THE PINHAS SAPIR CENTER FOR DEVELOPMENT
TEL AVIV UNIVERSITY

"Geography, Transparency and Institutions"

Joram Mayshar[1], Omer Moav[2], Zvika Neeman[3]

Discussion Paper No. 5-15

February 2015

We would like to thank The Pinhas Sapir Center for Development at Tel Aviv University for their financial support.

We are grateful for comment received from Daron Acemoglu, Bob Allen, Josh Angrist, Eddie Dekel, Diana Egerton-Warburton, Christopher Eyre, James Fenske, Oded Galor, Maitreesh Ghatack, Jeremy Greenwood, James Malcomson Andrea Matranga, Jacob Metzer, Stelios Michalopoulos, Motty Perry, Torsten Persson, Louis Putterman, Ariel Rubinstein, Yona Rubinstein, Matthew Spigelman, Yannay Spitzer, Nathan Sussman, Joachim Voth, David Weil and from seminar participants in Royal Holloway University of London, Ben-Gurion University, The Hebrew University of Jerusalem (department of economics), The Hebrew University of Jerusalem (department of archeology) , Bilkent University, Tel-Aviv University, University of Aberdeen, University of Edinburgh, ISET, LSE/UCL growth- development seminar, University of Warwick, University of Alicante, Oxford University, Harvard/MIT applied theory seminar, NYU Brown, Indiana, Di Tella, Copenhagen, Odense, Sant'Anna Pisa, and conference participants at The Royal Economic Society Conference, Royal Holloway, April 2011; CEPR Public Policy Symposium, Zurich, May 2011; Minerva Summer School in Economic Growth, Jerusalem, June 2011; Barcelona II Workshop "Towards Sustained Economic Growth" October 2011; NBER Summer institute, "Inequality and Growth", Cambridge Massachusetts July, 2012, Conference of the Economic History Association of Israel, Bar Ilan University, December 2012; CAGE/CEPR Conference Long-Run Growth: Unified Growth Theory and Economic History, University of Warwick, January 2013; CEPR Economic History Symposium; Perugia, April 2013 Warwick Summer School in Economic Growth, Coventry, July 2013.

[1] Joram Mayshar - Department of Economics, Hebrew University of Jerusalem. Email: msjoram@huji.ac.il
[2] Omer Moav - Department of Economics, University of Warwick, Department of Economics Hebrew University of Jerusalem. CAGE and CEPR. Email: omer.moav100@gmail.com. Moav's research is supported by the Israel Science Foundation (Grant No. 73/11).
[3] Zvika Neeman - The Eitan Berglas School of Economics, Tel Aviv University. Email: zvika@post.tau.ac.il

Abstract

We propose a theory by which geographic variations in the transparency of the production process explain cross-regional differences in the scale of the state, in its hierarchical structure and in property rights over land. The key linkage between geography and these institutions, we posit, is via the effect of transparency on the state's extractive capacity. We apply our theory to explain institutional differences between ancient Egypt and ancient Upper and Lower Mesopotamia. We also discuss the relevance of our theory to analyses of the deep rooted factors affecting economic development and to the growth of taxation in the modern age.

Keywords: Transparency, Institutions, Land Tenure, State Capacity, Hierarchy

JEL Classification numbers: D02, D82, H10, 043

1 Introduction

The protection of property rights has a paramount role in recent theories on the success of nations (North 1981). In that vein, Acemoglu and Robinson (2012) argue that extractive institutions that "are designed to extract incomes and wealth from one subset of society to benefit a different subset" comprise the most detrimental factor for economic prosperity. Besley and Persson (2009, 2010), on the other hand, consider the state's capacity to tax as a precondition for its ability to supply public goods and thereby enhance economic growth. These two theories only appear to conflict. Both posit that extracting resources by the state is beneficial, conditional on the taxes being predictable and not excessively distortionary and on the revenue being used for the provision of public goods rather than for consumption by the elite.

In this paper we offer a different view, which we reached by analyzing the characteristics of successful ancient states. We note that ancient states supplied the basic public good of defense; however, there is no evidence that these states were designed for anything other than to maximize the extraction of resources from the agricultural hinterland for consumption by the elite. As such they appear to have lacked the prerequisite conditions for success according to the above-mentioned theories. Yet Ancient Egypt, for example, had a rich civilization that enabled the construction of the great pyramids and was stable over several millennia – in spite of its extractive government and of the peasantry's lack of title to land.

Our prime motive, though, is not to challenge the theories on the institutional preconditions for modern prosperity, since our presumption that economic success went hand in hand with a strong, extractive government may have applied only to the agriculturally based economies of antiquity. Rather, our goal is to explain the major institutional differences that existed between extractive yet successful early states – and in particular between Egypt and Mesopotamia. Our basic argument is that what distinguished the successful nations of antiquity was the state's ability to appropriate revenue. This ability, in turn, was determined primarily by geographical and technological considerations that are overlooked by the recent studies on extractive institutions and state capacity. Accordingly, we attribute the success of Ancient Egypt to the facts that its staple cereal food required storage, and that farming activity was highly transparent. Transparency and storage enabled the central government to extract massive taxes using only a lean bureaucracy, and without generating significant distortions. Our general claim, that an increase in the transparency of production can account for a higher state capacity to tax, is applicable, we posit, also to the modern expansion of the state, and not only for antiquity. Moreover, since our theory offers an explanation for the emergence of different institutions across regions of the world, we believe that it is highly relevant for understanding the deep rooted factors that play a key role in the recent comparative development literature.[1]

In emphasizing the crucial role of geographical and technological factors, we follow a long and diverse scholarly tradition, including Childe ([1936], 1951), Braudel ([1949], 1976), Wittfogel (1957) and Diamond (1997). Our contribution to this scholarship is in identifying the specific factors that make tax capacity the key link between geography and institutions. To illustrate this contribution, consider Diamond's (1997) influential book in which he claims to explain the economic underdevelopment of New Guinea primarily by its inability to adopt the productive agricultural innovations that benefited Asia and Europe. We suggest, instead, that what explains the economic underdevelopment of New Guinea is not the lower productivity of the form of agriculture that was adopted there, but rather that this form of agriculture was not conducive to expropriation, and thus did not facilitate the emergence of a state. Indeed, New Guinea is known to have adopted agriculture early on. Yet, due to its tropical conditions, its main agricultural produce was based on roots and tubers whose non-seasonal nature and greater perishability (upon harvest) imply that storage is both inefficient and not strictly required.

The underlying reasoning for these claims appears in a related paper (Mayshar, Moav and Neeman, 2011) in which we propose that the transformative facet of the Neolithic Revolution that gave rise to the emergence of social hierarchy and the state was not the increase in productivity per se, but rather the induced radical change of the tax technology.[2] The seasonality of cereals meant that food storage became not only feasible (due to non-perishability) but also mandatory. This made farmers vulnerable to expropriation, both by robber bandits and by the state: in a single inspection, the would-be expropriator could impound a large fraction of a farmer's annual output.

In this paper we do not explore the emergence of states. Rather, we assume the state to exist, with a monopoly on the exercise of power. We focus on factors that contribute to state viability; and in particular on one environmental factor that contributes to a state's capacity to extract revenue: the degree of informational transparency of the production process. The phenomena that we seek to explain include: the scale of the state; its hierarchical structure, including the power of

[1] Spolaore and Wacziarg (2013) survey the relevant literature. Bockstette, Chanda and Putterman (2002) show that 'state antiquity' (an index of the depth of experience with state-level institutions) predicts income per capita, institutional quality and political stability in the present. Other recent papers that show empirically the effect of geographical factors on current economic outcomes include, among many others: Putterman and Weil (2010), Michalopoulos (2012) and Ashraf and Galor (2013).

[2] The notion of a 'tax technology' was proposed by Mayshar (1991); see also Slemrod and Yitzhaki (2002).

the center vs. the periphery (and the degree of urbanization); land tenure arrangements; and the existence of property rights to land. That is, we aim to explain variations in such institutions in antiquity across geographical regions, by variations in the transparency of the production process.

In section 2 we summarize the distinctive institutional features of Ancient Egypt and Mesopotamia that we seek to explain. We provide there also a brief literature review of the principal-agent model that we employ. The model, in which we interpret the agent as a representative farmer and the principal as an absentee land-owner representing the government, is presented and analyzed in section 3. The most distinctive feature of the model is that we focus on the effects of variations in the degree of transparency (or the degree of informational asymmetry). This degree of transparency, our main exogenous variable, is represented by the accuracy of a signal that the principal observes on the state of nature and from which she infers on whether the agent worked diligently.[3] Another distinctive feature of our model is that we limit the incentive scheme that the principal may employ. In addition to a bonus payment, a "carrot", on delivering high output, we allow only a "stick" in the form of possible dismissal, for suspected shirking.

Our main proposition is that the more accurate the public signal, the larger the role of the stick and the smaller the role of the carrot. That is, greater transparency increases the likelihood that the principal will dismiss the agent upon suspected shirking. At the same time it also reduces the agent's remuneration. In this sense, greater transparency induces a form of servitude. On the other hand, sufficient opacity may result in the state allowing the agent to retain a larger share of the output without any threat of dismissal. We interpret the latter result as showing that lack of transparency protects farmers' freedom and secures their property rights over the land that they cultivate. The farmers in this case de-facto own the land, even though the extractive state has an absolute power to expropriate both their output and their land.[4]

In a two-layered extension of this model (section 3.2), we examine the role of differential transparency at different levels of the governmental hierarchy. This allows us to explain the degree of state centralization. We show that when farming activity is sufficiently transparent, not only to the intermediary (governor) but also to the upper level of the hierarchy (king), the intermediary retains a smaller share of the revenue, and is subject to dismissal. On the other hand, if farming

[3] In this model the principal is assumed to observe the output but not the state of nature or the agent's effort. In appendix A we present an alternative framework that delivers the same qualitative results, in which the principal does not observe output and the moral hazard problem pertains to hiding (or misreporting) output by the agent, rather than to exerting little effort.

[4] We take the state's absolute extractive power as given. The possibility that such power requires resources to finance enforcement would reinforce our results: greater transparency implies larger revenue to facilitate the power to extract.

has low transparency for the king, the governor retains autonomy and a larger share of revenue. The way we see it, it was indeed the ability of the central authority to control the subordinated lords, and to extract revenue through them, that was the key for the success of early central states.

In section 4 we examine the applicability of these theoretical insights by studying the institutions that persisted for several millennia in the earliest civilizations of Mesopotamia and Egypt.[5] Among other things, we argue that the faster rise of a powerful central state in Egypt and its subsequent greater stability, compared to Mesopotamia, reflect the fact that Egyptian farming was highly transparent, both at the local level and at the state level. We argue that these features also explain the absence of land-owning peasantry in ancient Egypt and the weakness of its cities. We note further that our theory can explain also key differences between Upper Mesopotamia (Assyria) and Lower Mesopotamia (Babylonia). In the former, mostly rain-fed region, transparency was relatively low, while in the alluvial south, transparency was high at the local level of the city-state, but limited at the level of a more hierarchical central state. In Section 5 we consider the applicability of our model for explaining the modern growth of government. Section 6 concludes.

2 Background

2.1 Our motivating comparison: Ancient Egypt vs. Ancient Mesopotamia[6]

It is believed that intensive agriculture was first adopted in the highlands of Anatolia and northern Mesopotamia in the seventh millennium BCE. This was two and three millennia respectively before agriculture was taken up in the alluvial planes of southern Mesopotamia (Sumer) and in the Nile valley. It was in Sumer that the first major city-states were formed in the fourth millennium; an advance that has been described as a "takeoff" and as an "Urban Revolution" (Liverani 2006, Childe 1951).[7] And still, the first central territorial state was formed in Upper Egypt, at about 3000 BCE. These developments raise two puzzles: why did the Urban Revolution take place in Sumer, even though it was a relative late-comer to agriculture; and why did a strong and stable central state first develop in Egypt, even though it was an even later adopter of agriculture?

The rapidity of its formation and the subsequent stability of the central state in Egypt (Kemp

[5] While we focus on these two civilizations, and particularly on the fourth to second millennia BCE, these insights are clearly not restricted to these countries or to that early period.

[6] Further details and references for the issues discussed in this introductory sub-section are provided in section four below.

[7] Like many generalizations, this one has to be qualified somewhat. Archaeologists have identified substantial ancient city-mounds in Upper Mesopotamia from as early as the late fifth millennium (Ur 2010). The geographic span of control of these cities, however, was limited in comparison to that of Sumerian city-states like Uruk.

2006 and Wenke 2009) are just two of the key features that distinguish between ancient Egypt and Mesopotamia. Baines and Yoffee (1998, p. 268) conclude: "the two civilizations are profoundly different." Scholars have often noted additional distinguishing features (Trigger 1993, 2003). For example, even though there were fortified city-states in predynastic Egypt, after the formation of the central state Egyptian cities were no longer fortified and played a fairly limited role as administrative centers. This led Wilson (1960) to famously characterize Ancient Egypt as "a civilization without cities."[8] In contrast, up to the first millennium, Mesopotamia was ruled most of the time by rival and independent city-states, leading Adams (1981) to characterize southern Mesopotamia as "the Heartland of Cities." Additionally, commercial activities involving long-distance trade, the sale and lease of land, and extensive lending were widespread in Mesopotamia and amply recorded in numerous surviving documents. From the extensive records that survived from ancient Egypt it appears that commercial activity of this type was rather limited, at least up to the mid-first millennium. Accordingly, as Wilson (1960) noted, whereas written law codes are known in Mesopotamia already in the late third Millennium, they were absent in Ancient Egypt (Westbrook 2003).

We argue that these key differences between ancient Egypt and Mesopotamia stem primarily from differences in the degree of transparency of farming activity. In particular, we contend that the principal feature that distinguished Egypt was the transparency of its water supply. Indeed, scholars have known all along that the publicly observed level of the Nile revealed the 'state of nature' faced by individual farmers throughout Egypt with high accuracy.[9] We interpret this feature as implying high transparency at both the local and the state-wide levels. We argue that this feature enabled the local lords to raise revenue efficiently by incentivizing the cultivating peasants with minimal material reward, coupled with a threat of dismissal. This, we contend, explains why Egypt did not resort to owner-operated farming. More importantly, the Nile's transparency enabled the Pharaohs to employ a similar incentive scheme towards the district governors and down the chain of middlemen engaged in remitting taxes to the center. This enabled the pharaohs to siphon off

[8] Wilson contended (1960, p. 135) that "Ancient Egypt carried on her life through dozens of moderate-sized towns and myriads of agricultural villages;" but was "without a single major city." A number of Egyptologists object to Wilson's claim, emphasizing the existence of walled cities in the predynastic period and the prevalence of diverse population centers (see Kemp 2006). But this hardly contradicts Wilson's generalization, particularly if taken with the perspective of Mesopotamia (see Baines and Yoffee 1998, p. 209 and Redford 1997). Thus, whereas Van de Mieroop (1997, p. 1) writes: "Mesopotamia was not only the oldest urban civilization, but also the most urbanized society of antiquity," Eyre (1999, p. 35) summarized the situation in Egypt in Wilson's spirit: "Egypt was probably always a village society."

[9] See Cooper (1976) and further details in section 4 below.

a significant share of the country's agricultural product, via provincial administrative centers that had little independent power.

We argue further that the substantial differences between Upper (northern) and Lower (southern) Mesopotamia may be attributed to comparable differences in the transparency of farming. The southern alluvial plains lent themselves to centralized control more than the mostly rain-fed highland in the north. A differential degree of transparency between the lower and higher levels of hierarchy may also explain why the cities in Mesopotamia were powerful, unlike those in Egypt, and why successive attempts to unify Mesopotamia under a central state were prone to failure.

We conclude this overview by summarizing several alternative attempts to explain what made the ancient civilizations of Mesopotamia and Egypt so successful. The first theory to be examined is Wittfogel's (1957) "Hydraulic Theory" (based on ideas by Engels and Weber), which sought to explain why all the ancient civilizations emerged in riverine environments. The theory posits that large-scale irrigation infrastructure was necessary to realize the agricultural potential in such environments, and that a strong centralized state was an absolute prerequisite for the construction and management of such irrigation systems. As many critics have pointed out, however, the irrigation systems in ancient Egypt and Mesopotamia (and in China, Peru and most elsewhere) were initiated as local ventures, and constructed communally prior to the emergence of a strong central state. Moreover, even after a central state emerged, the management of these irrigation systems remained with the local elites. To the extent that central states in antiquity initiated large-scale irrigation projects, these were the consequence of central governments rather than the cause for their emergence.[10] Wittfogel's theory is now considered defunct; but this leaves the coincidence between these early civilizations and irrigated riverine environments largely unexplained. Our theory sheds new light on this coincidence.

The second theory argues that urban centers and the state emerged in order to provide the infrastructure required for facilitating trade. In this vein, Algaze (2005, 2008) downplays the significance of irrigation and agriculture and attributes the "takeoff" of the city-states in Lower Mesopotamia (Sumer) mostly to trade. In an attempt to apply theories on the growth of European medieval trading centers, Algaze emphasizes the diversity and richness of the ecosystems within Sumer and the relative ease of riverine transportation that gave rise to short-distance trade and,

[10] Wittfogel invoked a Hobbes-like scenario whereby farmers in these early civilizations (including China and Peru) elected to realize the productivity advantages of the "hydraulic potential," even though they fully recognized that it entailed the "loss of personal and political independence" (1957, pp. 15-19). Wittfogel's many critics include: Adams (1960, p. 280; 1981) concerning Mesopotamia, and Wilson (1960, pp. 130-131) and Butzer (1976, pp. 110-111) concerning Egypt.

later on, to significant long-distance trade with colonies up the Euphrates. This theory, though, can hardly be applied to Egypt, where trade was of secondary significance.[11]

The third theory emphasizes the role of government in coping with uncertainty. Adams (1981, p. 244; 2005) contends: "In the largest sense, Mesopotamian cities can be viewed as an adaptation to the perennial problem of periodic, unpredictable shortages. They provided concentration points for the storage of surpluses." Park (1992) makes a similar claim with regard to ancient Egypt. These scholars envision that the overriding problem that early farmers faced was due to aggregative food shortages, for which they were unable to insure themselves, other than through inter-annual storage at the state level.[12] There is indeed ample archaeological evidence for extensive storage at the state level in these two societies and in other early states. However, as we argue below, it is unlikely that inter-annual storage at the state level served as an important mechanism to sustain farmers in years of famine. As Sen (1981) argues in another context, peasants typically lack political clout. This makes it doubtful that they could expect support from the central government in years of overall shortage. Long-term storage at the center must have primarily served the urban elite and its immediate urban dependents.

The above three theories explain the emergence and success of early central states by the services that they presumably rendered to the population. Olson (1993) criticizes these teleological explanations for their "unhistorical assumption that government arose out of some kind of voluntary choice." According to him (and to his antecedents), the state emerged to serve the interest of the ruling elite – of "stationary bandits" who replace "roving bandits," in his terminology – in extracting resources from the population. This idea, though, hardly explains the early success of central states in Egypt and Mesopotamia, or the differences between these civilizations.

[11] For an alternative critique of Algaze's perception see Adams (2005). Bates (1983) proposes a similar trade theory to explain the existence of states in pre-colonial Africa. In his version of the theory, long-distance trade not only requires state institutions, but also supplies a main source of tax revenue to sustain the state (p. 30). Fenske (2013) provides evidence in support of the positive effect of trade on the emergence of more centralized states in pre-colonial Africa. But his analysis leaves mute the issue of whether the main mechanism for this causal effect is the benefit that states confer on trade or the effect of trade on the ability to tax. Given the vast difference in antiquity between the magnitude of the value of long distance trade (mostly in luxury goods serving the elite) and of agricultural production, it is evident why we consider the ability to tax agricultural output as the crucial consideration to understand early states.

[12] Halstead (1989) offers a similar theory for ancient Greece, indicating that this theory has little to do with riverine environments. We do not contest the applicability of this theory to the local village level, where direct bilateral relations prevailed and were central storage was advantageous due to returns to scale. We contest it only to the state level. Exogenous uncertainty plays an essential role also in our model. However, it serves only as a prerequisite for the existence of moral hazard. Moreover, in our context, increased exogenous uncertainty may be considered to be a key source for opacity of farming, and as reason for diffused, owner-occupied farming, rather than centralization.

Carneiro's (1970) "circumscription theory" provided an influential non-teleological theory, according to which ancient states could emerge only in circumscribed areas from which concurred farmers were unable to escape subjugation through migration. This theory was developed in an attempt to explain the South American experience, where states emerged in the circumscribed Andean valleys and not in the Amazon, but is often claimed to apply more generally.[13] Thus, Allen (1997) argues that Carneiro's circumscription theory is particularly applicable to desert-surrounded Egypt.[14] Moreover, Allen contends that Egypt's extreme circumscription provides the key factor that distinguished its institutions from those of Mesopotamia. This claim, though, indicates that Carneiro's geographical circumscription theory may not be applicable to Mesopotamia (or China), and reminds one that it provides also no explanation for the coincidence of early central states and irrigated riverine environments.[15]

Our non-teleological theory does not purport to explain the emergence of states, but it provides, we contend, a superior explanation for the on-going viability of states, once formed, and for related early institutions. In particular, it suggests that the coincidence between irrigated riverine environments and early central states was due not to that the construction and management of irrigation systems in arid riverine environments *necessitates* a despotic central state, as Wittfogel claimed.[16] Rather, we argue, it was due to that irrigation systems provide information and means of control, and thus *facilitates* on-going revenue extraction that is an essential prerequisite for the viability of

[13] This theory was adapted, among many others, by Mann (1986), who employed the term "caging" and referred more generally to social, rather than just geographical, confinement.

[14] We agree with much of Allen's analysis of ancient Egypt, though we differ on several key issues, and particularly with his main explanation of Egypt's uniqueness. Manning (2003, pp. 229, 28, 132-3), using terminology from both Carneiro and Mann, argues similarly to Allen: "What was unique about Egypt was its ecology, a narrow valley circumscribed by desert at both sides, which successfully caged the population. . . ." It should be noted that Carneiro's theory envisions a background of violence that is precipitated by population pressure, and that ends up with the essential enslavement of the subjected population. As Allen (1997) notes, there is no evidence for population pressure in the Nile valley in the fourth millennium BCE. More pertinently, it would be far-fetched to describe the peasantry in Egypt as enslaved.

[15] Carneiro's concern with emigration seems suspect also from a long-term Malthusian perspective (for evidence see Ashraf and Galor 2011): as long as the extracting elite is not too short-sighted and leaves the subjected farmers with subsistence food, fertile land will not remain uninhabited. Other scholars emphasize alternative aspects of the Nile valley to explain Egypt's 'uniqueness,' such as the ease of communication and control, due to the concentration of almost the entire population along the Nile, and the ease of transporting the tax revenue to the center. From our perspective, Egypt's circumscription may have helped its central state, not by inhibiting the population from escaping but, rather, by insulating Egypt from external threats. Unlike Mesopotamia, where nomadic bands posed a perennial security problem, there was no on-going need for fortified regional centers in ancient Egypt (other than at its southern tip).

[16] This coincidence is supported by Bentzen, Kaarsen and Wingender (2012). Instrumenting for actual irrigation with geographical factors that indicate its potential to increase yields, they show that irrigation-based societies have been less likely to adopt democracy.

any state.[17]

2.2 The pertinent economic literature: a selective review

Our paper touches upon wide-ranging economic issues. Rather than attempt a complete literature survey, we opt to cite some of the seminal references for these issues, and a few of the most pertinent studies. Greif's studies (1993, 2006) are the closest to our approach in employing formal game theoretic models to examine historical economic institutions. Greif (1993) analyzes how medieval merchants incentivized their agents with schemes that combined direct remuneration with a threat of dismissal. Greif (2006) studies the institutional foundations that underlie modern trade in impersonal markets. The historical context that concerns us, however, is entirely different, with emphasis on land tenure and government control, rather than on trade.

In emphasizing the role of asymmetric information in taxation we draw, of course, on Mirrlees' (1971) seminal contribution. However, whereas his framework is based on heterogeneity in unobserved skills, what we are after is the effect of alternative degrees of observability (transparency). The standard principal-agent problem that we adopt has often been applied to sharecropping contracts in agriculture (Stiglitz 1974). Sharecropping is known to have a major disadvantage relative to fixed rent contracts in that it distorts incentives. Ackerberg and Botticini (2000) provide a detailed survey of the literature and test three hypotheses concerning the potential advantages of share tenancy over fixed rent contracts in an historical medieval context: (1) risk-sharing: sharecropping offers a balance between incentives and risk sharing; (2) moral hazard: sharecropping better protects land from abuse; (3) imperfect capital markets: sharecropping allows poor tenants who face borrowing constraints to be profitably employed in agriculture. On the basis of a sample of landlords and tenants from medieval Tuscany, they find support only for the last two hypotheses. However, by controlling for the biases due to endogenous matching in the same data, Ackerberg and Botticini (2002) provide evidence suggesting that the first hypothesis – risk sharing – is also an important determinant of contract choice.[18]

[17] Our theory is supported also by evidence from the arid northern coast of Peru. Billman (2002) reports that early irrigation in the valley of Moche consisted of short canals that did not require much labor or leadership. Nevertheless, this simple form of irrigation created an opportunity for leaders "to control land and the flow of water" and enabled them "to finance the creation of centralized, hierarchical political organizations" – thus leading to the formation of an early territorial state in this valley (400 BCE–800 CE). By extending these irrigation canals, the leaders were able to consolidate their power through expanding the arable land and granting "use-rights to commoners in exchange for annual payments in labor or a percentage of the harvest" (p. 394).

[18] Only the third consideration plays a role in our model, since we assume tenants to be risk neutral and do not consider the possible abuse of land. It is the assumed imperfectness of capital markets that rules out fixed rental contracts that could eliminate the incentive problem in our context. Note though that if poor agents could borrow in

In the standard principal-agent model, the principal incentivizes the agent by designing a contract that consists of a 'carrot,' in the form of bonus output to be retained by the agent upon diligent work, and a 'stick' to be applied upon suspected shirking. We modify this standard model in three directions. First, we restrict the principal's ability to penalize the farmer upon suspected delinquency. The stick is assumed to be in the form of a threat of the agent's dismissal – rather than a low (or even negative) reward, corporal punishment or anything else that may jeopardize the agent's long-run ability to work efficiently.[19] Moreover, we assume that the dismissal of an agent, and his replacement by another, is costly to the principal. This modification is underlined by three presumptions: (a) the agent has to be maintained intact in order to be available for diligent work in the future;[20] (b) agents have work-locale specific knowledge so that their replacement is costly; (c) tenants have no wealth of their own and cannot borrow. In focusing on the sanction of dismissal, our model is related to the literature on unemployment as a discipline device and on "efficiency wages" (see Shapiro and Stiglitz, 1984).[21]

Our second modification of the standard framework is in embedding the principal agent relationship in a multi-period setting – the agent is employed repeatedly, as long as he is not dismissed.[22] Dismissal is a painful stick for the agent, since upon dismissal he is no longer employable in farming and has to relocate to the urban sector, were he enjoys no rents.[23] Our third modification is the extension of the model to a framework with a multi-tier hierarchy of principal-agent relations.

Our focus is on the effect of variations in exogenously-determined transparency. This idea is related to the literature on costly imperfect (endogenous) monitoring of the agent's actions by the principle. Such monitoring is known to affect the nature of the contract. Harris and Raviv (1979) established that costly monitoring will not be undertaken if agents are risk neutral and the principal has unrestricted ability to penalize a delinquent agent. We assume risk neutrality and assume that the observation of a signal on the state of nature (a form of monitoring) is costless; but, as noted,

order to meet fixed rent when output is low, the principal-agent problem would be shifted from the farming contract to the loan contract.

[19] Cheung (1969) shows that dismissal is common in sharecropping contracts and could increase efficiency.

[20] In addition to the landlord's interest to maintain the agent intact, one may presume that in years of famine farmers desisted from surrendering food to the landlord if that jeopardized their survival.

[21] In their framework, increased transparency leads to lower wages, as in our model, but it also leads to more employment, in contrast to our long-term predictions (studied in appendix E in which we analyze the case of endogenous plot size) where more transparency leads to less employed farmers.

[22] The bulk of the literature on multi-period principal-agent relations is concerned with the possibility of many agent types and with the principal learning the type of the agent over time (e.g., Holmström 1999). In our mode there is only one type of agents; the multi-period framework renders the disutility of dismissal to the agent endogenous.

[23] Wu (2012) presents historical evidence that cities were indeed a population sink with negative natural population growth, and were migration from rural areas served as a source of city population.

we restrict the penalty.

In examining alternative contracting schemes in the employment of labor and land our model relates to a large body of literature. Among others, it touches upon Domar's (1970) study of slavery and serfdom in farming in a pre-industrial setting. Domar, though, emphasizes the effect of the scarcity of labor relative to land, rather than of the transparency of farming.[24]

Our model is also closely related to the literature on property rights to land. Demsetz (1967) explains the emergence of private property rights on efficiency grounds, as a measure to resolve inefficiency due to externalities.[25] Following North (1981, pp. 7, 17, 22-23, 26, 28), in our alternative framework it is the state that specifies and enforces property rights, in order to maximize state revenue. In North's perception, securing the property rights of the non-elite serves as a commitment device by the state, designed to overcome the hold-up problem of ex-post expropriation in order to incentivize private investment. In our framework, asymmetric information (rather than investment hold-up) plays the key role. When transparency is high enough, it is in fact the threat of potential dismissal (implying lack of full property rights) that serves to incentivize the agent. On the other hand, given sufficient opacity (when the cost of erroneous dismissal and replacement of agents outweighs the benefits in incentives) the absolute state gives up the threat of dismissal and grants de-facto property rights. We are thus able to offer an explanation for cross regional differences in property rights to land.[26]

Finally our concern with governance and hierarchy recalls Coase's (1937) informational theory of the firm and Stiglitz's (1975) model of hierarchy and organization. The existing literature studies

[24] Following Domar, Lagerlöf (2009) explains the rise and decline of slavery on the basis of population dynamics. In the spirit of Domar, also Allen (1997) suggests that the lack of property rights to land in ancient Egypt was due to the relative abundance of land compared to the population. We note, however, that lack of owner-occupied farming persisted in Egypt until the nineteenth century (G. Baer 1962), when there was no scarcity of labor.

[25] In Demsetz' framework, communal land is divided into private land parcels by consensus (and without governmental involvement) in order to overcome its inefficient depletion by overcrowding. Dow and Reed (2013) offer a variant explanation for the emergence of (communal) property rights and for the creation of inequality during the Neolithic Revolution. They propose that after the adoption of agriculture, large enough coalitions of individuals who gained control over fertile land managed to exclude outsiders from sharing the land's rent. These elite insiders then employed the commoner outsiders (including their own expelled excess offspring) as wage laborers.

[26] Besley and Ghatak (2009) survey the literature on the determinants of the protection of property rights in a developmental contest. In a case close to the one studied here (their section 3.1.4) they examine a coercive authority that sets a rate of extraction to which it reputably commits overtime, in order to induce unobserved effort by agents. In another case (their section 3.1.6) they examine a static model, very similar to ours, were the authority may purchase a fully revelatory signal. In neither of these two cases do they consider the possibility of dismissal as an incentive device, nor the costs to the authority of employing that device. Thus they do not evaluate the possibility that an all-powerful authority may commit under some circumstances of opacity not to dismiss agents, and in effect grant them property rights to land, but to avoid such a commitment under other circumstances.

how the structure of hierarchy affects the flow of information and decisions within the organization (Sah and Stiglitz 1986).[27] Our multi-tier model is significantly simpler. It focuses on the effects of the extent of informational asymmetry on the allocation of power within different tiers of the hierarchy.

Thus, even though our model is motivated by an attempt to understand antiquity, it is very much in the spirit of the 'new institutional economics.'

3 An agency model with differential transparency

Following Olson (1993), we model the state as expropriating its subjects, without considering explicitly the possibility that state revenue may benefit the subjects through the provision of a public good.[28] The economy contains a given amount of land, divided into plots.[29] Each plot is allocated to one agent-tenant. We perceive the principal (the state) as an absentee landlord who designs a contract so as to maximize her expected periodic income. We first present a model with one tier of hierarchy. But since hierarchy plays a prominent role in our comparison between Egypt and Mesopotamia, we then extend our model to the case of two-tiered hierarchy.

The principal designs the contract. Her payoff is the output produced by all agents, less the payments to the agents and the costs of dismissal. Each agent decides how much effort to exert. His payoff is the payment received from the principal, less his cost of effort. We assume that the principal and the agent are both risk neutral. The assumption that farmers are risk neutral may seem problematic in the context of subsistence farming. In addition to the expediency of this assumption, in enabling us to obtain closed form solutions, risk neutrality is made more reasonable in our context since we explicitly assume that farmers' subsistence is guaranteed even in years of low output.

[27] The focus of the recent literature is on the design of hierarchy, accounting among others for the incentive problems inherent in the delegation of authority (Melumad, Mookherjee and Reichelstein 1995) and for communication and the acquisition of knowledge within the organization (Garicano 2000).

[28] This modeling strategy is not crucial for our arguments. Beyond the provision of security, which we take for granted since it serves the elite's interest, it seems as a rather reasonable simplification. Thus, Besley and Ghatak (2009, p. 4560) claim that starting at the 14th century CE, "Expropriations by government are a fact of historical experience"; Ma (2011) offers a similar perception of the state in imperial China.

[29] For now we take plot size as given, and as sufficient to maintain the tenant (and his family) even with low harvest. In appendix E we consider the plot size as endogenous.

3.1 The basic model

We assume that the output level produced by each agent can be either low or high: $Y \in \{L, H\}$, and that the agent's effort can also be either low or high: $e \in \{l, h\}$. The state of nature, too, is bi-modal and can be either good or bad: $\theta \in \{G, B\}$. Output is a function of the effort and the state of nature. In particular, output is high if and only if the state of nature is good and the agent exerts high effort:

$$Y = \begin{cases} H & \text{if } e = h \text{ and } \theta = G; \\ L & \text{otherwise.} \end{cases}$$

It is not evident how to set the timing of the effort relative to the realization of the state of nature.[30] We opt to assume that the agent chooses the level of effort before he learns the state of nature.[31] The ex-ante probability that the state of nature is good is denoted by: $p \in (0, 1)$. We assume that after the agent chooses the level of effort, both the agent and the principal observe a public signal: $\sigma \in \{\tilde{G}, \tilde{B}\}$ about the state of nature.[32] It is the accuracy of the signal which we take to represent the degree of transparency of production - our main exogenous variable. We identify the accuracy of the signal by a probability $q \in [1/2, 1]$, such that the conditional distribution of the signal satisfies:

$$Pr(\tilde{G}|G) = Pr(\tilde{B}|B) = q \ \ ; \ Pr(\tilde{G}|B) = Pr(\tilde{B}|G) = 1 - q.$$

The case of a perfectly revelatory signal is captured by: $q = 1$; and the case where it is uninformative is captured by: $q = 1/2$.

We denote the long-term periodic cost (in units of output) for maintaining the agent (and his family) intact into the next period, when he exert low effort, by $m \geq 0$. We assume that the long-term maintenance with high effort is $m + \gamma$, where $\gamma > 0$ is the periodic cost of exerting high effort. We assume that even if output is low, it is larger than the long-term maintenance cost with high effort: $L > m + \gamma$, and that it is efficient that the agent exerts effort: $p(H - L) > \gamma$.

[30] The relevant state of nature for agriculture is a vector whose components are distributed over the agricultural seasons – and so is effort. As is elaborated below, in Egypt the level of inundation is known before the agricultural year starts, but the spring heat waves occur towards its end. In Lower Mesopotamia the main natural uncertainties concern the magnitude of the (low) waters in winter, as well as spring heat waves and, particularly, the spring flooding. In rain-fed Upper Mesopotamia the main source of uncertainty is the idiosyncratic magnitude and timing of rains. In addition, there are of course all the other sources of natural uncertainties, like blight or locust or personal injuries.

[31] Appendix B examines the case where the agent learns the state of nature before exerting effort. We show that this weakly increases the payoff to the principal with no qualitative change to the model.

[32] The multiplicity of agents requires us to identify how the state of nature and the signal correlate across potential land plots. We address this issue in section 3.2.

The agent's only alternative employment outside of agriculture tenancy is perceived to be as a domestic servant outside of the farming sector, where his utility (in units of output) is normalized to zero. The agent's periodic utility, U, when engaged in agriculture equals his expected income, to be denoted by I, less the pertinent maintenance cost. In particular, when exerting high effort, this periodic utility is: $U = I - (m + \gamma)$.[33]

We denote the present value of the agent's utility from being employed in agriculture by V, and denote by $\delta \in (0, 1)$ the agent's discount factor. We presume that the agent has no additional sources of income or wealth, that he does not save from one period to the next (not even by storage) and cannot borrow.

The principal is assumed to rely on the following incentive scheme. If output is high, then the principal retains the agent with certainty and pays the agent (at harvest time) $\omega + b$, where $b \geq 0$ is a bonus payment. If output is low, then, independently of the signal, the agent is still paid a basic sustaining wage ω. This basic wage, we assume, has to sustain the agent (and his family) when he exerts effort, so that he will be employable in agriculture in the next period. It thus has to satisfy: $\omega \geq m + \gamma$.[34]

When output is low, however, if the signal indicates that the state of nature was bad ($\sigma = \tilde{B}$), the principal retains the agent. But if output is low and the signal indicates that the state of nature was good ($\sigma = \tilde{G}$), then the principal may suspect the agent of shirking and may replace him with another.[35] That is, we assume that the punishment that the principal can inflict on the agent upon suspected shirking is limited to dismissal.

We denote by $d = 1$ the strategy of dismissal upon low output and a signal indicating that the state of nature was good: $\sigma = \tilde{G}$, $Y = L$, retaining the agent otherwise, and by $d = 0$ the strategy

[33] The following analysis can be carried out also if we add to the utility function a fixed factor u_0 representing the psychic benefits for being with a family in agriculture. We assume that this utility level u_0 is sufficiently small, in particular, $u_0 \leq c(1-\delta)/p\delta$, for otherwise the threat of dismissal provides the principal with too great a leverage to control the agent and would complicate the analysis without adding any significant insight.

[34] By the nature of seasonal farming, this payment has to feed the agent and his family until the next harvest. Our assumption concerning the basic wage means that the agent is guaranteed survival even in bad states, and that his expected income is thus above the Malthusian threshold for keeping the population intact. This may be presumed to imply an increase in the farming population, but we assume that any excess of working population from the rural sector will be employed outside of farming, where the wage is low and does not fully guarantee reproduction. An alternative assumption that the basic wage ω has to cover only the maintenance cost m leads to qualitatively similar results as long as $u_0 > 0$.

[35] The replaced agent, we assume, still retains ω of the harvest, thus implying that the new agent has to obtain an advance ω to sustain him (with high effort) until the next harvest. We interpret the dismissed agent as one who has to leave the village forever, for employment as a domestic servant, where he may not be able to sustain a family. The new agent that the principal employs is presumed to be someone who has come of age and who would otherwise become a domestic servant.

of always retaining the agent. Thus we assume that the principal employs a pure strategy. She either dismisses the agent upon suspected shirking or not: $d \in \{0, 1\}$.[36] If the agent is dismissed, the principal incurs a fixed cost $x > 0$ (in units of output) representing the present value of lost output in training a new agent.[37] We assume that this cost is large enough to ensure that it will not be desirable to dismiss the agent when output is low ($Y = L$) and the signal indicates a bad state of nature ($\sigma = \tilde{B}$). This requires: $x > \delta p\gamma/(1-p)$. We describe the bonus, b, as a "carrot" and the dismissal, d, as a "stick." The employment contract in this environment can then be recast as seeking to incentivize through a balance between a carrot and a stick. This balance turns out to depend on the precision of the public signal q.

The above principal-agent problem is fully characterized by the following eight parameters: $(L, H, p, q, \gamma, m, x, \delta)$. The agent's value function V is determined as a function of these parameters and the endogenous contract variables b, d and ω. The fact that the agent is paid the wage ω in every period in which he is employed, regardless of the output produced, implies that under the optimal contract the principal would like to set ω as small as possible, so that $\omega = m + \gamma$. This implies that the employment contract can be fully described by two parameters only: $b \geq 0$ and $d \in \{0, 1\}$, the carrot and the stick.[38]

Given our normalization that the discounted utility of a dismissed agent is zero, in a stationary equilibrium the value of the employed agent's discounted utility, when he exerts high effort, has to satisfy:

$$V = [\omega + pb - (m + \gamma)] + [1 - Pr(dismiss|e = h)]\delta V. \tag{1}$$

For convenience, we denote the probability of a bad harvest and a good signal by $\mu = (1-p)(1-q)$. The probability of (wrongful) dismissal upon high effort is then $d\mu$. Given that $\omega = m + \gamma$, and this notation, V is determined by the contract parameters b and d and the parameters: μ, p and δ:

$$V(b, d) = \frac{pb}{1 - \delta(1 - \mu d)}. \tag{2}$$

[36] In appendix C we extend the model to the case in which the principal could randomize, where $d \in [0, 1]$ is the probability of dismissal upon suspected shirking ($\sigma = \tilde{G}, Y = L$). In appendix D we consider an alternative extension, were dismissal is binary, $d \in \{0, 1\}$, but the principal is allowed a series of warnings upon suspected shirking, prior to dismissal.

[37] This loss would include the advance payment to the new tenant for his labor in the following year and forgone output due to the loss of job-specific knowhow of the replaced tenant. In addition, x will include a measure of the extant by which the land's nutrients are likely to be depleted more by an agent subject to dismissal.

[38] One might argue that the principal has the incentive to renege, by avoiding paying the bonus to the agent, or by failing to dismiss him when specified by the contact. This possibility can be ruled out if we assume that the principal has a low time discount and realizes that the game is not only repeated, but is played simultaneously against many agents who are likely to believe that once reneging, the principal will always renege.

The principal's objective is to solve for the employment contract that maximizes her periodic expected payoff, denoted by π,

$$\pi = \underset{b \geq 0, d \in \{0,1\}}{max} p(H - b) + (1 - p)L - \mu dx - \omega,$$

subject to providing the agent with incentives to exert high effort:

$$\begin{aligned}&p(b + \delta V) + (1 - p)[q + (1 - q)(1 - d)]\delta V + \omega - (m + \gamma)\\&\geq\\&p(q(1 - d) + (1 - q))\delta V + (1 - p)[(q + (1 - q)(1 - d)]\delta V + \omega - m,\end{aligned}$$

where $V = V(b, d)$ as in (2).

Since $\omega = m+\gamma$, we can rewrite the principal's objective function (OF) and the agent's incentive constraint (IC) more succinctly as follows:[39]

$$\pi = \max_{b \geq 0, d \in [0,1]} p(H - L) + L - (m + \gamma) - pb - \mu dx, \tag{OF}$$

s.t.

$$pb + pqd\delta V(b, d) \geq \gamma. \tag{IC}$$

Given (2) and (IC) it is evident that the bonus b cannot be zero. It is also evident that the incentive constraint (IC) will be binding. By combining (2) and (IC) we can thus present the incentive constraint in the form:

$$b = b(d) = \frac{\gamma}{p}\left(1 - \frac{\delta pqd}{1 - \delta + \delta d(\mu + pq)}\right). \tag{3}$$

Had d been continuos, the constraint function $b(d)$ would be decreasing (and convex) in d, implying that, in incentivizing the agent, a stricter stick goes hand in hand with a smaller bonus.

3.1.1 Two Types of Contracts: 'Pure Carrot' or 'Stick and Carrot'

By selecting $d \in \{0, 1\}$ to maximize (OF) subject to (3), we obtain that the principal will design one of two contracts, depending on the degree of transparency, q. In particular, there exists a threshold level, $\hat{q}$, separating between the two contracts:

[39] The IC constraint implies that the agent benefits from exerting effort. To see that the left hand side indeed equals to the agent's benefit from exerting effort, note that its first component, pb, is the extra current expected income from high effort, while its second component is the current value of the decreased probability of dismissal. Note that according to a standard argument, the incentive constraint can be presumed to bind, since by decreasing b until it binds, the objective function can only increase.

if $q < \hat{q}$, it is a 'pure carrot' contract (subscripted by c), where:

$$d_c = 0,\, b_c = \frac{\gamma}{p}, \tag{4}$$

and

$$V_c = \frac{\gamma}{1-\delta}. \tag{5}$$

And if $q > \hat{q}$, it is a 'stick and carrot' contract (subscripted by s), where:

$$d_s = 1,\, b_s = \frac{\gamma}{p}\left(1 - \frac{pq\delta}{1-\delta\left(p+q-2pq\right)}\right), \tag{6}$$

with:[40]

$$V_s = \frac{\gamma}{1-\delta(p+q-2pq)}. \tag{7}$$

In the 'pure-carrot' contract, the agent is never dismissed and is incentivized only by the bonus. In the 'stick and carrot' contract, the agent is dismissed whenever output is low yet the signal is good ($Y = L$, $\sigma = \tilde{G}$). The selection of the type of contract depends on the degree of transparency.

At the threshold transparency $q = \hat{q}$, the principal will be indifferent between the two contracts. This implies that $\hat{q}$ is determined by the implicit condition:[41]

$$\frac{\hat{q}}{1-\hat{q}} = \frac{(1-p)\,x}{p\delta\gamma}[1-\delta(p+\hat{q}\ -2p\hat{q})]. \tag{8}$$

This quadratic equation in $\hat{q}$ has exactly one root strictly inside the unit interval $0 < \hat{q} < 1$. For some set of parameters, it also satisfies: $\hat{q} > 1/2$, meaning that there is a non-empty set of parameters for which the 'pure carrot' contract is optimal.

The switch of the optimal contract from 'pure carrot' to 'stick and carrot' when the quality of information improves captures the essence of our claims.[42] It is important to understand the logic behind it. A principal relying on a 'stick' to incentivize the agent has to incur the fixed cost of dismissal, x, whenever a dismissal takes place. Given that the agent is in fact incentivized to

[40] Since for any non-negative p and q in the unit interval: $p+q-2pq < p+q-pq \leq 1 \leq 1/\delta, V_s$ is guaranteed to be positive.

[41] At the threshold $\hat{q}$, $\pi_s(q) - \pi_c(q) = pq\delta V(b_s, 1) - \mu x = 0$.

[42] In the more general case of $d \in [0,1]$ examined in appendix C, there will be in general three zones. In addition to the two zones discussed here, there will be an intermediate zone where $1 > d > 0$. But the qualitative results on the effects of changes in transparency remain the same. These qualitative implications obtain also in the case of possible warning prior to dismissal examined in appendix D. In that case, as transparency improves the number of warnings increases. In the range of high transparency, the agent's bonus and value are just as in (6) and (7). And if the transparency is sufficiently low (and other supporting conditions hold), the number of warnings approaches infinity, meaning that the agent is never dismissed, and the agent's bonus and the value are as in (4) and (17).

exert effort, the actual dismissal upon suspected shirking is thus always in a sense erroneous. It occurs when the state of nature is bad and the signal is deceptive ($\theta = B$ and $\sigma = \tilde{G}$) with the probability $\mu = (1-p)(1-q)$. The probability of dismissal, and thereby the expected cost of dismissal, thus decrease when the quality of information, q, improves. Thus, when the dismissal is sufficiently costly to the principal – sufficiently large x – incentivizing the agent with the stick is worthwhile for the principal only when q is large enough. The threshold $\hat{q}$ balances between this expected cost of dismissal μx and the expected savings to the principal in reduced expected bonus payments $p(b_c - b_s)$ due to the threat of dismissal.

3.1.2 The effect of transparency on the allocation of income

Under a 'pure-carrot' regime, where $q < \hat{q}$, the expected income of the agent and the principal are:

$$I_c = m + 2\gamma, \tag{9}$$

and

$$\pi_c = p(H-L) + L - (m + 2\gamma). \tag{10}$$

And their combined expected income is:

$$I_c + \pi_c = p(H-L) + L. \tag{11}$$

It is readily observable that within the range of low transparency, where the principal refrains from ever dismissing the agent, the economy is efficient and the expected income of both the principal and the agent is independent of q. These features are lost in the range of high transparency. At this range of 'stick and carrot,' where $q > \hat{q}$:

$$I_s = m + 2\gamma - \frac{pq\delta\gamma}{1-\delta(p+q-2pq)}, \tag{12}$$

$$\pi_s = p(H-L) + L - (m+2\gamma) + \frac{pq\delta\gamma}{1-\delta(p+q-2pq)} - \mu x, \tag{13}$$

and

$$I_s + \pi_s = p(H-L) + L - \mu x. \tag{14}$$

The expected total income indicates that the 'stick and carrot' contract entails the inherent inefficiency of dismissing the agent even though he works diligently. This efficiency loss equals the expected cost of dismissal, μx, and declines as accuracy improves. When the signal is fully accurate ($q = 1$), the 'stick and carrot' regime become efficient. By construction, the principal's payoff is continuous at the threshold of transparency and increases with q thereafter. The gains to the principal from a rise in q above $\hat{q}$ are derived both from a rise in total income (as a result of the improved accuracy of the signal) and from a decline in the agent's income. Indeed, it is the agent who bears the entire burden of the 'stick and carrot' regime: at the threshold accuracy $\hat{q}$ his expected income I (and his expected periodic utility) drops by the expected cost of dismissal $\mu x = (1-p)(1-\hat{q})x$. After that threshold, his expected periodic income (and periodic utility) continues to decline with q. At this range, the benefit that the agent sustains due to the reduced probability of dismissal enables the principal to reduce the bonus payment b, while still maintaining the incentive constraint.

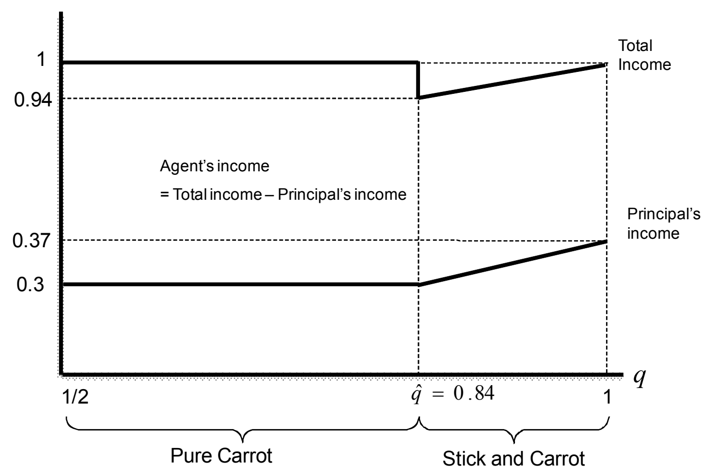


Figure 1: Periodic expected income as a function of signal accuracy

These features are summarized by Figure 1. The principal's expected income π as a function of accuracy q is depicted by the lower solid line. Total expected income $I + \pi$ is depicted by the upper solid line; with the balance between these two lines representing the agent's expected income. We attempt here a simple illustrative calibration of the model. The expected crop size of each plot is set to one: $E(Y) = pH + (1-p)L = 1$.[43] In particular, we set: $H = 1.1$, $L = 0.6$ and $p = 0.8$, so that a bad harvest that significantly reduces the crop occurs about every five years. To be consistent

[43] One should think of this unit as representing about 1.5 tons of grain of output, net of the grain that is needed for seed (typically assumed to be about 15 percent of the crop) and also net of expected spoilage in storage (typically assumed to be another 10-20 percent). For a more elaborate attempt to calibrate early Near Eastern farming see Hunt (1987), even though he neglects the labor input (our γ).

with tenants' output share of about two thirds and with the relative high cost of maintaining a family throughout the year, we posit that $m = 0.5$ and $\gamma = 0.1$. In addition, we set $x = 2$, so that the present value loss of replacing an agent is two expected crops. This includes the advance of $m + \gamma = 0.6$ to the new agent, as the dismissed agent will retain the low crop $L = 0.6$ when leaving the land. It should be noted that the same qualitative results are achieved with $x = 1$. However, in our robustness checks presented in the appendices a higher x is required for the sake of obtaining the full set of results, and thus for consistency we elect here $x = 2$.

Finally, to avoid giving the principal too much leverage over the agent by setting the discount rate too high, and as consistent with a prevailing interest rate (in grain) of one third or more, we set $\delta = 0.75$.

It is instructive to compare the outcome when the signal is fully accurate ($q = 1$) with the outcome when the signal is highly inaccurate ($q < \hat{q}$). As seen in Figure 1, in both cases the economy is efficient, since the diligent agent is never dismissed. However, the distribution of income is quite different. When the signal is fully accurate, the bonus that the agent requires in order not to shirk is minimal. Indeed, the agent's (gross) income falls from $I_c = m + 2\gamma$ in the range of the opaque signal to: $I_s = m + 2\gamma - p\delta\gamma/[1 - \delta(1 - p)]$, when $q = 1$. The agent's utility – income net of effort and subsistence costs – from being engaged in Agriculture, $U = I - (m + \gamma)$, in this case dissipates entirely if he were fully patient ($\delta = 1$).[44]

Let us summarize our findings thus far in an informal proposition:

Proposition 1

(a) If the transparency level is sufficiently low, the agent-tenant is in a 'pure carrot' regime, in which he is never dismissed. In such a regime the agent can be considered as a de-facto owner of the land that he cultivates. While he may not be able to sell the land, he might be able to transfer the right to cultivate it to another. This de-facto ownership is not due to any explicit legal rights, or to the benevolence of the principal (the state) or to any impediments that prohibit dismissal. Rather, the agent's de-facto rights to the land stem from the fact that, given the low transparency, it is optimal for the principal to refrain from ever dismissing the agent.[45]

[44] We note that the agent remuneration in both states is independent of the output levels L and H, and that the effort cost γ and the replacement cost x were taken as though independent of L and H. We address these issues partially in Appendix E, where we generalize the model to include an endogenous population and an endogenous plot size.

[45] The range of 'pure carrot', however, may be empty ($\hat{q} \leq 1/2$). Its existence depends on a combination of sufficiently high level of x, and sufficiently low discount rate δ. The latter is required, since when the agent cares a great deal about the future, the potency of a threatened dismissal increases.

(b) If the transparency level is sufficiently high, then the agent-tenant is in a 'carrot and stick' regime. In this regime, when a farmer produces low output he is subject to dismissal, and thus cannot be considered to have ownership rights to the land that he cultivates. In that range, as transparency increases, the principal relies more on the 'stick' and less on the 'carrot'. That is, under the optimal 'carrot and stick' contract, when the accuracy of the signal that the principal obtains is higher, even though the probability of dismissal declines, the threat of dismissal provides a stronger incentive (as the probability of a random unjust dismissal is smaller), and the tenant's share of output is lower, while the state's revenue is higher.

These findings, and in particular the general effect of transparency on the optimal combination of the 'stick' and 'carrot', are robust and do not depend on our specific modeling assumptions. The credible threat of using the 'stick' of dismissal reduces the cost of incentivizing the agent with the 'carrot' as more of it would be required otherwise. However, to maintain credibility of the 'stick', dismissal must be used (even if unjustifiably), whenever output is low and the signal is good. Since the probability of such dismissal declines with transparency, it follows that the expected cost of including a 'stick' in the contract declines with transparency. It is the fixed costs of dismissal, and our assumption of a bimodal dismissal probability $d \in \{0, 1\}$ that accounts for the existence of a threshold transparency level $\hat{q}$. As transparency increases beyond this threshold, the stick becomes an ever more effective incentivizing mechanism, since increased transparency reduces the probability of dismissal when the agent exerts high effort and increases the probability of dismissal if the agent exerts low effort (off equilibrium). This logic also provides the basis of our prediction that sufficient opacity protects the agent, granting him both higher income and de-facto land ownership.

3.2 A Two-Level Hierarchy Model

We now extend our model to two tiers of government. Extending the model further to n tiers of hierarchy along the same lines is rather straightforward. For the relations between the upper echelon (the king) and lower level of hierarchy (the village governor), we elect to employ the model from Appendix A, where the governor may hide output, rather than exert a low level of effort. For the relations between the governor and the farmers in the village under his control we keep the basic model.

We now attach a subscript 1 or 2 to variables in each level of the hierarchy, from the bottom up.[46] We assume that there are two independent state variables that determine the state of nature

[46] In order not to clutter the presentation we omit additional indices for separate plots in each village or for separate

in each plot of land: $\theta_1 \in \{G, B\}$ at the level of each plot of land, and $\theta_2 \in \{G, B\}$ at the level of each village. The plot-specific states are assumed to be independent across plots, and the village specific states are assumed to be independent across villages. We denote by $p_1 \in (0, 1)$ the probability for each plot of land to sustain a plot-specific good state, and by $p_2 \in (0, 1)$ the corresponding probability for the entire village.

As in the basic model, output in each plot can be either low or high: $Y_1 \in \{L_1, H_1\}$ and the agent's effort can be either low or high: $e \in \{l, h\}$. Plot output is assumed to be high if and only if the agent exerts high effort and the plot's 'total' state of nature is good. Here we extend our former analysis by assuming that the state of nature in a plot is good if and only if both its plot-specific state variable and its village specific state variable are good ($\theta_1 = \theta_2 = G$). Thus, the state of nature in a specific plot is good with probability $p_1 p_2$, and otherwise it is bad.[47]

We assume that the village governor and the farmers know the village specific state of nature, θ_2, after the decision whether to exert effort is made. In addition, if the village specific state is good ($\theta_2 = G$), the governor receives plot-specific signals σ_1 for each plot, regarding its specific state variable (θ_1). These signals are accurate with probability $q_1 \in [1/2, 1]$ and are assumed to be independent across plots.

At the higher level of the hierarchy, between the village governor and the King, we assume an analogous information structure. The king does not know the specific states θ_1 of individual plots, nor the states θ_2 for any of the villages. But he receives an independent signal σ_2 about the latter, whose accuracy is denoted by the probability $q_2 \in [1/2, 1]$.

The contract selected by the governor will have the same structure as in the basic model. It pays a basic wage ω_1, which covers the subsistence cost of the tenant m_1 and the cost of effort γ: $\omega = m_1 + \gamma$. In addition it pays a bonus b_1 if the farmer's output is high. The governor could also incentivize the agent by a threat of dismissal $d_1 \in \{0, 1\}$ upon suspected shirking, at a cost of x_1 to the governor. Suspected shirking occur if output is low ($Y_1 = L_1$) but is anticipated to be good – that is, when the village state is good and the plot specific signal is good ($\theta_2 = G, \sigma_1 = \tilde{G}_1$). Otherwise, the agent is retained. Thus, subject to the incentive constraint holding, the agent is dismissed (even though he exerts high effort) with probability: $Pr(dismiss|e = h) = (1 - p_1)\, p_2\, (1 - q_1)\, d_1$.

villages.

47 An alternative (perhaps more natural) modeling formulation is that the village state of nature determines the probability of success of individual plots in the village in a conditionally i.i.d. way. That is, if the village is in a good state, individual plots produce high output with a high probability (provided high effort, of course) and if the village is in a bad state then individual plots produce high output with a low probability (which is zero in the model we do analyze). This formulation, however, will render the analysis significantly more complicated, without any effect on the qualitative results.

That is, the agent is dismissed when the village state is good, the plot state is bad, and the signal is misleading.

In analogy to (2) above, the value for the tenant of being employed is:

$$V_1(b_1, d_1) = \frac{p_1 p_2 b_1}{1 - \delta_1 \left(1 - \left(1 - p_1\right) p_2 \left(1 - q_1\right) d_1\right)}, \tag{15}$$

were δ_1 is the farmer's time preference.

The governor's objective is to find the employment contract $\{b_1, d_1\}$ that maximizes her periodic expected payoff, now denoted by π_1,

$$\pi_1 = \max_{b_1 \geq 0, d_1 \in \{0,1\}} p_1 p_2 \left(H_1 - b_1\right) + \left(1 - p_1 p_2\right) L_1 - p_2 \left(1 - p_1\right) \left(1 - q_1\right) d_1 x_1 - \omega_1, \tag{16}$$

subject to providing the agent with incentives to exert high effort (noting that $\omega_1 - (m_1 + \gamma) = 0$):

$$\begin{aligned} & p_1 p_2 b_1 + \left(1 - \left(1 - p_1\right) p_2 \left(1 - q_1\right) d_1\right) \delta_1 V_1 \\ & \geq \\ & \left[p_1 p_2 \left(q_1 \left(1 - d_1\right) + 1 - q_1\right) + \left(1 - p_2\right) + p_2 \left(1 - p_1\right) \left(q_1 + \left(1 - q_1\right) \left(1 - d_1\right)\right)\right] \delta_1 V_1 + \gamma, \end{aligned} \tag{17}$$

where $V_1 = V_1(b_1, d_1)$ as in (15).

We can rewrite the agent's incentive constraint (17) more succinctly, and with equality, as follows:

$$b = \gamma / p_1 p_2 - q_1 d_1 \delta_1 V_1(b_1, d_1). \tag{18}$$

This maximization problem is a minor variant of the principal's problem in the basic model, in which $p_1 p_2$ substitutes for p as the probability of a high output, and in which the probability of dismissal is $p_2 \left(1 - p_1\right) \left(1 - q_1\right) d_1$ instead of $(1 - p)(1 - q)d$. Thus, we do not report the outcome in detail. The governor will choose a 'pure carrot' contract ($d_{1c} = 0$) if transparency is below some threshold $\hat{q}_1$, and a 'stick and carrot' contract, if transparency is sufficiently high. Above $\hat{q}_1$, the expected income of the governor will increase with q_1.[48]

We turn now to the study the king's problem. For that purpose we assume that the number of plots in each village is sufficiently large so that the effects of the plot-specific states on the total village revenue can be substituted by its expected value. The revenue obtained by the village governor is then limited to two possible outcomes only, depending on the village-specific state of

[48] The corresponding bonus payments become: $b_{1c} = \gamma / \left(p_1 p_2\right)$ with $V_{1c} = \gamma / (1 - \delta_1)$; and $b_{1s} = \gamma / \left(p_1 p_2\right) - q_1 \delta_1 V_1(b_{1s}, 1) = \gamma / \left(p_1 p_2\right) \{1 - p_1 p_2 q_1 \delta_1 / [1 - \delta_1 (1 - p_2) - \delta_1 p_2 (p_1 + q_1 - 2 p_1 q_1)]\}$. If $p_2 = 1$, this reverts to the results of the basic model. We note that when $q_1 = \delta_1 = 1$, then $H_2 - L_2 = N_1 p_1 (H_1 - L_1)$.

nature θ_2. We denote by L_2 and H_2 the governor's income in a village-bad year (when $\theta_2 = B$) and good year (when $\theta_2 = G$) respectively. If N_1 is the number of plots in a village, then:

$$\begin{aligned} L_2 &= N_1 \left[L_1 - (m_1 + \gamma) \right], \\ H_2 &= H_2(q_1) = N_1 \left[p_1(H_1 - L_1 - b_1) + L_1 - (1 - p_1)(1 - q_1) d_1 x_1 - (m_1 + \gamma) \right]. \end{aligned}$$

The parameters b_1 and d_1 are those selected by the governor, determined as above, by the local transparency level q_1 and the additional parameters (like x_1 and δ_1). Beyond the threshold $\hat{q}_1$, the good-year revenue H_2 thus increases in q_1.[49]

Recall that the king receives a signal σ_2 regarding the village state θ_2, whose accuracy is denoted by q_2. The game played at this stage is that the governor may underreport the revenue collected. That is, she may report collecting L_2, even though she in fact collected H_2. The king is assumed to employ an analogous two-edged incentive scheme to the one above: a bonus b_2 if the governor reports collecting H_2, and a threat of dismissal $d_2 = 1$, at a cost of x_2 to the king, if the governor's report is that the revenue is L_2 but the signal σ_2 indicates that a good village harvest is to be expected.

Let m_2 be the maintenance cost of the governor, and assume $\omega_2 = m_2$. As in appendix A, the king maximizes:

$$\pi_2 = \max_{b_2 \geq 0, d_2 \in \{0,1\}} p_2(H_2 - b_2) + (1 - p_2)[L_2 - (1 - q_2) d_2 x_2] - m_2.$$

Under the presumption that the incentive constraint for the governor to report revenue truthfully is binding, we obtain:

$$b_2 = (H_2 - L_2) - q_2 d_2 \delta_2 V_2,$$

where $\delta_2 V_2$ is the discounted value of the governor for keeping her position in the next period. Under the assumption that the governor considers losing her office to have a value of zero, we obtain in analogy to (1) in the basic model:

$$V_2 = p_2 b_2 + [1 - d_2(1 - p_2)(1 - q_2)] \delta_2 V_2,$$

from which one can obtain $V_2(b_2, d_2)$ as in (2), and then solve explicitly for the incentive scheme $\{b_2, d_2\}$ (with $d_2 \in \{0, 1\}$), that will be selected by the king.

[49] We note that when $q_1 = \delta_1 = 1$, then the king obtains all surplus and $H_2 - L_2 = N_1 p_1 (H_1 - L_1)$.

Thus, if the village farming is sufficiently opaque to the king (and subject to some supplementary parameter restriction, among others on x_2 and δ_2), the governor enjoys a carrot regime, in which she is autonomous in the sense that she is never dismissed:[50]

$$if\ q_2 < \hat{q}_2,\ \text{then: } d_{2c} = 0; b_{2c} = H_2 - L_2; \pi_2 = L_2 - m_2.$$

In this regime, the king's revenue is independent of the state of nature; the governor retains for herself the maintenance cost m_2, and in addition also the balance $H_2 - L_2 = b_{2c}$, whenever the state of nature is good.

On the other hand, when the village farming is sufficiently transparent to the king ($q_2 > \hat{q}_2$), a stick and carrot regime prevails: $(d_2 = 1)$. In this case, the governor is dismisses whenever the king is led by the signal to expect high revenue, but the governor reports that the collected revenue is low. This occurs with probability $(1 - p_2)(1 - q_2)$.[51] Following the analysis in Appendix A, in this regime:

$$\text{if } q_2 > \hat{q}_2,\ \text{then: } d_{2s} = 1; b_{2s} = (H_2 - L_2) - q_2\delta_2 V_{2s},$$

where

$$V_{2s} = \frac{p(H_2 - L_2)}{1 - \delta_2(p + q_2 - 2pq_2)}.$$

The king's expected revenue in this case becomes:

$$\pi_2 = (L_2 - m_2) + pq_2\delta_2 V_{2s} - (1 - p)(1 - q_2)x_2.$$

The threshold transparency level $\hat{q}_2$ guarantees the continuity of the king's objective function, and is determined by the implicit condition:

$$\frac{\hat{q}_2}{1 - \hat{q}_2} = \frac{(1 - p)x_2}{p\delta_2 V_{2s}},$$

where V_{2s} is evaluated at $\hat{q}_2$.

As in the basic model, this transparency threshold increases with the costs of dismissal x_2 and decreases with the governor's discount factor δ_2.

[50] While the patience parameter δ_2 may be expected to be higher for the governor than for the individual agents, the dismissal costs x_2 may be expected to be much higher as well.

[51] We assume that the individual farming agents do not save from year to year and cannot borrow. This assumption is not likely to hold for the governor. In that case, if the governor has an indication of the king's signal, she may be able to avert being dismissed by supplying the shortfall of the revenue expected from her from alternative sources. This possibility opens up complexities that we prefer not to go into.

In the range of high transparency, the king's revenue, once again, increases with transparency. As in the main model, if the economy is very transparent ($q_2 = 1$) and the governor is very patient ($\delta_2 = 1$), the state bureaucracy becomes extremely lean: the governor receives no bonus ($b_{2s} = 0$) and retains only m_2, while the king receives all the rest of the revenue collected from the farming sector. If, however, the economy is nontransparent to the king, the governor becomes autonomous, in the sense that she is not subject to dismissal, and that she retains a larger fraction of the wealth extracted from the farmers.

We summarize our findings in this section in an informal proposition:

Proposition 2

In analogy to the relationship between the farming agent and the principal land owner, the balance of power between the central king and the provincial governor depends on the transparency of the provincial economy to the king.

(a) When the local conditions are sufficiently opaque to the center, the intermediary governor enjoys substantial autonomy in that she pays a (relatively low) fixed tribute and is, otherwise, not subject to dismissal.

(b) When the transparency of the local provincial economy to the king is sufficiently high, the governor is subject to dismissal, upon delivering lower revenue than the level expected of her. The governor also retains a relatively lower share of the revenue collected, and when transparency is very high and the governor is very patient, she retains little beyond her minimal maintenance costs.

3.3 The Urban Sector

Our model, we argue, captures the essence of hierarchies in general and of the ancient agricultural state in particular. Two items were nevertheless taken as exogenous, even though they are clearly not so. We have emphasized geography, but hardly addressed the issue of space. The first missing item is the size of each farmer's land plot, which in turn is related to the number of farms and farmers in any geographical location. We extend the model to include endogenous plot size in Appendix E. The second missing item is the span of control of each governor, or the number of plots under her supervision. We plan to expand the model to include this element in future work. At this stage, we seek to derive additional insights from our model concerning the urban sector.

We consider all those who are not engaged in farming as belonging to the urban sector. To simplify, we assume further that the urban sector does not trade with the farming sector. That is, the provision of protection and the collection of tribute ('protection' revenue) is the only interaction between the two sectors. We also simplify here by reverting to a single tier of government, thus equating the governor with the king. The food collected by the governor is evidently not consumed entirely by her. This food revenue provides the means for maintaining the guards who provide protection to the farming sector (ensuring the governor's monopoly in extracting revenue). This food sustains also the artisans who supply various amenities (including luxury items) for the governor and his dependents, and may possibly serve to exchange for prestige goods from abroad. Since some of the food that reaches the urban sector is in some sense wasted on sumptuary meals or on imports, the ratio of the average food collection to the food required for long-term maintenance (m) can give us an upper estimate of the size of the urban sector that is supported by the farming sector.[52]

More significant than the relative sizes of the two sectors is the very different uncertainty in food supply that they face. The essence of this issue can be clarified by considering what happens in bad years. At the level of the individual farmer bad years occur with probability $1-p_1p_2$. At the governor's level, however, they occur less frequently, conditioned on a bad village-year ($\theta_2 = B$), and thus with a lower probability of $1 - p_2$. This reflects the fact that the governor's revenue bundles together the revenue from many independent plots, and thus provides an insurance against idiosyncratic plot bad states. However, our model also identifies a difference in the severity of bad harvests due to village bad states. In this case, by our assumptions the output of each farmer is L_1, and the revenue collected by the governor is $L_2 = N_1 [L_1 - (m_1 + \gamma)]$. It so happened that in the numerical calibration that was selected above: $L_1 = m_1 + \gamma$. This implies that the remuneration retained after a bad harvest enables farmers to survive until the next harvest, but the governor and the urban sector obtain no revenue at all. This extreme result is clearly due to our simple model and to this particular calibration; but it reflects a general phenomenon: a larger share of the farming output remains in the periphery after bad harvests. This is likely to capture another important and ill-understood aspect of ancient economies. Without attempting to specify the exact conditions in our model under which the severity of bad harvests is higher for the urban sector, we would like to consider the implications of this phenomenon. Proposition 3 summarizes our conclusions from this section:

Proposition 3

The urban sector is likely to be more vulnerable to downward shocks to output. This implies

[52] If farmers are employed (during the summer) in building monuments, and are paid for their extra effort by the state, as was customary in Egypt, this too would have to be taken into account.

that hunger and starvation are likely to be concentrated particularly among the lower strata of the urban sector: servants, small artisans and the like. This implication is in line with our presumption that this segment of society is demographically vulnerable, and may not have reproduced on its own, other than through an inflow from the farming sector. In addition, under the circumstances assumed here, the vulnerability of the urban sector implies that whereas farmers need only store food within the year, inter-annual storage is an absolute necessity in the urban sector, as a buffer for years when the harvest is low. This inter annual storage, however, should not be considered as providing insurance for the farming sector, but rather as serving the urban sector.[53]

4 Application

Our goal in this section is to demonstrate that the theoretical insights obtained in the preceding section are consistent with the institutions of the major civilizations of the ancient world. We shall focus on three regions: Egypt, Upper Mesopotamia and Lower Mesopotamia, and on the fourth to the second millennia BCE, although we believe that our analysis is applicable to other regions and other periods as well.

In 1961 ninety four percent of the population in Bangladesh was rural; and as late as 1980 eighty nine percent of the population in Ethiopia was engaged in agriculture.[54] A similar pattern must have prevailed in antiquity. Yet, as scholars are well aware of, our archaeological and textual sources about antiquity pertain almost exclusively to the urban centers and to the elite. The reconstruction of society from these sources is thus heavily biased towards the perspective of the urban elite. As argued by Hicks (1969), theory may thus be necessary in any attempt to offer a more balanced reconstruction of antiquity. We view our contribution in this spirit. This means that, among others, our paper has little to say about the role of religion and related ideologies that serve to legitimate authority. Still, we argue that our model provides a useful framework within which one may better understand key social institutions that persisted over many centuries, even millennia, and that help to distinguish between the different civilizations of antiquity.

We briefly summarize the distinctive environmental features of the farming technology in each region, as a prelude to examining its social institutions. Even though agriculture originated in the proximity of northern Mesopotamia and reached the Nile valley much later, we shall start with

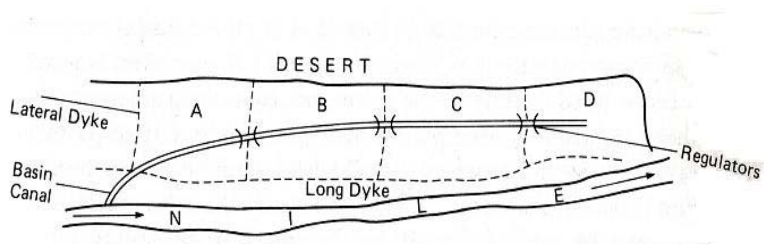
[53] This conclusion is consistent with the predominant archaeological finding of storage pits and granaries in ancient urban centers, but is inconsistent with the common presumption (see for example Adams (1981, p. 244; 2005), cited above), that urban central storage served the entire population and was possibly the main service that the state provided to the countryside.

[54] FAO statistics (http://faostat.fao.org); 1980 is the earliest year with data on agricultural population.

Egypt, for the simple reason that from our perspective of transparency, Egypt presents a polar case.

4.1 Egypt

Agriculture in Egypt originated away from the Nile valley, and was adapted into the valley in Upper (southern) Egypt only in the late fifth millennium BCE.[55] From this relatively late start, and from a core in Upper Egypt, a central state was formed along the entire length of the Nile at about 3000 BCE. The distinctive farming technology that underlined Egyptian agriculture in the Nile valley is known as 'flood basin irrigation.' It prevailed with relatively minor modifications for more than five millennia until the construction of the first Aswan Dam at the beginning of the twentieth century.[56] Figure 2, depicting an independent system of four connected flood basins, assists in understanding the essential features of this agricultural method. The four basins in this figure are sealed to the north (right side) by a natural protrusion of the desert that also bounds the basins to the west. The basins are separated from the river itself by the river's elevated natural levee, and are separated one from another by man-made lateral dykes. Each of these basins could have been populated by several villages, located at the desert's edge, on the river's embankment, or on raised natural mounds within the basin that turned into islands during the inundation season.[57]



[55]The term Upper Egypt refers to the Nile valley south of Cairo; Lower Egypt refers to the Nile delta; the term Middle Egypt is sometimes applied to northern Upper Egypt.

[56]Our description of the flood basin system is based on Butzer (1976) and on Willcocks's (1899) detailed account on the irrigation system that operated in Upper Egypt in the late nineteenth century. The land in the Delta (Lower Egypt) was also irrigated through flooding, but slightly differently. These two areas account for the bulk of Egypt's cultivable land (Butzer, p. 83). Two additional areas were irrigated without flooding: the Fayum depression, and the raised land on the Nile's riverbanks and along the desert margins. The latter land was irrigated by extended low gradient longitudinal canals (parallel to the river) or by water hauled from the Nile. These longitudinal canals were developed and extended by far-sighted Pharaohs in the second and first millennia BCE, and facilitated the cultivation of orchard gardens and two annual crops.

[57]The width of the Nile valley varies from about two kilometers in southern Upper Egypt to about eighteen kilometers in Middle Egypt. In areas where the valley was wide, longitudinal dikes, parallel to the river, were set up to make the basin width manageable. According to Willcocks, in the 1880s there were 212 basins in Upper Egypt, with an average size of about 27 square kilometers. Based on Hassan (1994, p. 165), about 1,400 families would have populated a basin of this average size in ancient times.

Figure 2: A schematic flood basin system typical for Upper Egypt. Source: Lloyd (1983, p. 327), adapting Willcocks (1899, p. 610).

The Nile receives its water mainly from the early-summer monsoon rain in eastern Africa. When its water level rose sufficiently during the inundation period (in mid-August), the local inhabitants would breach the Nile's embankment to open up the basin canal and flood the basin system (see Figure 2). The basins would then be filled up sequentially, from the lowest (D) to the highest (A), where after each basin was filled, the inlet regulator was sealed. This created a system of terraced pools, each one covered by about 1.5 meters of standing muddy water. The water was retained in the basins for about 40 days, depositing its rich mineral nutrients and soaking the soil. Then (in early October), after the Nile receded, the basin system would be gradually drained back into the Nile by opening the escape channel that leads from the lowest section (D) to the river.[58] After the soil dried somewhat, the peasants would plow and sow their fields (mostly with barley). The moisture trapped in the soil served as the sole source of water during the growing season. Five months later (in late March) the farmers' bottleneck season arrived, when they had to harvest the ripened crops before the hot winds of April and May could parch the grain stalks and cause the seeds to disperse to the ground.

The unpredictable annual fluctuations in the Nile's inundation level necessitated adjustments to the regular mode of operation. With low inundation levels, some higher basins would not be filled up sufficiently and food production would falter. Occasional high inundation levels could be even more damaging, overflowing the riverbanks, destroying the dykes and occasionally flooding the villages and spoiling the locally stored grain.[59] Even in normal years, the operation could be disrupted, for example if one of the lateral dykes was breached, or if the spring scorching winds arrived too early. Still, it is evident that in any given year, the conditions that farmers faced were fairly homogeneous within each basin, and even across basins.

The above description explains the significant role of the village headmen. He had to organize the villagers to perform the necessary tasks of opening and sealing the basin dykes during the inundation season, as well as cleaning the canals of silt and maintaining the dykes in the dry

[58] Willcocks (1899, p. 59) explains that holding the water in the basins for too long engendered worms that destroyed the crops and delayed the ripening of the grain. The drainage of the water back to the Nile also had the effect of flushing away the salt that would be naturally created in irrigated soil. This, in combination with the fertilization by the deposited nutrients enabled perennial cultivation without the need to fallow the land.

[59] Normally, the slow pace of water increase during the inundation would have enabled farmers to take precautions and move their property to the outer margins of the valley if they foresaw potential flooding. But if dykes collapsed, the villages on the mounds within the flood basins could be flooded all of a sudden, with the peasants losing their belongings, and, in particular, their private stores of grain for the entire ensuing year.

season, after the harvest. This description explains also the critical role of the local chieftain who supervised the entire basin system, determined the timing of the inflow of the water and its outflow, and initiated earthwork to repair and improve the basin system. All these decisions, and the attendant earthwork, had to be determined, organized and supervised locally, and had to be performed cooperatively by the local peasants, sometimes with very short notice.

The local-public good nature of the basin irrigation system explains why local leadership was essential, with the power to extract labor services on an on-going basis from the farmers. It is thus likely that already the earliest basin systems operated under local leadership. The nature of the system must have served these local leaders also to gradually amass power and to become the mayor-kings of incipient city-states.[60] After only a few centuries of feuding city-states in southern Upper Egypt in the late fourth millennium BCE, Egypt was united under a single ruler. The description above clarifies, however, that even after a central state was formed, the management of the flood basins could not have been centralized and had to remain in local hands. Some substantial irrigation projects that extended intensive agriculture into the higher periphery of the Nile valley were probably initiated by subsequent Pharaohs. However, this hardly supports Wittfogel's thesis that a strong despotic central government was an essential prerequisite for the farming system of Egypt. In fact, the above description explains that central management (even at the district level) was almost irrelevant for farming in Egypt. This accords with Eyre's (1994, pp. 72-73) dismissal of the "received view" that "a strong centralized bureaucracy exercising detailed and uniform control over the productive resources [of Egypt]."[61]

As we noted, farming activity throughout Egypt was conditioned to a major extent on one variable: the maximal level of the inundation, which became known in mid-September, prior to the agricultural year. As a result, farming activity was highly transparent to the central government, irrespective of the substantial ex-ante non-predictability of the Nile's inundation. Scholars have long noted evidence that records of the Nile's level of inundation were kept already in the third millennium BCE (see Kemp 2006, p. 64). This was not a record for the sake of history; it must have served as a control device for the Pharaohs. The most detailed description of how taxes were levied in Upper Egypt is available, however, only from much later periods. Cooper (1976, p. 366)

[60] The narrowness of the Nile valley in southern Upper Egypt may explain why the basin system was apparently invented there, and why it was there that the earliest city-states were formed in the late fourth millennium – since relatively short lateral dykes were required. In addition, it may be that in the fourth millennium BCE the water table in the Middle and Lower Egypt was still high, causing swamps and impeding the penetration of agriculture.

[61] Taking it to the extreme, Eyre (1997, p. 375) quotes, approvingly, an expert's assessment of the management of the water system in modern Egypt: "The central state pretends to regulate everything and in fact regulates nothing." For a similar view of the role of the central government see Butzer (1976, p. 43).

describes the taxation of agriculture in the middle ages: "Agriculture was so well regulated in Egypt that, on the basis of the Nile flood recorded by the Nilometer, the government knew in advance what revenue to anticipate." In particular, "The height of the Nile flood determined how much and in what manner the tax assignments were made in each district." We conjecture that this was generally the case also in antiquity.[62]

The high transparency of the state of nature in every district (Nome) and in every flood basin system within each district, explains why the Pharaohs could be so powerful, without any direct involvement in the management of the farming sector. As long as the Pharaohs monopolized physical power and could dismiss subordinates at will, they could effectively siphon off much of the tax revenue and run a highly lean state bureaucracy – consistent with Proposition 2. Indeed, at least in the early part of the Old Kingdom, the positions of governors and state bureaucrats were by a revocable appointment, and non-hereditary.[63] The absolutist power of the Egyptian Pharaohs from early on may explain also why they, unlike their counterparts in Mesopotamia, were considered as god incarnates.[64]

As Eyre (1994, p. 74) summarizes, the Egyptian central state's ability to tax was closely aligned with the decentralized organization of farming: "The crucial factor for the central power was its ability to enforce fiscal demands and political control. ... [P]ower lay in control over the ruling class ... not in the detailed administration of the individual peasantry."

The leanness of the intermediating bureaucracy explains, in turn, why the capital cities of the different districts remained essentially administrative centers and did not amass substantial independent wealth to rival the predominance of the capital city. This, we argue, explains Wilson's above cited observation concerning Egypt's anomaly as "a civilization without cities."[65]

[62] The transparency of Egyptian farming was due not only to the observation of the height of the inundation in advance of the cultivation season, but also to the relative ease of monitoring farming activity in real time by traveling by boat along the Nile. Evidence that this was done in practice is provided by the Wilbour Papyrus (see Kemp 2006, p. 256a) that contains the minutes of a monitoring expedition from about 1140 BCE. This expedition recorded rent assessments for more than 2,000 large plots of land in Middle Egypt that were owned by temple institutions.

[63] Baines and Yoffee (1998, p. 206) write: "The king's most powerful influence was probably on the elite. Their status and wealth depended on him – often on his personal favor and caprice." Eyre (1999, p. 48) wrote similarly that in the Old Kingdom "the government appears to be an elite overlay" above the villages (p. 48), consisting of official appointees who were charged with channeling tax revenue to the center.

[64] The Mesopotamian kings (with a single exception of Sargon in early Akkad) were only considered as envoys of the gods (Baines and Yoffee 1998). If the first Pharaohs were brutal in enforcing their power, religion, ideology and custom may have subsequently substituted to some extent the need to employ direct coercion. This may partially explain why later Pharaohs were willing to share some of their wealth with the temples. The temples became substantial landowners, but posed no threat for the Pharaohs. In various periods, particularly since the Persian occupation in the mid first millennium BCE, large private estates proliferated as well.

[65] Redford (1997, p. 216) provides an alternative explanation for the absence of multiple rival cities in ancient Egypt.

Scholars have been puzzled by the rapidity of the formation of a strong central state in Egypt and by its remarkable subsequent stability. We believe that our model provides the explanation for these puzzles: Egyptian farming was relatively easy to control from afar, and to raise taxes from, since it was highly transparent at all levels of hierarchy.[66] Indeed, the above description of the basin system clarifies not only how the Pharaoh could efficiently extract tax revenue from the provinces, but also how each governor could determine how much tax was due from the local chieftain who supervised individual irrigation basin systems. Likewise, the homogeneity within basins enabled the local chieftain to have effective control over the revenue received from the village headmen. Even when local idiosyncratic accidents occurred, like a collapsed dyke, these were clearly known at the local level. That is, in terms of our model, farming activity was highly transparent at the local level (q_1 was very high). In addition, the homogeneity of the land within the basins and the commonality of the infrastructure implies that there was little field-specific know-how or capital.[67] This implies that the costs of replacing any individual farmer (x_1) were rather low. The agrarian system at the local level thus fits very well the 'stick and carrot' regime of our model.

It was the village community as a whole that was responsible for paying taxes. For that purpose, the village headman exercised tight control over the village land and in effect determined the land assignments of the individual farmers. By custom, though, the same fields were assigned to the same farmer annually, or to his heir. But this assignment was conditional on diligent work. In particular, farmers did not have any secure tenure and the village head or the estate manager (in the case of temple land) could reassign fields as they saw fit.[68]

Scholars have often debated the extent of private land ownership in ancient Egypt. It is well recognized that a notion prevailed all along that the entire land of Egypt belonged to the Pharaoh

He attributes it to the special historical circumstances of Egypt's rapid and brutal unification that annihilated the opposition from the very beginning. However, this hysteresis explanation can hardly account for the fact that ancient Egypt remained essentially a one-city civilization even after some kings relocated their capital cities, and even after what are known as the "intermediate periods" in which Egypt was temporarily divided between rival kings.

[66] We note two different meanings of the term 'control.' The first, in the sense of on-going management of the irrigation system, as is invoked by Wittfogel, is irrelevant for Egypt's Pharaohs. The second, in the sense of subjugation, or the power to command and extract, captures the essence of Egypt's supposed uniqueness.

[67] The transparency of farming and the fact that the land could not in effect be over-exploited are manifested in that in the known few cases (from the mid-first millennium BCE) where private land was leased to tenants, the lease contracts were for one year only (Hughes 1952).

[68] See Eyre (1997, p. 378; 1999, pp. 51-52). G. Baer (1969, p. 17) reports that this state of affairs prevailed still in the early part of the nineteenth century CE. Baer argues that the Egyptian village (particularly in Upper Egypt) was characterized by three features: (a) village land was periodically redistributed by the village-head to the peasants; (b) the village was collectively responsible for tax payments; (c) the village as a whole was held responsible also for maintaining irrigation work and providing labor for public works.

(Baines and Yoffee 1998, p. 206). At the same time, it has been established that a significant fraction of the land was de-facto "owned" in various periods by landlords other than the state. This land typically consisted of large estates by the temples, by various lay organizations and by powerful noblemen.[69] From our perspective, the more significant feature of the land tenure system in Upper Egypt is that even when the land was private, it was owned by absentee landlords, and not by the tenant-peasants who cultivated the fields. This was the case, we argue, in accordance with Proposition 1, because the high transparency of farming robbed these peasants of the main protection that is otherwise offered by the existence of moral hazard.[70]

The lack of transferable and inheritable title to land and the hierarchical nature of land management explain in turn the limited extent of real-estate transactions in ancient Egypt – at least in comparison to Mesopotamia.[71] The absence of legal title to land implied also that land could

[69] Hughes (1952, pp. 1-2) summarizes that in the first two millennia of the historic period (up to the late New Kingdom, or about the first millennium BCE) there was never "a large body of small landholders who managed and worked their plots themselves ... the lowest classes were largely serfs on the domains of Pharaoh, the wealthy and the temples." He also states: "in theory all the land belonged to Pharaoh throughout Egyptian history;" but then notes that much land belonged to the temples and that, at various periods, estates were assigned to court officials and to military officers. Manning (2003, pp. 65-98) surveys the land tenure regime in Upper Egypt; but his main focus is on the Ptolemaic period.

[70] It was apparently indignation over the fact that Egyptian farmers did not own the land that they cultivated that led the authors of the Old Testament to refer to Egypt as a "house of bondage," even though slavery was no more pervasive in ancient Egypt than elsewhere in the Ancient Near East. Pharaonic Egypt thus seemed to the Israelites as the epitome of a 'bad' society. Eyre (1997) contends that the divorce between land-ownership and actual farming was endemic to Egypt and persisted essentially until the mid-twentieth century. Legal scholars often note that the concept of 'private ownership' refers to a bundle of rights. To the extent that the custom was that peasants were able to continue to cultivate the land that was allotted to them in the past, or were able to transmit this privilege to their sons, they had some form of 'ownership' (that conferred value); and if they could sublet that land it was yet another such right – even if land remained inalienable – see Ghalwash (1998). G. Baer (1962, pp. 1-70; 1969, pp. 62-78) contends that land became private in Egypt only during the nineteenth century (though this claim is somewhat disputed). Following the elimination of private tax farming, a series of legislative measures aimed to place the responsibility for tax payments directly upon individual cultivators. These measures established peasants' right (that were until then but a custom, and mostly in Lower Egypt) not to be evicted from the land that they cultivated and to transmit it to their male heirs, or even to sell it – as long as the taxes were paid to the state. Land for which taxes were in arrears, or where the cultivating peasant died without a mature male heir, reverted to the state and was then reassigned to someone who was able to pay the taxes – see Ghalwas (1998). G. Baer (1962, 1969) reports about the wide-ranging effects of the redistribution of land and the establishment of legal title to private land. Among others, these measures eroded the village community and the traditional power of the village heads. On the other hand, even though these measures led initially to the diffusion of land holdings, by the end of the nineteenth century, these measures led, in fact, to the concentration of land in the hands of major private landholders who leased the land to peasants. Baer attributes this re-agglomeration of land to the commercialization of farming activity. We argue that, in addition to whatever economic advantage there was in agglomeration, it was aided by the particular weakness in Egypt of the key disadvantage to land concentration: the moral hazard inherent in the relationship between the absentee landlord and the cultivating tenant.

[71] Using records on land sales from the New Kingdom and later, K. Baer (1962, p. 25) contends "private individuals could own farm land at all periods of ancient Egyptian history." But the limited extent of the evidence that he was

not be used to secure loans, and explains the paucity of early records for land-secured loans. These factors, and the little surplus that remained in the countryside (including in the regional administrative centers) explain the lesser extent of commercial activity – once again, in comparison to Mesopotamia. The village leadership resolved local legal disputes within the village, applying traditional common law (Eyre 1999, p. 44). Disputes among those higher up in the hierarchy were similarly resolved by the authority just above the disputants. These features help explain why the earliest known law codes from ancient Egypt are from the second half of the first millennium BCE, whereas such codes were promulgated in Mesopotamia about two millennia earlier.[72]

Finally, we seek to examine the applicability of our third proposition: that long-term interannual storage was particularly important in the urban center; that it primarily served the urban sector itself, rather than the countryside; and that it were the urban poor who were the most vulnerable to starvation. The archaeological evidence on extensive granaries and storage in the capital city and in the temples is well known (Kemp 2006, pp. 257, 322, 356-360). This evidence, though, does not address the issues of whether these granaries supplied grain to the peasants in the countryside in years of famine as a form of insurance, as some scholars contend (see above); nor does it negate our prediction on the particular vulnerability of the urban poor. We are not familiar with direct evidence on these issues. According to the insurance hypothesis, however, years of shortage were ones in which the center was particularly strong relative to the periphery, as is illustrated by the biblical story about Joseph's central food storage and the subsequent distribution of food during the prolonged years of famine (Genesis 47).[73] According to our contrasting scenario, the opposite happened: in years of severe overall shortage, the share of total food that remained in the countryside increased, shifting the balance of power away from the center to the periphery. As indirect evidence in support of our prediction, we note that the few episodes in which the central state in Egypt disintegrated were apparently associated with periods of sustained shortages (see Butzer, 2012).[74]

able to marshal pales in comparison to the abundant data on land transactions from Mesopotamia (this cannot be attributed solely to perishabilty of the papyrus records, given the abundance of other textual evidence from Egypt).

[72] For surveys of the legal institutions of ancient Egypt and Mesopotamia see Westbrook (2003).

[73] Redford (1970) characterizes the biblical story of Joseph (Genesis 37-50) as a novelette that was composed in the mid-first millennium BCE by someone with little familiarity with Egypt. The pertinent verses in Genesis 47:13-26, he argues, were added even later and were not part of the original Joseph story (pp. 168, 182).

[74] Ascribing the abrupt collapse of the Old Kingdom to extremely low flood levels (around 2180 BCE), Hassan (2007) quotes a contemporary report on the resulting breakdown in governmental control. That source describes how tax revenue ceased to reach the Pharaoh's palace and the seats of the provincial governors, leaving the craftsman unemployed and causing rebellion in the towns and disintegration of the country through infighting.

4.2 Mesopotamia

The economic and social institutions in Mesopotamia were significantly different from Egypt's. Upper (northern) Mesopotamia was mostly rain-fed. In Lower (southern) Mesopotamia, agriculture was based on the rivers; however irrigation was not by flooding of the alluvial plain, as in Egypt, but rather by canals in elevated sloped areas on the banks of abandoned past river courses.

Wilson (1960, p. 128) provided an insightful comparison of the irrigation systems in Egypt and Lower Mesopotamia. He compared the basin irrigation system along the Nile to leaves that grow out directly out of a single "stalk of grass or bamboo," and suggested that the settlement pattern in (Lower) Mesopotamia "might be likened to a tree, with the river as the trunk and the canals as strong branches thrusting out from the trunk." Figure 3 (from Wilkinson 2003) presents a similar attempt to encapsulate the different settlement patterns within Mesopotamia, between the rain-fed north (a) and the irrigated south (b).

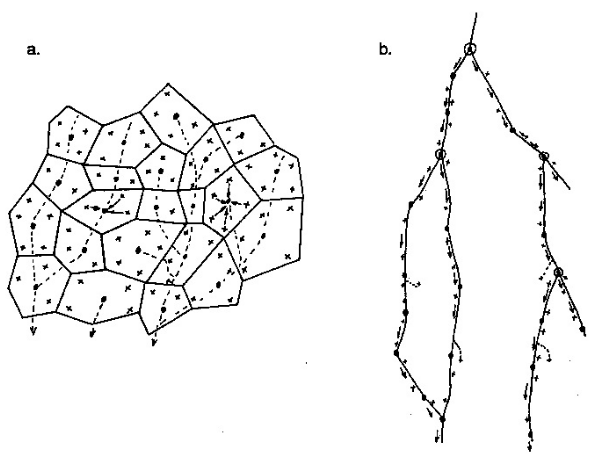


Figure 3: Schematic Settlement patterns in Upper Mesopotamia (a) and Lower Mesopotamia (b); Solid circles represent central settlements; crosses represent minor settlements (villages). Source: Wilkinson (2003, p. 211).

From our perspective, Upper Mesopotamia can be considered as closer to the polar case of opacity, while irrigated Lower Mesopotamia (south of modern Baghdad) presents an intermediate case with high transparency only at the local level. We shall consider the latter region first.

4.2.1 Lower Mesopotamia

Agriculture in the arid alluvial plains of Lower Mesopotamia (Babylonia) depends entirely on irrigation from either the Euphrates or the Tigris.[75] Since the Euphrates and the Tigris are fed by winter rains and by spring melting snow from the mountains of modern Turkey and Iran to the north and east, both rivers have low water levels in October-December, and both swell in April-May. This seasonal pattern posed two major problems of mismatch.[76] First, the winter and early spring sowing and cultivation had to be conducted when the rivers were low. This required extended canals and also mechanisms for the distribution of water, when water was in short supply and could not irrigate all the potentially cultivable land.[77] Second, the cultivated land was endangered by spring flooding, exactly at harvest time.[78] To avoid the latter danger it was necessary to divert the excess spring water away from the fields into the plain at the lower end of the cultivation zone.

It was the difficulty in overcoming these two problems that apparently delayed the adoption of extensive agriculture in Babylonia until long after agriculture flourished in Upper Mesopotamia and irrigation systems were established in southwest Iran (Wilkinson 2003, pp. 72-76).[79] But, once irrigation agriculture was introduced, it led to a relatively rapid development of civilization. More than thirty major city-states have been identified in Babylonia in the fourth and third millennia BCE. The largest of these cities was Uruk. It was at Uruk at about 3200 BCE, when its population reached about twenty thousand, that writing is believed to have originated (Yoffee 2005, p. 43).

[75] Within Lower Mesopotamia scholars often distinguish between northern Babylonia (also identified as Akkad) and southern Babylonia (Sumer).

[76] See Adams (1981, pp. 3-6) and Postgate (1994, p. 178).

[77] Adams (1981, p. 6) estimates that due to the shortage of water, only 8,000-12,000 square kilometers could be cultivated out of a potential estimated by Wilkinson (2003, p. 76) at about 50,000 square kilometers. Thus, unlike rain-fed Upper Mesopotamia, a potentially arable piece of land could have been worthless without an attached water allocation. This emphasizes the power of those who could deny water. The relative abundance of land may explain also the prevalence of the rotating fallow in Lower Mesopotamia.

[78] In addition, harvesting in April-May was subject to the similar hot parching winds that affected Egypt.

[79] During the seventh to the fifth millennia, the Persian-Arab Gulf extended more than two hundred kilometers inland. Much of Sumer was then covered by estuaries and swamps, and rudimentary agriculture was conducted in dispersed settlements on natural mounds. As argued by Nissen (1988, pp. 55-56), increased aridity during the early fourth millennium and the receding water table due to the retraction of the Gulf's transgression may help explain the relatively late extension of extensive farming into Sumer.

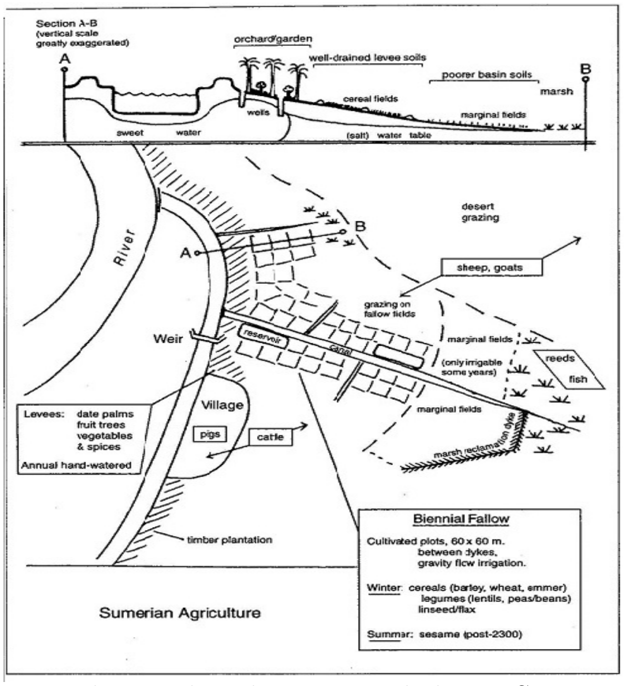


Figure 4: A schematic irrigated agricultural area in Babylonia; Source: Wilkinson (2003, p. 92), following Postgate (1994, p. 175).

Figure 4 presents a schematic reconstruction of an agricultural zone in Lower Mesopotamia.[80] Farming was critically facilitated by the elevated riverbed of the Euphrates relative to the nearby alluvial terrain. The fields were irrigated by an intricate system of canals that diverted water from the elevated river to the outer slopes of the river's levees and to the levees of abandoned past courses of the river. Date orchards and vegetables were grown on the raised levees of the canals. Feeding canals directed water to the fields, where cereals were cultivated. In spring time, these feeding canals had to be turned off and the spring floods were directed to the swampy alluvial plain below the cultivated fields.[81]

Liverani (2006, p. 28) acknowledges the transformational nature of the "secondary agricultural

[80]Postgate (1994, p. 175) points out that the gradient in the cross sectional diagram is grossly exaggerated. According to Wilkinson (2003, p. 77) the levees rise by only 2-3 meters above the flood-planes. The square fields indicate that what is reconstructed here is northern Babylonia, since in Sumer the fields were elongated, stretching all the way from the feeding canal to the marshland (see below).

[81]See Adams (1981, p. 245) and Wilkinson (2003, p. 89). This pattern may explain not the priority of the Euphrates over the Tigris, but the priority of Sumer over Akkad in the early millennia. As Nissen notes (1988, pp. 144-145), "the Tigris cuts deeply into the land," while "the river [Euphrates] was especially low-lying in northern Babylonia," flowing "through a relatively narrow plain." In contrast, irrigation from the Tigris, and from the Euphrates up north, required long canals to capture the river water upstream and channel it to the cultivable areas in a gradient lower than the river's. These irrigation projects were more elaborate, and probably required organized public initiative, rather than simpler local entrepreneurship.

revolution" of locally administered irrigation agriculture in Lower Mesopotamia. But somewhat surprisingly, he does *not* attribute the "urban revolution" to irrigation. Rather, he attributes it to subsequent technical agricultural innovations: the cultivation of elongated fields with a deep plowing technique, employing a seeding plow that was pulled by oxen.[82] Steinkeller (1999) offers a similar view. These eminent scholars argue that the introduction of these farming techniques was a source of substantial economies of scale that contributed not only to higher productivity, but also to the centralization of farming. The claim of economies of scale, though, seems to us implausible. In analogy to our proposed explanation of the rapid state formation in Upper Egypt, we argue that it was the introduction of irrigation farming itself, and its high transparency and amenability to control, that provides the key explanation for the formation of the early city-states in Lower Mesopotamia.[83]

As Figure 4 illustrates, farming in Lower Mesopotamia was significantly more intricate than the flood basins of Upper Egypt. Unlike the homogeneity of fields in Egypt, in Lower Mesopotamia even fields within the same zone varied in quality, depending mostly on how high they were above the saline water table that was created by the adjacent marsh. Unlike in Egypt, where the salt that accumulated in the topsoil was flushed away naturally when the flood basins were drained, salination of the topsoil in Babylon necessitated the fallowing of the fields every alternate year.[84]

[82] See Liverani (2006, pp. 14-19). The diminished significance attached to the role of irrigation reflects Liverani's explicit aversion to Wittfogel's ideas, which he describes as "notorious" and "abnormal" (2006, p. 15). Potts (1997, pp. 70-89) provides a useful summary of Mesopotamian farming techniques. The cultivation of narrow and very long fields (sometimes 1.5 kilometers long) – unlike the square fields in Figure 4 – and the plowing and sowing of these elongated fields using a pair of oxen was prevalent in Sumer. This cultivation method provided several important advantages that contributed to the increased productivity of Sumerian farming. The long fields were arranged with the narrow side next to the feeding canal and with the length of the field sloping down towards the marshy plain. Deep furrows were plowed in the length of each field with wide spacing between furrows. Using a seeder plow (instead of the standard broad-throw method) led to seeding only within these deep furrows. This saved seed and also enabled the efficient utilization of scarce water during the cultivation season, since only the furrows had to be watered. The deep furrows implied also that salinity of the topsoil was concentrated mostly away from the seeds.

[83] Algaze (2001, 2005, 2008) similarly emphasizes the relatively high productivity of agriculture in Sumer; but he downplays the significance of irrigation and agriculture altogether and attributes the "takeoff" of the city-states in Sumer mostly to trade. Along with Adams (2005), we are not swayed by his arguments. The long-distance imports of Uruk served mostly the local elite and gained prominence only after the city was well established. We doubt also that the early Sumerian cities served as trading hubs for exchange between local farmers and fishermen. It is more likely that these early urban centers served primarily as control hubs for the local elite. In particular, Algaze's arguments fail to explain the source of the elite's ability to extract surplus from the countryside.

[84] Another recurring difficulty emanated from the pastoralist inhabitants of marginal grazing land and the swamps. The existence of pastoralists in their vicinity was probably a mixed blessing to the farmers. It may have contributed to the farmers' welfare through trade and through supply of labor in the peak harvesting season, but it also exposed the farmers to the threat of banditry. The latter consideration may explain the phenomenon that Adams (1981, pp. 137-145) identifies as the "implosion" of the rural settlements in the late fourth millennium, when the rural farming

In particular, cultivation in Lower Mesopotamia depended absolutely on rationed water that was controlled upstream and could be directed elsewhere.[85] It was the elite families who managed the water supply and controlled the water direction at various canal junctures. This form of control, in turn, enabled the elite to extract surplus from the cultivating peasants. In terms of transparency, we thus interpret the farming activity in Lower Mesopotamia as highly transparent to the local elite.

We contend that this high local transparency explains why powerful early city-states, controlled by elite families, were able to form in Sumer. According to our first proposition, this transparency explains also why owner-cultivated farming was practically nonexistent in Sumer. Just as in Upper Egypt, cultivation in Sumer was conducted by peasant families of sharecroppers who were controlled by a hierarchy of managers, under the ultimate control of the dominant elite families. This system has been described as "a pyramid of individual families" (Steinkeller 1999, p. 293) or as an "institutional household" (Renger 1995). Scholars have long maintained that the bulk of the agricultural land of the Sumerian city-states was owned, at least nominally, by each city's temple. The temple affairs, and the cities, were in turn managed by a coalition of the dominant elite families. The conditions of the cultivating peasants in Lower Mesopotamia were thus rather similar to those in Egypt. But, unlike in Egypt, the land was essentially privately owned by members of the elite (sometimes through the city temples) in the urban centers.[86]

The prevalence of elite-owned private land in Lower Mesopotamia is closely related to other features that sharply distinguish between this region and Egypt.[87] As noted above, the operation of the complex irrigation system in Lower Mesopotamia required skilled managers with "thorough knowledge of local conditions on a day-to-day basis" (Hunt 1987, p. 172). As a result – unlike

population sought to reside in the walled cities (Yoffee 1995, p. 54). The rural sector apparently recovered during the reign of Akkad and of the Ur III dynasties that could better control the pastoralists (Steinkeller 2007).

[85] The shortage of water at the critical cultivation season is evidenced by the use of irrigation fees, already in the late third millennium BCE (Ouyang 2010).

[86] There is extensive literature on the extent of land ownership in Sumer by private families (Renger 1995). From our perspective, the important observation is that these private lands were not owned by the actual cultivators. The land tenure system in the northern part of Lower Mesopotamia (Akkad) was a mixture of the one that prevailed in Sumer and the one in Upper Mesopotamia, where the temples played only a minor role in agriculture. The land tenure system in Northern Babylonia in the third and early second millennium was examined by Renger (1995), Steinkeller (1999) and Goddeeris (2002). This region attained dominance by the first millennia (due possibly to the changing course of the Euphrates in Sumer and the increased salinity of the irrigated land there.) The records of numerous land transactions from the second and first millennia attest to the significant extent of private land ownership in Northern Babylonia. Jursa (2010) provides a comprehensive overview of the Babylonian economy in the first millennium, including its land tenure institutions.

[87] The different institutions in Egypt and Lower Mesopotamia indicate that the ease of navigation offered by the Nile could not have been the prime source of Egyptian uniqueness.

the case of Egypt were there was less need for specialized knowledge – the local managing elite in Lower Mesopotamia were indispensible and irreplaceable (high x_2 in our extended model). In other words, we interpret the farming activity in Lower Mesopotamia as transparent to the local elite, but as relatively opaque to remote potential authority. According to our second proposition, we contend that this explains why the local elite in the urban centers in Lower Mesopotamia were extremely powerful and resilient. This explains why strong cities were one of the most distinctive features of the Mesopotamian civilization – unlike Egypt. In turn, this also explains why several aggressive attempts to unify Lower Mesopotamia under one of the rival city-states ended in failure – once again, in marked distinction to the quick and durable unification of Egypt.

The rival city states of Lower Mesopotamia fought each other periodically for a millennium, before they were first consolidated under Sargon of Akkad in about 2350 BCE. Sargon's central state lasted, however, less than two centuries and started to disintegrate well before that. In about 2100 BCE another territorial state was formed, under the third dynasty of the city of Ur. This highly oppressive and bureaucratic central state lasted only one century, before it too crumbled. The next territorial state was established by Hammurabi of Babylon in about 1800 BCE, but it too did not last long. Thus, until the first millennium, Mesopotamia was ruled most of the time by rival, city-states, with only brief intermittent periods of a central territorial state.[88] We contend that our second proposition helps account for this apparent non-viability of central territorial states in Lower Mesopotamia. Our perspective suggests that this was because the early central rulers in Lower Mesopotamia could not dispense with the local elites in the city-states that they managed to concur. Unlike Egypt, where knowledge of the inundation level provided the Pharaoh with effective control over all the districts along the Nile, Lower Mesopotamia did not offer such a control mechanism to emasculate the power of the local elites. We note that, in contrast to Wittfogel's theory, in both Egypt and Lower Mesopotamia the maintenance and the management of irrigation was conducted at the local level. It was, however, the real-time specific knowledge of the local elites in the Mesopotamian cities that made them indispensible. This analysis is consistent with Yoffee's (2005) description of the fate of Sargon's earliest central state. According to Yoffee (2005,

[88] Starting in the first millennium BCE, the successive empires of Assyria, Babylonia and then Persia, like the subsequent Greeks and Romans, developed administrative methods that managed to subject formerly independent city-states; yet, even under these empires, these cities typically retained much of their former autonomy. Thus, the Mesopotamian cities in the first millennium BCE continued to conduct their internal affairs under an assembly of the local citizens, and were managed by a local mayor who served alongside the appointed governor (see Van de Mieroop 1997, pp. 128-139). We do not pretend to explain here the secret for the success of these empires, but venture to suggest that the issue of how tax revenue was funneled to the center (and the switch to the remittance of taxes in coins that started with the Persians) provides a key perspective for understanding the viability of these empires.

p. 37), Sargon of Akkad was well aware of the intermediation problem when he ascended to power, and sought "to disenfranchise the old landed aristocracy." But, after conquering the diverse city states in Lower Mesopotamia, he ruled them nevertheless through appointed "royal officials, who served alongside the traditional rulers of the conquered city-states" (2005, p. 142). It was this "uneasy sharing of power ... [that] led to a power struggle" and to the ultimate demise of Sargon's territorial states (Yoffee 1995, pp. 292-293; 2005, p. 143).

4.2.2 Upper Mesopotamia

Due to the uncertainty and idiosyncratic nature of rainfall and the relative unevenness of the terrain, farming in the highlands of Upper Mesopotamia was rather opaque.[89] Early urbanization was identified in Upper Mesopotamia during the late fifth and fourth millennia BCE, but was somehow aborted in the later part of the fourth millennium – when the first city-states started to flourish in Sumer.[90] Wilkinson (1994; 2003, p. 211) concludes that the settlement pattern in Upper Mesopotamia was generally characterized by a scatter of a large number of roughly equivalent, nucleated administrative units with a radius of about five kilometers around each one (see Figure 3). Each unit was administered by a central settlement that relied on agricultural revenue from its satellite villages, within a radius of control determined by the "constraining effect of land transport and the convenience of being within one day's round trip of the center" (1994, p. 503). Wilkinson (2003, p. 211) attributes this nucleated settlement pattern to the fact that no site had an "overwhelming situational or demographic advantage." Without disputing the factual basis of this observation, we contest the logic behind this argument. By the winner-takes-all (increasing returns to scale) nature of violent conflicts, a-priori advantage is not a prerequisite for the formation of larger central states under those leaders who happened to defeat their neighbors. From our perspective, the key to the nucleated pattern of semi-autonomous administrative units in early Upper Mesopotamia was the inability of the winner of any territorial conflict to subjugate and extract on-going revenue from distant conquered areas. The localized nature of the early states in this region can be explained, we posit, with our second proposition. That is, due to the opacity

[89] See Wilkinson (1994) and Jas (2000). We restrict our attention here to Upper Mesopotamia, but the opacity of its agriculture was probably less pronounced than that in other rain-fed regions of the Ancient Near East. Noy-Meir (1973) demonstrates the effects of extreme spatial variations in micro-climate and in terrain quality on desert plant populations. Mediterranean economies in the arid hilly regions, like that of ancient Israel, are clearly closer to this extreme end of the spectrum. The early institutional evidence on owner-occupied farming and on the weakness of states from these regions accords with the implications of low transparency in our model.

[90] The large size of these early cities, and the architectural remains of the dwellings, suggest that these cities were inhabited not only by the elite, but also by the farming peasants (Ur 2010).

of farming activity (low q_2) and a high dismissal cost (x_2), the span of control of early city states in Upper Mesopotamia was rather limited.

Extensive cuneiform evidence about farming in Upper Mesopotamia is available only from the mid-second millennium BCE, from the vicinity of Nuzi. This evidence reveals that, unlike the situation in Lower Mesopotamia, the temples did not possess much economic power, but the kings and the elite owned large estates. Much land was also owned by nuclear families who worked their patrimonial land. The evidence from Nuzi reveals in fact that land ownership in Upper Mesopotamia was in a constant flux, with small landholders regularly losing their ancestral fields to rich families through debt and sale under duress (Zaccagnini 1999; Jas 2000). The persistence of owner-occupied farming in Upper Mesopotamia indicates, however, that the process of land consolidation due to indentured farmers losing their patrimonial land must have been matched by an opposite process of the gradual dissolution of large, presumably less efficient, estates.[91] The prevalence of owner-cultivated private farming in Upper Mesopotamia is consistent with our first proposition, given the presumed low transparency of farming in that region.

Thus, we interpret the distinctive institutions of Upper Mesopotamia in the fourth to second millennia BCE as consistent with our first two propositions.

5 Application to the modern growth of the state

In this section we argue that our analysis of how the tax technology affects state institutions is applicable not only to antiquity but also to the modern phase of human history. According to our theory, the transition to agriculture was ultimately responsible for the emergence of the state because it facilitated appropriation of food by non-producers. A similar theory, we contend, may explain the sharp and unprecedented increase in the scope of the state (as a share of total product) in the past century and a half, following several millennia in which the share of government has been about five to ten percent.

The main explanations in the literature for the increase in the scope of the state since the nine-

[91] The Nuzi records from the mid-first millennium reveal that existing laws sought to preserve patrimonial land, and in particular, the sale of private land to non-relatives was apparently forbidden. This is evidenced by many extant cuneiform contracts in which land sales were disguised as "adoptions," where the seller adopted the purchaser as a son, and then became a tenant on his former land (Zaccagnini 1984). A similar pattern of land tenure with significant owner-occupied farming persisted throughout the rain-fed regions of the ancient Near East. Jas quotes Warriner (1948, pp. 21, 104) who noted that the older land tenure regimes in North and South Mesopotamia persisted to the modern era: "In the north, the forms of tenure are similar to those of Syria, with a class of small proprietors taking some but not all, the land. In the south large owners or sheiks own virtually all the land, letting it to share-tenants, through a series of intermediary lessees."

teenth century concern public goods and state-provided redistribution of income. Starting with Wagner ([1883], 1958), mainstream public finance theory attributes the growth of the government's share of output to increased demand for public goods (which is presumed to rise more than proportionately when per capita income increases). On the other hand, the political economy literature emphasizes the redistributive element of government spending (Meltzer and Richard 1981). It is argued that the dominant interest groups in democracy tend to increase the scale of government in order to redistribute income towards the median-income voter (whose income is lower than that of the voter with average income).

Both of these explanations apply to the modern world the type of functionalist explanations that were proposed to explain the emergence of the state in antiquity (section 2.1). Underlying both of these explanations is a premise that increased democratization in recent centuries has led governments to give more attention to the public's welfare. There is little doubt, indeed, that the expansion of the state in recent centuries coincided with the expansion of democracy. The perspective in this paper, however, leads us to wonder what could have led the autocratic monarchs of previous centuries to curtail the state's tax revenue (as a share of total income), in comparison with the democracies that replaced them. Indeed, it is customary to assume (as we do) that the autocratic governments of antiquity were Leviathans who sought to maximize long-term revenue. It is unlikely that the appetite of the Bourbons or the Romanovs for tax revenue was less than that of the Pharaohs of ancient Egypt. Holding all else constant, we would have thus expected that democratization and greater attentiveness to public welfare would have reduced the relative burden of taxation, rather than the reverse.[92]

This argument suggests that the democratization explanation for the increased scope of the state, taken on its own, conceals the crucial role of the Industrial Revolution and its induced transformation of the tax technology. The shift away from agriculture almost eliminated the significance of the environmental factors that are so prominent in our theory of the extractive capacity of ancient states. However, the mechanism of increased transparency that we emphasize here is applicable, we argue, to understanding how the Industrial Revolution induced a transformation of the tax technology, which then led to the increased scope of the state.

The idea that a shift in the tax technology was the root cause for the modern growth of government is not new. It accords with Olson's non-teleological theory of the state and with

[92] Our argument against the democratization explanations accords with the findings by Mulligan, Gil and Sala-i-Martin (2004) who found no significant difference between modern democracies and non-democracies in the pattern of economic and social policies.

Parkinson's (1960, p. 3) Second Law: "[State] expenditure rises to meet [tax] income." This idea accords also with Kau and Rubin's (1981) proposal that the growth of the modern state should be attributed to a decline in the cost of collecting taxes, due to increased market production and the related increase in record keeping. We prefer to describe this increased efficiency of the tax technology as resulting from the increased transparency of production. The latter was due in part to the shift to mass production by hired labor in large corporations – a shift that was accompanied by a massive accounting paper trail (see Kleven, Kreiner and Saez 2009).[93] This paper trail exposed productive activity to the state and transformed the state's ability to tax, among others by turning private companies into efficient tax collection agencies, and by facilitating the taxation of income. This explanation accords very well with our analysis of how increased transparency of production contributes to state power.

Thus, we contend, our theory that changes in the scope of the state are due to increased transparency of production applies both to antiquity and to the modern world.

6 Concluding thoughts

Stigler (1961) argued that "knowledge is power." We apply this adage to theorize about the role of information asymmetry in shaping the institutions of early state societies. We contribute thereby to the recent literature on the deep rooted factors that play a role in comparative development. Our overarching contention is that through its effect on the tax technology, the transparency of production is a major causal determinant of the scale of the state, its hierarchical structure, and land tenure arrangements.

Based on this theory we conduct a comparative analysis of the salient institutions of ancient Mesopotamia and Egypt. In particular, we argue that the rapid rise of the powerful central state in Egypt, its subsequent resiliency, the weakness of its cities and the lack of land-owning peasantry can all be explained by the high transparency of Egyptian farming, both at the local level and at the state level. The same paradigm also explains key institutional differences between Upper Mesopotamia, where transparency was low, and Lower Mesopotamia, where it was high at the local level but much lower at the central level.

[93] Peacock and Wiseman (1961), Tilly (1990) and Gennaioli and Voth (2010) argue that tax capacity increased since the Middle Ages (and the Commercial Revolution), and particularly in the twentieth century, due to the necessity of financing wars. This theory, though, fails to explain why wars throughout history prior to the Middle Ages did not increase the relative scope of the state. The two approaches for explaining the modern growth of taxes and government can be reconciled in that increased transparency served to augment the tax potential, and wars served as a trigger for utilizing that increased potential.

We conclude by placing our geographic theory of early institutions in the perspective of alternative attempts to understand antiquity. Reflecting the ideas of Polanyi (1944) and Finley (1999), anthropologists, archaeologists and historians of antiquity used to maintain that the ancient economies were fundamentally different from modern ones, and that economic theory, with its focus on the central role of markets, cannot be applied to the study of antiquity. This common perception has eroded somewhat in recent years, with increased application of the ideas from the new institutional economics (see Bang 2009). However, the underlying grounds for the earlier rejection of the applicability of economic theory to the study of the ancient world have not changed. The role of markets and private property – which is still paramount in the new institutional economics – was apparently rather limited in antiquity. Thus, Van de Mieroop (1999, 2004), while acknowledging the applicability of modern economic theory, downplays the significance of markets: "the fundamental difficulty lies in determining the relative importance of exchange through a market within the totality of the ancient Mesopotamian economy" (1999, pp. 117-118). He notes, instead, the significance of hierarchical, non-market relations, and in particular the role of "public institutions of temples and palaces" in determining the distribution of resources (1999, p. 118).[94] North (1977, p. 706) concedes similarly: "Economic historians have not even begun to account for such non-market allocative systems, and until they do, they can say very little about societies in which markets had very little allocative effects." North contends further: "Karl Polanyi's challenge must be met head on if economic history is to provide us with improved insights about our ancient past," and suggests: "transaction cost analysis is a promising analytical framework to explore non-market forms of economic organization" (pp. 707, 709).

Our analysis of hierarchical relations in antiquity addresses this agenda in the spirit of North (1981), but while avoiding any reference to markets or to voluntary trade. Instead of employing the somewhat vague notion of transaction costs, we study the institutions of antiquity by applying the

[94] In the quoted passage, Van de Mieroop employs Polanyi's terminology that the ancient Mesopotamian economy was characterized by "redistribution." We note that scholars who apply this term to ancient Egypt and Mesopotamia typically do not provide evidence for any redistribution to the peasantry that would support Polanyi's grand vision of "the patriarchal family ... on an enormously enlarged scale" (1944, p. 30). Rather, they refer to remittances to state functionaries, or to wages-in-kind to various workers and craftsmen employed by the state and the temples. Similarly, when these scholars discuss how the ancient state obtained its revenue, they often use terms like 'tribute' which may imply a voluntary contribution. Thus, in examining the extensive system of transfers known as bala between the Ur III state and its provinces, Steinkeller (1987, p. 28) claims: "the bala institution functioned as a central redistribution system, integrating all the provinces into one interdependent whole." In reexamining this terminology and the evidence on the bala system, Sharlach (2004) argues: "In a redistributive system, the entities expect to receive as much as they give, or, at least, to have their needs provided for by the central authority" (p. 21). She then concludes: "the best characterization for the bala system is the term 'tax,' because the primary function of the system was the forced contribution of the province's wealth to support the central government" (p. 21).

economic theory of moral hazard under informational asymmetry – the theory that transformed the literature on economic development in recent years.

Our environmental theory of the institutions of antiquity sheds important light also on key economic and political concerns, unrelated to antiquity or to the impact of the environment. In particular, it underscores the role of differential degrees of informational asymmetry for understanding modern taxation and for understanding the architecture of hierarchical institutions. Whereas the prevailing perception is that asymmetry of information hinders efficiency and encumbers taxation, our proposed framework reveals that the lack of transparency of individuals' activities (also known as 'privacy') may also have a silver lining in that it protects individuals' freedom, and may even promote their material well-being.

Appendix A: The agent could hide output

In this appendix we consider a variant of the basic model, in which effort is costless, but the agent may hide output. In particular, the agent may report that output is low even when it is high. The principal provides the agent with a bonus b if reported output is high, and could dismiss the agent if the reported output is low and the signal indicates that the state of nature is good ($d = 1$). The basic wage in this case covers subsistence: $\omega = m$.

The incentive scheme, $b > 0, d \in \{0, 1\}$, for the agent to accurately report when output is high should satisfy the incentive constraint:

$$b + \delta V \geq (H - L) + [(q(1 - d) + (1 - q)]\delta V. \tag{a1}$$

When binding, this incentive compatibility constraint implies:

$$b = (H - L) - q\delta dV. \tag{a2}$$

The value function $V(b, d)$ of always reporting accurately is as (2) in the basic model:

$$V(b, d) = \frac{pb}{1 - \delta(1 - \mu d)}. \tag{a3}$$

Substituting (a3) in (a2) one obtains the incentive constraint in the form comparable to (3), where $(H - L)$ substitutes for γ/p:

$$b = b(d) = (H - L)\left(1 - \frac{\delta pqd}{1 - \delta + \delta d\,(\mu + pq)}\right). \tag{a4}$$

The principal's objective is now:

$$\pi = \underset{b,d\in\{0,1\}}{max}\, p(H - L) + L - pb - \mu dx - m, \tag{a5}$$

subject to (a4).

This generates by construction two types of contracts, where the threshold transparency level $\hat{q}$ that distinguishes between them satisfies:

$$\frac{\hat{q}}{1 - \hat{q}} = \frac{(1 - p)\,x}{p\delta(H - L)}[1 - \delta(p + \hat{q}\ - 2p\hat{q})]. \tag{a6}$$

A pure carrot contract applies when $q < \hat{q}$ and has:

$$d_c = 0, b_c = H - L, \text{ and } V_c = p(H - L)/(1 - \delta), \tag{a7}$$

and a stick and carrot contract, applies for $q > \hat{q}$ and satisfies:

$$d_s = 1, b_s = (H-L)\left(1 - \frac{\delta pq}{1-\delta(p+q-2qp)}\right), \qquad V_s = \frac{p(H-L)}{1-\delta(p+q-2qp)}. \tag{a8}$$

These results reveal that the analysis of the main model is qualitatively robust to this alternative scenario of the moral hazard problem.

Appendix B: The agent knows the state of nature prior to exerting effort

The purpose of this appendix is to examine the case in which the agent knows the state of nature before exerting effort and to demonstrate that it has no qualitative effect on the model's outcomes. When the agent knows the state of nature, the incentive to entice him to exert effort is relevant only when the state of nature is good, $\theta = G$.

In a stationary equilibrium the value of the employed agent's discounted utility, when he exerts high effort when the state of nature is good, and obtains a periodic utility of γ when the state of nature is bad (as he doesn't exert effort) has to satisfy:

$$V = [pb + (1-p)\gamma] + [1 - Pr(dismiss)]\delta V, \tag{b1}$$

where $\omega = m + \gamma$. Given that the probability of dismissal under the 'stick and carrot' contract is $(1-p)(1-q)$, V is determined by b for $d = 1$:

$$V_s = \frac{pb + (1-p)\gamma}{1 - \delta(1 - (1-p)(1-q))}. \tag{b2}$$

The principal's objective function remains unchanged and the incentive constraint becomes, under 'stick and carrot':

$$b + \delta V \geq (1-q)\delta V + \gamma,$$

and under 'pure carrot':

$$b \geq \gamma.$$

This implies that the principal sets the bonus so that:

$$b_c = \gamma \text{ and } b_s = \gamma - q\delta V.$$

By replacing $b_s = \gamma - q\delta V$ in (b2) we obtain that

$$V_s = \frac{\gamma}{1 - \delta(p+q-2pq)},$$

as in the basic model, and

$$b_s = \gamma \left(1 - \frac{q\delta}{1 - \delta \left(p + q - 2pq \right)} \right). \tag{b3}$$

The difference between the cost of incentivizing the agent under $d = 0$ and $d = 1$ is

$$p\gamma - \left[pb_s + (1 - p)(1 - q) x \right], \tag{b4}$$

and hence, since $b_s < \gamma$, for a sufficiently large x there exists a threshold $\hat{q}$ above which the principal will choose $d = 1$ and below $d = 0$. In particular, replacing (b3) in (b4), $\hat{q}$ is given by:

$$\frac{\hat{q}}{1 - \hat{q}} = \frac{(1 - p)\, x}{p\delta\gamma} [1 - \delta(p + \hat{q} \ - 2p\hat{q})],$$

which is identical to the threshold in the basic model. These results reveal that the analysis of the main model is qualitatively robust to this alternative scenario of the moral hazard problem.

It should be noted that aggregate surplus in the economy is higher when the agent knows the state of nature as effort is not exerted when the state of nature is bad. Because the incentive constraint is binding, the value of employment V in this appendix is identical to the value in the basic model and thus all the additional surplus is kept by the principal.

Appendix C: The basic model with probabilistic dismissal

In this appendix we retain the basic model in the text, but allow for probabilistic dismissal: upon low output and a good signal, the principal dismisses the agent with probability $d \in [0, 1]$. The problem above can be recast as that of minimizing the principal's discretionary expenditure:

$$\varphi = \min_{d \in [0,1], b} pb + \mu x d, \tag{c1}$$

subject to the incentive constraint (3):

$$b = b(d) = \frac{\gamma}{p} \left(1 - \frac{\delta pqd}{1 - \delta + \delta d \left(\mu + pq \right)} \right). \tag{c2}$$

Given that $b(d)$ is convex over $d \in [0, 1]$, and the objective function is liner, we are guaranteed a unique optimum, characterized by tangency, other than possibly at the corners.

The optimum d can be solved for explicitly:

$$d^* = Max\{1, Min[0, d^*(q)]\}, \tag{c3}$$

where,

$$d^*(q) = \frac{1-\delta}{\delta(\mu + pq)} \left(\left(\frac{\delta\gamma q}{(1-\delta)\mu x} \right)^{1/2} - 1 \right). \tag{c4}$$

Thus, in general, there will exist $\hat{q} \in (\check{q}, 1)$ such that: $d(q) = 0$ if $q \leq \check{q}$; $d(q) = d^*(q)$ if $\check{q} \leq q \leq \hat{q}$ and $d(q) = 1$ if $\hat{q} \leq q$. To guarantee that $\check{q} \geq 0.5$ (a range of pure carrot exists), we need once again an additional condition: $(1-p)(x/\gamma) \geq \delta/(1-\delta)$. This requires that the dismissal cost x be large enough and/or that δ be small enough, to reduce the impact of the threat of dismissal on the agent.

For the same parameters as in the example in the main text figure 5 depicts the optimal dismissal probability d as a function of transparency q:

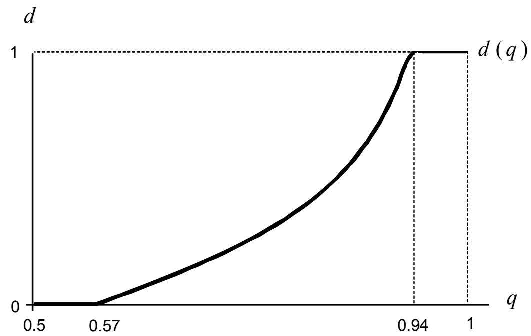


Figure 5: The optimal dismissal probability, $d \in [0, 1]$, as a function of transparency q

As in the basic case, the agent's bonus is maximal when $q < \check{q}$. In the range above $\check{q}$, as the probability of dismissal increases, the bonus decreases – since the increased threat of dismissal is used as a substitute incentive device for the bonus. The bonus, decreases when transparency increases also in the range were $q > \hat{q}$, when the dismissal probability reaches its upper limit ($d = 1$). At the same time, the principal's net expected revenue (taking into account the costs of dismissal) increases monotonically as q rises above the lower threshold $\check{q}$.

Appendix D: Enabling a set of warnings prior to dismissal

In this appendix we extend the basic model in a different direction. We avoid the possibility of probabilistic dismissal (studied in Appendix C), but enable the principal to delay dismissal, by providing a set number of warnings prior to dismissal. A warning would be issued whenever an incident of suspected shirking occurs: the output is low, but the signal that the principal receives indicates a good state of nature ($Y = L, \sigma = \tilde{G}$). That is, we assume that the principal selects an integer n, interpreted as the number of such incidents prior to dismissal. The basic model can be interpreted as one were n is restricted to the set $\{1, \infty\}$, being either one (presumably, when q is sufficiently high) or infinity (when q is sufficiently low).

Let $V(n)$ be the value function of being employed in agriculture for an agent with n warning left, upon low output and a good signal ($Y = L, \sigma = \tilde{G}$). Thus, if $n = 1$, the agent is dismissed the next time $Y = L$ and $\sigma = \tilde{G}$ and from our normalization that the value of a dismissed agent is zero: $V(0) = 0$. Let $b(n)$ be the bonus that the agent obtains when $Y = H$.

In analogy to the derivation of (2) in the basic model, we now obtain:

$$V(n) = pb(n) + \mu\delta V(n-1) + (1-\mu)\delta V(n).$$

And in analogy to the incentive constraint in (IC) we obtain the condition:

$$pb(n) = \gamma - pq\delta[V(n) - V(n-1)]. \tag{d2}$$

By combining (19) and (d2) we obtain a recursive structure for $V(n)$ in the form:

$$V(n) = A + BV(n-1), \tag{d3}$$

where the constants A and B conflate the parameters of the model:

$$A = \frac{\gamma}{1 - \delta + \delta(\mu + pq)}; B = \frac{\delta(\mu + pq)}{1 - \delta + \delta(\mu + pq)}. \tag{d4}$$

It is evident that $A > 0$ and $1 > B > 0$.

Given that $V(0) = 0$, (d3) solves explicitly to:

$$V(n) = \frac{A(1 - B^n)}{1 - B}. \tag{d5}$$

Thus we obtain also:

$$b(n) = \gamma/p - q\delta AB^{n-1}. \tag{d6}$$

It is evident that the bonuses increase with n: the more remote the dismissal possibility the higher is the bonus. It can be observed that $b(1)$ and $V(1)$ are equal to b_s and V_s of the basic model, while b_c and V_c are equal to the results obtained in (d5) and (d6), when n equals infinity.

We now move to the principal's problem of selecting the (stationary) number of warnings n that maximizes her expected discounted net revenue (given the required incentives are provided for the agent to be diligent). For that purpose we first denote by δ_2 the principal's discount factor, and denote by $\varphi(j,n)$ the discounted expected discretionary costs for the principal (including bonus payments and dismissal costs) of employing an agent who has j warnings left until dismissal, when adopting a stationary strategy of allowing every new agent with n such warnings. For $j = 1$ we have:

$$\varphi(1,n) = pb(1) + \mu[x + \delta_2\varphi(n,n)] + (1-\mu)\delta_2\varphi(1,n). \tag{d7}$$

And for $1 < j \leq n$:

$$\varphi(j,n) = pb(j) + \mu\delta_2\varphi(j-1,n)] + (1-\mu)\delta_2\varphi(j,n). \tag{d8}$$

These two equations simplify to:

$$\varphi(1,n) = \alpha b(1) + \beta x/\delta_2 + \beta\varphi(n,n); \tag{d9}$$

$$\varphi(j,n) = \alpha b(j) + \beta\varphi(j-1,n)], \tag{d10}$$

where the two constants a and β are given by:

$$\alpha = \frac{p}{1-\delta 2+\mu\delta 2}; \beta = \frac{\mu\delta 2}{1-\delta 2+\mu\delta 2}. \tag{d11}$$

Using (d9) and (d10) recursively we can now obtain the explicit solution:

$$\varphi(n,n) = \frac{\gamma}{1-\delta_2} + \frac{\beta^n x}{\delta_2(1-\beta^n)} - \frac{\alpha q\delta A(B^n-\beta^n)}{B-\beta}. \tag{d12}$$

The last two components of the discounted costs vanish when n equals infinity. The first component thus represents the discounted costs of paying every period the expected highest bonus γ for someone with a perpetual tenure, who can be incentivized only through such a bonus. The second component is the discounted expected dismissal costs, of adopting a strategy in which newcomers are offered n warnings (that is incidents of $Y = L, \sigma = \tilde{G}$) before dismissal. The last component in (d12) represents the discounted expected savings in paying bonuses that reduced from the maximum, when the number of warnings declines over time (from n to 1).[95]

The following diagram plots $n^*(q)$, the optimal n for each transparency level q, for the same parameters that were used to illustrate the basic model, with the additional parameter: $\delta_2 = 0.98$.

[95] It is instructive to confirm that $\varphi(1,1) = \varphi(\infty,\infty)$ defines exactly the threshold q^* defined in the text, and is independent of the principal's discount factor δ_2.

[96]

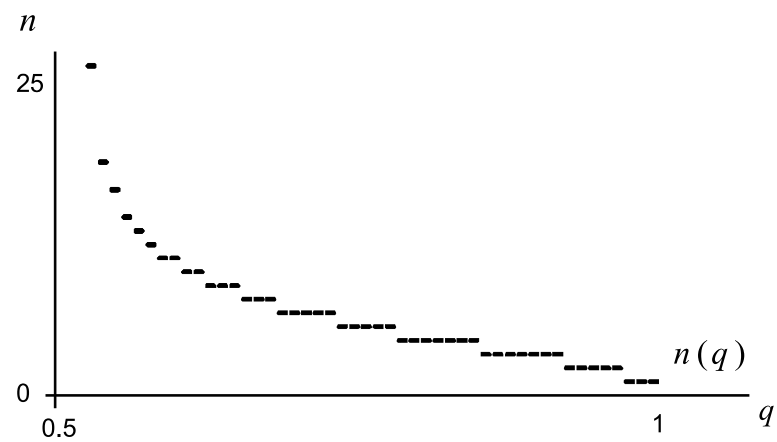


Figure 6: The optimal number of warnings before dismissal, n, as a function of transparency q

This analysis confirms the robustness of our basic results. There may be a range with sufficiently low transparency where permanent tenancy is provided. In that range, the total cost to the principal is highest and the bonus payments are maximal. As transparency increases, the optimal n decreases. In that range, as the information improves, the principal relies more and more on the threat of dismissal to incentivize the agent (in the sense of providing a decreasing number of prior warnings n) and at the same time also provides lower bonuses. Thus, once again opacity of production provides the tenant with both a form of de-facto property rights and greater reward for exerting effort.[97]

Appendix E: Endogenous Population

In this appendix we enable the principal to adjust the plot size that we denote by λ. The appendix illustrates that our qualitative results from the main model with an exogenous number of agents are unchanged. In addition, it generates new predictions regarding the effect of transparency on population size.

The output from a plot of size λ is assumed to be:

$$Y(\lambda) = \begin{cases} \lambda H & \text{if } e = h \text{ and } \theta = G; \\ \lambda L & \text{otherwise.} \end{cases}$$

[96] A lower discount rate for the principal reduces the discounted cost of dismissal and shifts the curve $n^*(q)$ downwards.

[97] In our calibration, where the probability of a warning to be issued (upon exerting effort) is $\mu = 0.2(1-q)$, even when q is close to one half, a warning will not be issued more frequently than every five years. Five warnings might then be considered as much longer than the expected life span of an adult farmer, and as effectively equal to infinity.

The agent's cost of exerting high effort to obtain λH output from a plot of size λ is denoted by $\gamma(\lambda)$. This cost function is assumed to satisfy $\gamma(0) = 0$ and to be increasing and sufficiently convex. A larger plot size is likely associated also with a larger loss from training a new agent. We represent this feature by assuming that the replacement loss is given by $x(\lambda) = \lambda x$.

If the principal's total land area is T, the number of plots (and agents) is T/λ. The principal is assumed to maximize her expected payoff from the entire land under her control. Thus, her problem becomes:

$$\Pi = \max_{\lambda>0, b\geq 0, d\in\{0,1\}} (T/\lambda)[p(\lambda H - \lambda L) + \lambda L - (m + \gamma(\lambda)) - pb - (1-q)d\lambda x],$$

s.t.

$$pb + qd\delta V \geq \gamma(\lambda),$$

$$\omega \geq m + \gamma(\lambda).$$

For any given plot size λ, the previous analysis will carry over. That is, both the subsistence constraint and the incentive constraint will be binding and in particular: $\omega = m + \gamma(\lambda)$. When the signal is uninformative (q is sufficiently low), a 'pure carrot' contract will apply:

$$d_c = 0; \qquad b_c = \gamma(\lambda)/p. \tag{e1}$$

The principal's problem in this range is equivalent to selecting λ to minimize $T(m + 2\gamma(\lambda))/\lambda$. Given the assumed convexity of $\gamma(\lambda)$, the optimal λ_c is determined by the unique solution to the first order condition:

$$m/2 = \gamma(\lambda_c)\eta(\lambda\gamma). \tag{e2}$$

Here $1 + \eta(\lambda)$ is the elasticity of the cost function: $\eta(\lambda) = \lambda\gamma'(\lambda)/\gamma(\lambda) - 1$.

Similarly, in an informative environment where q is sufficiently high, the optimal contract is of the 'stick and carrot', and according to (6):

$$d_s = 1; \qquad b_s(q,\lambda) = \frac{\gamma(\lambda)}{p} - \frac{q\delta\gamma(\lambda)}{1 - \delta(p + q - 2pq)}. \tag{e3}$$

The principal's problem, reduces then to selecting λ to minimize $T(m + \gamma(\lambda) + pb_s(q,\lambda)]/\lambda$. Upon substituting from (e3), this is equivalent to minimizing: $T[z(q) + \gamma(\lambda)]/\lambda$, where we define:

$$z(q) = \frac{m[1 - \delta(p + q - 2pq)]}{2[1 - \delta(p + q - 1.5pq)]} = \frac{m}{2}\frac{1 + 0.5pq\delta}{1 - \delta(p + q - 1.5pq)}. \tag{e4}$$

The optimal plot size in this case, λ_s, can be solved from:

$$z(q) = \lambda_s \gamma'(\lambda_s) - \gamma(\lambda_s) = \gamma(\lambda_s)\eta(\lambda_s). \tag{e5}$$

Under our concavity assumption on $\gamma(\lambda)$, the right hand side of (e5) increases with plot size λ. Since $z(q)$ increases in q, and since by (e4) $z(q) > m/2$ for any q, conditions (e2) and (e5) imply that plot size λ_s in the 'stick and carrot' regime will increase with transparency q, and will be larger than λ_c. This is due to the fact that when the stick is in use, it costs less to incentivize the agent, and so the principal may as well assign the agent a larger plot size and economize on the fixed cost of agents' maintenance. The larger plot size implies, of course, a smaller population. The extra decision variable leads to higher expected revenue to the principal, in comparison with the case of a fixed plot size. To better evaluate the impact of endogenous plot size, we now assume that the cost function has constant elasticity, and calibrate it so that $\gamma(1) = \gamma$ and so that the optimal plot size under the 'pure carrot' regime remains equal to one ($\lambda_c = 1$). This guarantees that under 'pure carrot' every aspect of the economy with fixed land size remains unchanged. However, the greater revenue of the 'stick and carrot' regime implies that the new threshold transparency $\hat{q}_\lambda$ for switching into this type of contract will be lower than before. At this transparency threshold, agents are made discretely worse off when switched from a 'pure carrot' contract to a 'stick and carrot' contract. But beyond this point, since each agent's net periodic utility depends positively on the expected bonus payment pb for high effort, the larger plot size implies that agents are made better off as transparency increases. Moreover, beyond the old threshold level $\hat{q}$ (for fixed plot size case), agents would be better than under the fixed plot case. This is compatible with increased revenue to the principal, since the number of agents is smaller.

These results, if depicted, are similar to the plot in Figure 1, if we normalize $T = 1$, so as to depict the principal's total expected income, or her expected income per unit of land, rather than expected income per plot (or per agent). The change in the figure once plot size is endogenous is that the threshold $\hat{q}$ is smaller and the principal's income above the threshold is higher. It should be noted that in a figure that captures the principal's income when plot size is endogenous the vertical difference between the two lines does not represent each agent's expected income, since this (as noted above) is in fact increasing, due to the larger plot size.[98]

We obtain here that, with everything else held constant, as economic activity becomes more transparent, the smaller is the agent's output share and the lower is population density. These

[98] Assuming $c(\lambda) = c\lambda^{1+\eta}$ and positing that $\lambda_c = 1$, implies that $\lambda_c^{1+\eta} = m/2c\eta$ and requires: $\eta = m/(2c)$.

may seem like confirmable predictions; however the presumption that all else remains the same is daunting.

References

[1] Acemoglu Daron and James A. Robinson (2012), *Why Nations Fail: The Origins of Power, Prosperity, and Poverty,* Random House.

[2] Ackerberg, Daniel. A. and Maristella Botticini (2000), "The Choice of Agrarian Contracts in Early Renaissance Tuscany: Risk Sharing, Moral Hazard, or Capital Market Imperfections?" *Explorations in Economic History*, 37, 241-257.

[3] Ackerberg, Daniel. A. and Maristella Botticini (2002), "Endogenous Matching and the Empirical Determinants of Contract Form," *Journal of Political Economy*, 110, 564-591.

[4] Adams, Robert M. (1960), "Early Civilizations, Subsistence, and Environment," in Carl H. Kraeling and Robert M. Adams (eds.) *City invincible: a Symposium on Urbanization and Cultural Development in the Ancient Near East,* University of Chicago Press.

[5] Adams, Robert M. (1981), *Heartland of Cities: Surveys of Ancient Settlement and Land Use of the Central Floodplain of the Euphrates*, University of Chicago Press.

[6] Adams, Robert M. (2005), "Critique of Guillermo Algaze's 'The Sumerian Takeoff'," *Structure and Dynamics* 1, article 10.

[7] Algaze, Guillermo (2001), "Initial Social Complexity in Southwestern Asia: The Mesopotamian Advantage," *Current Anthropology,* 42, 199-233.

[8] Algaze, Guillermo (2005), "The Sumerian Takeoff," *Structure and Dynamics*, 1, article 2.

[9] Algaze, Guillermo (2008), *Ancient Mesopotamia at the dawn of Civilization: The Evolution of an Urban Landscape,* University of Chicago Press.

[10] Allen, Robert C. (1997), "Agriculture and the Origins of the State in Ancient Egypt," *Explorations in Economic History,* 34, 135–154.

[11] Ashraf, Quamrul and Oded Galor (2011), "Dynamics and Stagnation in the Malthusian Epoch," *American Economic Review*, 101, 2003-2041.

[12] Ashraf, Quamrul and Oded Galor (2013), "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review,* 103, 1-46.

[13] Baer, Gabriel (1962), *A History of Landownership in Modern Egypt, 1800-1950,* Oxford University Press.

[14] Baer, Gabriel (1969), *Studies in the Social History of Modern Egypt*, University of Chicago Press.

[15] Baer, Klaus (1962), "The Low Price of Land in Ancient Egypt," *Journal of the American Research Center in Egypt*, 1, 25-45.

[16] Baines, John and Norman Yoffee (1998), "Order, Legitimacy, and Wealth in Ancient Egypt and Mesopotamia," in Gary M. Feinman and Joyce Marcus (eds.) *Archaic States,* School of American Research Press, 199-260.

[17] Bang, Peter Fibiger (2009), "Review Article: The Ancient Economy and New Institutional Economics," *Journal of Roman Studies,* 99, 194-206.

[18] Bates, Robert, (1983), *Essays on the political economy of rural Africa*, University of California Press.

[19] Bentzen, Jeanet Sinding, Nicolai Kaarsen and Asger Moll Wingender (2012), "Irrigation and Autocracy," University of Copenhagen, Dept. of Economics Discussion Paper No. 12-06.

[20] Besley, Timothy and Maitreesh Ghatak (2009), "Property Rights and Economic Development," in Dani Rodrik and Mark Rosenzweig (eds) *Handbook of Development Economics*, North Holland.

[21] Besley, Timothy and Torsten Persson (2009), "The Origins of State Capacity: Property Rights, Taxation and Politics," *American Economic Review*, 99, 1218-1244.

[22] Besley, Timothy and Torsten Persson (2010), "State Capacity, Conflict, and Development," *Econometrica,* 78, 1-34.

[23] Billman , Brian R. (2002), "Irrigation and the Origins of the Southern Moche State on the North Coast of Peru," *Latin American Antiquity*, 13, 371-400.

[24] Bockstette, Valerie, Areendam Chanda, and Louis Putterman (2002), "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth*, 7, 347-69.

[25] Butzer, Karl W. (1976), *Early Hydraulic Civilization: A Study in Cultural Ecology,* University of Chicago Press.

[26] Butzer, Karl W. (2012), "Collapse, Environment, and Society," *PNAS*, 109, 3632–3639.

[27] Braudel, Fernand ([1949], 1976), "The Mediterranean and the Mediterranean world," in *The age of Philip II*, translated by Siân Reynolds, Harper & Row.

[28] Carneiro, Robert L. (1970), "A Theory of the Origin of the State," *Science,* 169, 733-738.

[29] Cheung, Steven N.S. (1969), *The Theory of Share Tenancy*, University of Chicago Press, Chicago.

[30] Childe, V. Gordon ([1936], 1951), *Man Makes Himself*, Watts & Co.

[31] Coase, Ronald (1937), "The Nature of the Firm," *Economica,* 4, 386–405

[32] Cooper, Richard S. (1976), "The Assessment and Collection of Kharāj Tax in Medieval Egypt," *Journal of the American Oriental Society,* 96, 365-382.

[33] Demsetz, Harold (1967), "Toward a Theory of Property Rights," *The American Economic Review,* 57, 347-359.

[34] Diamond, Jared (1997), *Guns, Germs, and Steel: The Fates of Human Societies*, Norton, New York.

[35] Domar, Evsey (1970), "The causes of slavery or serfdom: a hypothesis," *Economic History Review,* 30, 18-32.

[36] Dow Gregory K. and Clyde G. Reed (2013), "The Origins of Inequality: Insiders, Outsiders, Elites, and Commoners," *Journal of Political Economy,* 121, 609-641.

[37] Eyre, Christopher J. (1994), "The Water Regime for Orchards and Plantations in Pharaonic Egypt," *Journal of Egyptian Archaeology,* 80, 57-80.

[38] Eyre, Christopher J. (1997), "Peasants and 'Modern' Leasing Strategies in Ancient Egypt," *Journal of the Economic and Social History of the Orient,* 40, 367-390.

[39] Eyre, Christopher J. (1999), "The Village Economy in Pharaonic Egypt," in Alan K. Bowman and Eugene Rogan (eds.) *Proceedings of the British Academy, 96: Agriculture in Egypt from Pharaonic to Modern Times*, Oxford University Press, 33-60.

[40] Fenske, James (2013), "Ecology, Trade and State in Pre-Colonial Africa," *Journal of the European Economic Association,* Forthcoming.

[41] Finely, Moses I. ([1973], 1999), *The Ancient Economy*, updated edition, University of California Press.

[42] Garicano, Luis (2000), "Hierarchies and the organization of knowledge in production," *Journal of Political Economy,* 108, 874-904.

[43] Gennaioli, Nicola and Hans-Joachim Voth (2010), "State Capacity and War," working paper.

[44] Ghalwash, Maha A. (1998), "Acquisition by the Peasants of Mid-Nineteenth Century Egypt: the Ramya System," *Studia Islamica,* 88, 121-139.

[45] Goddeeris, Anne (2002), *Economy and Society in Northern Babylonia in the Early Old Babylonian Period*, Peeters Publishers.

[46] Greif, Avner (1993), "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders," *American Economic Review,* 83, 525-548.

[47] Greif, Avner (2006), *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*, Cambridge University Press.

[48] Halstead, Paul (1989), "The Economy has a normal Surplus: Economic Stability and Social Change among Early Farming Communities of Thessaly, Greece," in Paul Halstead and John O'Shea (eds.) *Bad Year Economics: Cultural Responses to Risk and Uncertainty*, Cambridge University Press, 68-80.

[49] Harris, Milton and Arthur Raviv (1979), "Optimal incentive contracts with imperfect information," *Journal of Economic Theory,* 20, 231-259.

[50] Hassan, Fekri A. (1994), "Population Ecology and Civilization in Ancient Egypt," in Carole L. Crumley (ed.) *Historical Ecology: Cultural Knowledge and Changing Landscapes,* School of American Research Press, 155-182.

[51] Hassan, Fekri A. (2007), "Droughts, Famines and the Collapse of the Old-Kingdom; Re-Reading Ipuwer," in Zahi A, Hawass and Janet Richards (eds.) *The Archaeology and Art of Ancient Egypt: Essays in Honor of David B. O'Connor*, vol. 1, Supreme Council of Egyptian Antiquities, 357-377.

[52] Hicks, John (1969), *A theory of Economic History*, Oxford University Press.

[53] Holmström, Bengt (1999), "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies,* 66, 169-182.

[54] Hughes, George Robert (1952), *Saite Demotic Land Leases,* University of Chicago Press.

[55] Hunt, Robert C. (1987), "The Role of Bureaucracy in the Provisioning of Cities: A Framework for Analysis of the ancient Near East," in McGuire Gibson and Robert D. Biggs (eds.) *The Organization of Power: Aspects of Bureaucracy in the Ancient Near East*, The Oriental Institute of the University of Chicago, 161-192.

[56] Jas, Remko M. (2000), "Land Tenure in Northern Mesopotamia: Old Sources and the Modern Environment," in Remko M. Jas (ed.) *Rainfall and Agriculture in Northern Mesopotamia*, Nederland Historisch-Archaeologisch Instituut.

[57] Jursa, Michael (2010), *Aspects of the Economic History of Babylonia in the First Millennium BC: Economic Geography, Economic Mentalities, Agriculture, the Use of Money and the Problem of Economic Growth*, Ugarit-Verlag.

[58] Kau, James B. and Paul H. Rubin (1981), "The size of Government," *Public Choice,* 37, 261-274.

[59] Kemp, Barry J. (2006), *Ancient Egypt: Anatomy of a Civilization*, Second edition, Routledge.

[60] Kleven, Henrik Jacobsen, Claus Thustrup Kreiner and Emmanuel Saez (2009), "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries," *NBER* Working Paper No. 15218.

[61] Lagerlöf, Nils-Petter (2009), "Slavery and Other Property Rights," *Review of Economic Studies,* 76, 319-342.

[62] Liverani, Mario ([1998], 2006), *Uruk: The First City*, Edited and translated by Zinab Bahrani and Marc Van de Mieroop, Equinox Publishing.

[63] Lloyd, Alan B. (1983), "The Late period: 664-323," in Bruce G. Trigger et al. (eds.) *Ancient Egypt: A Social History*, Cambridge University Press, 1983, 279-364.

[64] Ma, Debin (2011), "Rock, Scissors, Paper: the Problem of Incentives and Information in Traditional Chinese State and the Origin of Great Divergence," London School of Economics, Economic History Working Paper No.152.

[65] Mann, Michael (1986), *The Sources of Social Power, Volume I: A History of Power from the Beginning to A.D. 1760*, Cambridge University Press.

[66] Manning, Joseph G. (2003), *Land and Power in Ptolemaic Egypt: The Structure of Land Tenure*, Cambridge University Press.

[67] Mayshar, Joram (1991), "Taxation with Costly Administration," *Scandinavian Journal of Economics,* 93, 75-88.

[68] Mayshar, Joram, Omer Moav and Zvika Neeman (2011), "Transparency, Appropriability and the Early State," CEPR Discussion Papers 8548, C.E.P.R. Discussion Papers.

[69] Meltzer, Allan H. and Scott F. Richard (1981), "A Rational Theory of the Size of Government," *Journal of Political Economy,* 89, 914-927.

[70] Melumad, Nahum D., Dilip Mookherjee and Stefan Reichelstein (1995), "Hierarchical Decentralization of Incentive Contracts," *The RAND Journal of Economics,* 26, 654-672.

[71] Michalopoulos, Stelios (2012), "The Origins of Ethnolinguistic Diversity," *American Economic Review,* 102, 1508-1539.

[72] Mieroop, Marc Van de (1997), *The Ancient Mesopotamian City*, Oxford University Press.

[73] Mieroop, Marc Van de (1999), *Cuneiform Texts and the Writing of History*, Routledge.

[74] Mieroop, Marc Van de (2004), "Economic Theories and the Ancient Near East," in Robert Rollinger and Christoph Ulf (eds.) *Commerce and Monetary Systems in the Ancient World: Means of Transmission and Cultural Interaction*, Franz Steiner Verlag, 54-64.

[75] Mulligan, Casey B., Ricard Gil and Xavier Sala-i-Martin (2004), "Do Democracies Have Different Public Policies than Nondemocracies?" *Journal of Economic Perspectives,* 18, 51-74.

[76] Nissen, Hans J. (1988), *The Early History of the Ancient Near East: 9000 – 2000 B.C.*, Translated by Elizabeth Lutzeier with Kenneth J. Northcott, University of Chicago Press.

[77] North, Douglass. C. (1977), "Markets and Other Allocation Systems in History: The Challenge of Karl Polanyi," *Journal of European Economic History,* 6, 703-719.

[78] North, Douglass. C. (1981), *Structure and Change in Economic History*, W.W. Norton & Co.

[79] Noy-Meir, Imanuel (1973), "Desert Ecosystems: Environment and Producers," *Annual Review of Ecology and Systematics,* 4, 25-51.

[80] Olson, Mancur (1993), "Dictatorship, Democracy, and Development," *American Political Science Review,* 87, 567-576.

[81] Ouyang, Xiaoli (2010), "Administration of the Irrigation Fee in Umma During the Ur III Period (ca. 2112–2004 BCE)," in L. Kogan et al. (eds.) *City Administration in the Ancient Near East*, Eisenbrauns, 317-349.

[82] Park, Thomas K. (1992), "Early Trends toward Class Stratification: Chaos, Common Property, and Flood Recession Agriculture," *American Anthropologist,* 94, 90-117.

[83] Parkinson, C. Northcote (1960), *The law and the Profits,* John Murray.

[84] Peacock, Alan T. and Jack Wiseman (1961), *The Growth of Public Expenditure in the United Kingdom*, Princeton University Press.

[85] Polanyi, Karl ([1944], 1957), *The Great Transformation: The Political and Economic Origins of Our Time*, Beacon Press.

[86] Postgate, J. Nicholas (1994), *Early Mesopotamia: Society and Economy at the Dawn of History*, Routledge.

[87] Potts, Daniel T. (1997), *Mesopotamian Civilization: The Material Foundation*, Athlone Press.

[88] Putterman, Louis and David Weil (2010), "Post-1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality," *The Quarterly Journal of Economics,* 125, 1627-1682.

[89] Redford, Donald B. (1970), *A Study of the Biblical Story of Joseph: (Genesis 37-50)*, Brill.

[90] Redford, Donald B. (1997), "The Ancient Egyptian 'City': Figment or Reality," in Walter E. Aufrecht et al. (eds.) *Urbanism in Antiquity: From Mesopotamia to Crete,* Sheffield University Press, 210-220.

[91] Renger, Johannes M. (1995), "Institutional, Communal, and Individual Ownership or Possession of Arable Land in Ancient Mesopotamia From the End of the Fourth to the End of the First Millennium B.C.," *Chicago-Kent Law Review,* 71, 269-319.

[92] Sah, Raaj Kumar and Joseph E. Stiglitz (1986), "The Architecture of Economic-Systems - Hierarchies and Polyarchies," *The American Economic Review,* 76, 716-727.

[93] Sen Amartya (1981), *Poverty and Famines: An Essay on Entitlement and Deprivation*, New York: The Clarendon Press, Oxford University Press.

[94] Shapiro, Carl and Joseph E. Stiglitz (1984), "Equilibrium Unemployment as a Worker Discipline Device," *The American Economic Review,* 74, 433-444.

[95] Sharlach, Tonia M. (2004), *Provincial Taxation and the Ur III State*, Brill-Styx.

[96] Slemrod, Joel B. and Shlomo Yitzhaki, (2002), "Tax Avoidance, Evasion, and Administration," in Auerbach, Alan J. and Martin Feldstein (eds.) *Handbook of Public Economics*, Volume 3, Elsevier, 1423-1470.

[97] Spolaore, Enrico and Romain Wacziarg (2013), "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature,* 51, 325-369.

[98] Steinkeller, Piotr (1987), "The Administrative and Economic Organization of the Ur III State: The Core and the Periphery," in McGuire Gibson and Robert D. Biggs (eds.) *The Organization of Power: Aspects of Bureaucracy in the Ancient Near East*, The Oriental Institute of the University of Chicago, 19-41.

[99] Steinkeller, Piotr (1999), "Land-Tenure Conditions in Third Millennium Babylonia: The Problem of Regional Variation," in *Michael Hudson and Baruch A. Levine (eds.) Urbanization and Land Ownership in the Ancient Near East*, Peabody Museum, Harvard University, 289-329.

[100] Steinkeller, Piotr (2007), "City and Countryside in Third-Millennium Southern Babylonia," in Elizabeth C. Stone (ed.) *Settlement and Society: Essays Dedicated to Robert McCormick Adams*, Cotsen Institute of Archaeology University of California, Los Angeles and The Oriental Institute of the University of Chicago, 185-211.

[101] Stigler, George (1961), "The Economics of Information," *Journal of Political Economy,* 69, 213-225

[102] Stiglitz, Joesph E. (1974), "Incentives and risk sharing in sharecropping," *Review of Economic Studies,* 41, 219-255.

[103] Stiglitz, Joesph E. (1975), "Incentives, Risk, and Information: Notes Towards a Theory of Hierarchy," *Bell Journal of Economics,* 6, 552-579.

[104] Tilly, Charles (1990), *Coercion, Capital and European States, AD 990-1992*, Blackwell.

[105] Trigger, Bruce (1993), *Early Civilizations: Ancient Egypt in Context*, American University in Cairo Press.

[106] Trigger, Bruce (2003), *Understanding Early Civilizations: A Comparative Study*, Cambridge University Press.

[107] Ur, Jason A. (2010), "Cycles of Civilization in Northern Mesopotamia, 4400–2000 BC," *Journal of Archaeological Research,* 18, 387-431.

[108] Vollrath, Dietrich (2009), "The Agricultural Basis of Comparative Development," working paper.

[109] Wagner, Adolph ([1883], 1958), "Three Extracts on Public Finance," in Richard A. Musgrave and Alan T. Rowley (eds.) *Classics in the Theory of Public Finance*, Macmillan, 1-15.

[110] Warriner, Doreen (1948), *Land and Poverty in the Middle East*, Royal Institute of International Affairs.

[111] Wenke, Robert J. (2009), *The Ancient Egyptian State: The Origins of Egyptian Culture (c. 8000-2000 BC)*, Cambridge University Press.

[112] Westbrook, Raymond, editor (2003), *A History of Ancient Near Eastern Law*, volume one, Brill.

[113] Wilkinson, Tony J. (1994), "The Structure and Dynamics of Dry-Farming States in Upper Mesopotamia," *Current Anthropology,* 35, 483-520.

[114] Wilkinson, Tony J. (2003), *Archaeological Landscapes of the Near East*, The University of Arizona Press.

[115] Willcocks, William (1899), *Egyptian Irrigation*, Second edition, London.

[116] Wilson, John A. (1960), "Egypt through the New Kingdom: Civilization without Cities" in Carl H. Kraeling and Robert M. Adams (eds.) *City invincible: a Symposium on Urbanization and Cultural Development in the Ancient Near East*, University of Chicago Press.

[117] Wittfogel, Karl A. (1957), *Oriental Despotism: A Comparative Study of Total Power*, Yale University Press.

[118] Wu, Lemin (2012), "Does Malthus Really Explain the Constancy of Living Standards?" Working paper, University of California, Berkeley.

[119] Yoffee, Norman (1995), "Political Economy in Early Mesopotamian States," *Annual Review of Anthropology,* 24, 281-311.

[120] Yoffee, Norman (2005), *Myths of the Archaic State: Evolution of the Earliest Cities, States and Civilizations*, Cambridge University Press.

[121] Zaccagnini, Carlo (1999), "Economic Aspects of Land Ownership and Land use in Northern Mesopotamia and Syria from the Late Third Millennium to the Neo-Assyrian Period," in Michael Hudson and Baruch A. Levine (eds.) *Urbanization and Land Ownership in the Ancient Near East* edited by, Peabody Museum, Harvard University, 331-352.