as perimeter walls that made Sumerian cities large-scale fortifications.\(^{15}\) Walls were the endogenous, artificial substitute for the missing natural protection that

\(^{14}\)One might think that having nearly 50% of surrounding territory constituting a desert, as in some Sumerian cities, may offer good defense. But a few pathways could suffice for enemies to attack, explaining why historians describe Sumer as insecure. Thus, we restrict attention to comparative statements between Egypt and Sumer.

\(^{15}\)Figure G.1 in the Online Appendix includes illustrations of four Sumerian cities. All of them had walls.
was present in Egypt (where cities did not typically have walls).

According to Remark 2, building artificial defenses required a high initial income $v_1$. We now consider the empirical approximation of this parameter.

(3) **Initial income - Rivers and ecology as directly productive resources.** Egyptian economic life has been characterized as strongly dependent on the Nile, which had at least two important properties: a yearly flood that fertilized the soil, and a two-way navigability that facilitated exchange along the entire valley. According to Bradford (2001: 9), “[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...” Although the Tigris and Euphrates had less attractive properties, Sumer combined the alluvium with an unparalleled initial endowment of plant and animal domesticates. According to Trigger (2003: 281), domesticated animals afforded large gains in labor productivity, and may help explain why Sumer and then Egypt were the first areas in the world to develop civilization.\(^{16}\)

In addition, the rivers offered variation in terms of diet. The aforementioned properties, fertility, easy exchange, ecological diversity, map into a relative high $v_1$ in our model. We take these factors into account in our empirical proxy for baseline income. To capture the presence of rivers we use a measure of river flow accumulation, to proxy for agricultural suitability we use the caloric potential of pre-Columbian exchange crops, to track variation in plants we use a measure of ecological diversity (used previously by Fenske 2014) and to account for variation in fauna we use indicators of the habitat of wild progenitors of domesticated species (from IUCN 2012) according to Driscoll et al. (2009). The index is constructed as an average of the percentiles of the four factors we consider.\(^{17}\)

In fact, virtually every city in ancient history had walls. According to van de Mieroop’s (1997) study of Mesopotamian cities, “*Perhaps the presence of walls was the main characteristic of a city in the eyes of an ancient Mesopotamian.*” Both walls and moats have been estimated to involve large investments (e.g., 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 (1997): 76)).

Diamond (1997: Ch. 8) states that all eight founder crops in the Neolithic were present in the area as well as four of the five most important domesticated animals. He further states (1997: 135) *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.* Olsson and Hibbs (2005) present empirical evidence that corroborates this observation.

16Diamond (1997: Ch. 8) states that all eight founder crops in the Neolithic were present in the area as well as four of the five most important domesticated animals. He further states (1997: 135) *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.* Olsson and Hibbs (2005) present empirical evidence that corroborates this observation.

17Tracking ancient conditions is challenging. Soils have changed over time and estimated yields are sensitive to assumptions on the sophistication of land management. A partial solution to the first problem is to use measures that do not take into account soil constraints. In the Online Appendix F.3 we show that our results are robust to using these measures. The second problem can be partially addressed by assuming
Our quantification exercise again matches the historical narrative and the predictions from the theory. Remark 2 indicates that if defense capabilities are low but growth capabilities are high (Sumer’s case), high levels of income are needed to make civilization likely. Figure 2 b) shows that initial income in the locations of Sumerian cities was high. This is not due to a strong positive correlation between initial income and growth capabilities, as shown in Figure 2 b) where we shade the parametric space to indicate the density of cells. Moreover, most of the cells in the Old World (the darker cloud of background points) lie in locations with relatively low income, emphasizing how atypical the locations of the first cities were.

4.2 False negatives and false positives

One potential concern with our exercise is that some of the locations we deem as failures may be false negatives: those locations could be homes to other civilizations that developed relatively soon after Egypt and Sumer.

The immediate observation is that only two other pristine civilizations emerged in the Old World—in China and the Indus valley—, which drastically reduces the number of potential false negatives. In addition, we can locate the cities from these civilizations in our parametric space to determine whether they arose in areas that the theory predicts as failures.

The Indus valley gave rise to a multi-city civilization in locations both upstream (e.g., Harappa) and downstream (e.g., Mohenjo Daro), along the Indus and nearby rivers, reaching its peak in the 2600-1900BC period. Civilization in China is believed to have emerged in the Yellow River, giving rise to the Erlitou and Erligang cultures in the 1900-1200BC period. Roughly at the same time, the Sichuan basin gave rise to the largely independent Sangxingdui culture. Both the Chinese and Indus civilizations emerged, like Egypt and Sumer, in riverine valleys.

Figure 2 a) shows that both the Chinese and Indus civilizations do not lie in failure areas. They resemble Sumer in that the Chinese and Indus locations lacked high natural defenses but had high initial income with which to mount defenses (the only exception is Dholavira in the Indus Valley, located in an island inside a lake). As we explain in Section H of our Online Appendix, archaeological research suggests that both civilizations lied in areas where—as we do—the lowest sophistication in inputs and management allowed by the FAO-GAEZ data. Online Appendix F further discusses our measures and robustness.
defense was a concern, initial income was high, and investments likely had high returns.

A second concern is that while Egypt and Sumer match the predictions of the model for civilizational success, many other locations in the Old World may also match the predictions, constituting false positives. A first indication against this possibility is present in Figure 2, where ancient cities appear to have emerged in statistically rare areas. In other words, not many locations other than Egypt and Sumer were as well-suited for civilizational success.

We now corroborate that message in an alternative way. As we have seen, the window for civilization “Egypt-style” was a combination of high natural defense and high growth capability, while the one “Sumer-style” was a combination of high growth capability with high initial income. These conditions can be defined more or less stringently, depending on what cutoff level of each parameter is considered to separate high from low levels. If one defines the cutoffs to be rather low, most locations in the world’s map will appear to satisfy both conditions. As cutoffs are raised, fewer locations will appear to satisfy either condition.

Figure 3 shows three maps, each identifying cells that satisfy none, one, or both conditions. We raise the cutoffs that separate high from low levels of the parameters as we move from the top to the middle and low maps. Consequently, fewer cells in each map satisfy the conditions as the conditions get more stringent. The map at the top shows that cities from all four pristine Old World civilizations fall in regions satisfying at least one condition. However, there are other locations that feature similar characteristics, but yielded no known early civilization. The following (middle) map imposes higher cutoffs for the proxies of the model. Egypt and Sumer are still colored, as are regions around the cities in the Indus valley and China. Many of the suitable locations without civilizations, however, disappear. The bottom map imposes extremely stringent cutoffs, and a tiny set of cells is shown to survive, including Sumer and Egypt (as well as part of the Indus valley). This highlights the exceptionality of the locations for Sumer and Egypt, the two earliest human civilizations. Online Appendix I displays maps using other cutoffs.

5 The end of the Bronze Age

After identifying conditions for the rise of civilizations, we now address their fall. For a period of almost 400 years, multiple states emerged in the Eastern Mediterranean that prospered and were able to defend their wealth against “barbarian” populations. This set of thriving
Figure 3: Black (white) pixels represent high initial income and high growth capabilities (high natural defense and high growth capabilities). The green circles and polygons represent ancient cities. Within a map, cutoffs differ for better visualization: e.g., the bottom map displays defense and growth capabilities in the top 5 and top 1 percent (resp.) as no location has both capabilities in the top 1 percent.
states included the city-ports of the Levant, the kingdoms of Anatolia, the Egyptian empire, and the city-states of Mesopotamia and Cyprus. But a collapse epidemic swept across the entire region around 1200BC. As Cline puts it (2014: 241), “...the world as they had known it for more than three centuries collapsed and essentially vanished”. According to 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.”

A long debate on the causes of the collapse at end of the Bronze Age has considered earthquakes (Schaeffer 1948), droughts and famines (Carpenter 1968), internal rebellions (Zuckerman 2007 and Carpenter 1968), and innovations in military technology (Drews 1993). The hypothesis of earthquakes has been discredited in the face of new archaeological evidence showing that most urban destruction was caused by humans. Hittites and Egyptians left unequivocal testimonies of attacks by the “Sea Peoples,” as the Egyptians called them, a diverse array of intruders with different origins (Sandars 1987). The same evidence challenges a pure internal rebellion story. The possibility of invasions remains, but begs the question of what caused the invasions. Two hypotheses consistent with available evidence are:

1. A severe climate shock (drought), which caused famines, and compelled populations in the periphery to invade. 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, which tipped the military balance in favor of nomadic intruders. According to Drews (1993: 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... 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”? Chrissanthos (2008: 11) summarizes them: “...better and lighter body armor, [...] lighter and smaller round shields, [and] revolutionary longer, stronger swords [...] They also invented 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

---

18See Langgut, Finkelstein, and Litt (2013) for a paleobotany study supporting the idea of a drought, and an interpretation involving rebellions.
We now use our model with exogenous defense capabilities as a tool for short-run analysis given the (relative) sudden nature of the shocks at the end of the Bronze Age. Consider the challenger’s valuation to be parameterized as \( h \times v \) \( (h > 0 \text{ for hunger}) \), and the challenger’s military capability \( \kappa_c \geq 1 \), so the challenger’s expected benefit reads \( \frac{b_v}{a_t+b_t} h V_{t+1} - \frac{b_t}{\kappa_c} \). The historical debate has sometimes considered changes in the challenger’s motivation to attack \( (h) \) and its military effectiveness \( (\kappa_c) \) as the rival explanations (1) and (2) above. But the “hunger” and the “military innovation” hypotheses, while historically distinct, are formally identical in our model. At the margin, \( h \) and \( \kappa_c \) affect the aggressiveness of the challenger in the same way. Therefore, fixing \( h = 1 \) and studying the comparative statics of \( \kappa_c \) illuminates the effect of changes in both the motivation and aggressiveness of challengers.

The parameter \( \kappa_c \) was assumed equal to 1 in the baseline model. We now consider a move to \( \kappa_c > 1 \). How will the incumbent fare when facing a tougher challenger? In other words, how does a higher \( \kappa_c \) affect the partition of the parameter space derived in Proposition 1? The following proposition yields the answer.

Proposition 3 For any point in the \((\kappa_1, \rho)\) space where either security or prosperity (or both) are attained, a higher \( \kappa_c \) implies that security, prosperity or both may be lost. A higher \( \kappa_c \) reduces the area of \( PS \) (where both security and prosperity obtain). In addition, \( SI \), which yields insecurity and stagnation, grows at the expense of all others, making it less likely that the incumbent capabilities \((\kappa, \rho)\) are compatible with civilization.

The comparative statics of \( \kappa_c \) are intuitive: a more aggressive challenger reduces the set of parameter values for which the incumbent attains both growth and security, and enlarges the set in which he attains neither. Figure D.1 in the Online Appendix includes a graphical representation of the comparative static effects of \( \kappa_c \).

---

\(^{19}\)Localized climate shocks and military innovations are not mutually incompatible causes and can be part of a systemic collapse (Cline 2014). Late Bronze Age societies were part of a commerce network. Thus it is possible that climate- and technology-induced invasions that devastated only a critical number of nodes set off domino effects.

\(^{20}\)The parameter \( h \) could also track the differential ability of the challenger at “operating” the asset under the interpretation that a successful challenge leads to replacement. 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) but a low valuation for the asset due to an inability to operate it. These are interesting variations left for future research.
6 Conclusion

We built a model to investigate the rise of civilization, arguably the most important transformation in the history of humankind. Civilizational breakthrough requires that agricultural settlements solve the problem that economic growth causes voracity by surrounding non-agricultural societies. Economic growth prior to civilization is a self-liquidating enterprise. The choice of components in the model, informed by the anthropological and historical literatures, captures the essential environmental parameters and the minimalistic strategic dilemma facing proto-civilizations.

The theory contributes three messages. First, a weak form of the Hobbesian tenet that security matters for prosperity holds, because without a substantial degree of security productive investments, and hence civilization, are inviable. Second, a strong form of the Hobbesian idea does not hold, because full security is not necessary for civilization. Substantial investments and growth are possible even in the face of some insecurity. Third, the tension between prosperity and security—which we have called the paradox of civilization—is resolved not by institutions seen as formal rules of the political game, but by tangible assets shaped by the geographic environment.

The role of geographic advantages has been emphasized by analytic perspectives in anthropology, especially in connection with productive potential. The role for defense and the resulting tradeoffs in the allocation of resources have been less prominent, despite the fact that defense infrastructure is a major component of archaeological data. Our theory integrates productive and defense fundamentals through the lens of positive political theory, in a way that we believe has some empirical traction.

The theoretical prediction is clear-cut: civilization requires a combination of high defense and growth capabilities, or, if initial defense capabilities are low, a combination of high growth capabilities and high levels of initial income. High initial income or growth capabilities are not individually sufficient conditions for civilization. Information on the location of the first two human civilizations matches each of these predictions. Our illustration relies both on historical narrative and a quantification of the parameter space of the model. We use our predictions not only to rationalize the emergence of civilization, but also the first macro-episode of civilization collapse.

Quantifying the model’s parametric space offers a way to enrich the historical analysis
beyond the narrative approach. Our exercise does not amount to a formal statistical test, and we did not use it to rule out alternative theories. Future work should address these limitations, and relax the simplifications we have imposed on both the theory and the historical illustrations. Hopefully this lays out exciting possibilities for future research.

References


