Transparency, Appropriability and the Early State*

Joram Mayshar[†] Omer Moav[‡] Zvika Neeman[§]

August 12, 2011

Abstract

We propose a general theory that explains the extent of the state and accounts for related institutions as byproducts of the state's extractive technology. We posit further that this extractive technology is determined by the transparency of the production technology. This theory is applied to examine two principal phases in the evolution of the early state. First, we argue that the common explanation of the emergence of the state as a consequence of the availability of food surplus due to the Neolithic Revolution is flawed, since it ignores Malthusian considerations. In contrast, we suggest that what led to the emergence of the state was a transformation of the tax technology that was induced by the greater transparency of the new farming technology. We then apply our theory to explain key institutional features that distinguished ancient Egypt from ancient Mesopotamia, and, in particular, to explain their different land tenure regimes.

KEYWORDS: Transparency, Appropriability, The Early State, Institutions, Land Tenure

JEL CLASSIFICATION NUMBERS: D02, D82, H10, O43

*We are grateful for comment received from Eddie Dekel, Diana Egerton-Warburton, Oded Galor, Jacob Metzer, Ariel Rubinstein, Yona Rubinstein, Yannay Spitzer, Nathan Sussman, Joachim Voth, and from seminar participants in Royal Holloway University of London, Ben-Gurion University, The Hebrew University of Jerusalem, Bilkent University, Tel-Aviv University, University of Aberdeen, University of Edinburgh and conference participants at The Royal Economic Society Conference, Royal Holloway, April 2011, CEPR Public Policy Symposium, Zurich, May 2011, and Minerva Summer School in Economic Growth, Jerusalem, June 2011.

†Department of Economics, Hebrew University of Jerusalem. Email: msjoram@huji.ac.il.

‡Department of Economics, Hebrew University of Jerusalem, Royal Holloway University of London, and CEPR. Email: omer.moav100@gmail.com; Moav's research is supported by the Israel Science Foundation (Grant No. 73/11).

§Eitan Berglas School of Economics, Tel-Aviv University, Email: zvika@post.tau.ac.il

1 Introduction

The emergence of the state is commonly associated with the increase in productivity that accompanied the Neolithic Revolution. The standard argument is that the transition from foraging to agriculture created food surplus, and the availability of surplus facilitated through various channels the advance of an elite that did not engage in food production, leading ultimately to the emergent state. We argue that this explanation is deeply flawed. The protracted rise in productivity during the Neolithic period by itself could not have generated any surplus, since population size would have adjusted endogenously to prevent its creation.[1] Rather, it was the forerunners of the early state that generated surplus through expropriation, thus, in part, curtailing the increase in population. More specifically, we argue that the transition to agriculture made expropriation more rewarding and thus induced banditry and a demand for protection. From our perspective it does not matter whether that demand was met by roving bandits who turned stationary (as argued by Olson 1993) or by leaders from within (as is more commonly assumed). Yet, due to the public-good nature of protection, it is unlikely that the farmers who sought protection eagerly departed from their produce to finance it. Thus, it was the same innovation in the appropriating technology that invited robbery to begin with, we contend, that also holds the key for understanding the emergent state.

Our more specific proposal is that the feature of the Neolithic that made expropriation more rewarding was the cultivation of cereals and the ensuing requirement to store grains. Such storage is fully consistent with a Malthusian subsistence regime and with the lack of any long-run excess of food.[2] Yet it was this appended feature of the new production technology, we argue, that induced a fundamental change in the 'tax technology,' facilitating the regular appropriation of food-stuff by bandits and by the emergent state.[3]

[1] Richerson, Boyd and Bettinger (2001:388-389) argued along similar lines against the idea that population pressure forced humans to adopt agriculture, claiming that agriculture was induced by climatic changes that made it feasible, and that subsequent population growth rates could easily have kept pace with the technological improvements. Their interest in the origins of agriculture ends, however, where ours begins. Bellwood (2006:14-19) summarizes supporting evidence of phenomenal population growth rates through reproduction (including one historical case of a seven-fold increase in 66 years) that resulted from the introduction of agriculture to some frontier environments. Given such a potential human growth rate, the often proposed rate of human population growth during the Neolithic period, in the order of magnitude of 0.01 percent, does not indicate the existence of surplus or population pressure.

[2] The literature that relates storage to early stratification is reviewed in section 2. We note that in the archaeology literature, storage is sometimes considered as synonymous, or at least as a prime indicator of surplus. But this generates a semantic confusion. We will consistently refer to surplus as food in excess of long-term survival needs. Thus, inter-annual grain storage for the purpose of buffering against future famine is not an indicator of surplus.

[3] The notion of a 'tax technology' was proposed by Mayshar (1993); see also Slemrod and Yitzhaki (2002). Implicitly we assume that the food that was appropriated by the elite did not translate into elite population, but was rather

We seek to take here one important further step. The binary existence or non-existence of stored food cannot explain fundamental regional variations among the early states. We propose that the more general aspect of the transition to agriculture, that underlies both the ultimate emergence of the state and regional variations among the early states, was the transparency of the production process. According to this perception, grain storage is but an example (albeit an important one) of greater transparency of the production technology. It is the broadly interpreted 'transparency' of production, we claim, that induces the 'tax technology.' We thus aim to study the impact of production transparency on the scale of tax receipts and on related social institutional. We elect to focus here on the earliest civilizations of Mesopotamia and Egypt in the fourth and the third millennia BCE.[4] But, as indicated in the concluding section, we contend that our theory is applicable also to the modern age, in the wake of the industrial and the informational revolutions.

To formally analyze the implications of transparency of economic activity we employ a simple principal-agent model. We interpret the agent as a representative farmer and the principal as an absentee land-owner representing the government. Agricultural output is assumed to depend on the agent's unobservable effort and on the state of nature. Output, we assume, is observed by the principal, but not the state of nature, for which the principal has only a signal.[5] Thus, the principal cannot accurately infer the effort level exerted by the agent. To handle this moral hazard problem, the principal offers the agent a contract that contains a 'carrot' and a 'stick.' The carrot is in terms of output to be retained by the agent; the stick is the threat of the agent's dismissal. We assume that dismissal is costly for both the agent and the principal. The accuracy of a public signal about the state of nature is the main exogenous variable in our model, representing the degree of transparency.

Our principal proposition is that the more accurate the public signal is, the larger the role of the stick and the smaller the role of the carrot. That is, greater transparency increases the probability that the principal will use the strategy of dismissing the agent if output is low, and at the same time lowers the agent's remuneration. It is in this sense that greater transparency induces a form of 'servitude.' On the other hand, opacity results in what may be described as 'freedom,' where the farmer is never dismissed and the land is de-facto owned by farmers who pay some of their output as taxes. We propose to interpret this result as showing that lack of transparency helps protect the freedom of farmers, secures their property rights over the land they cultivate, and increases their

spent on servants and prestige commodities produced by artisans or imported.

[4] From here on, all references to millennia are to be understood as BCE.

[5] Appendix B presents an alternative framework in which output is not observable by the principal and the moral hazard problem pertains to hiding (or misreporting, or sheltering) output. The qualitative results remain unchanged.

share of the product, at the expense of the state.

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

The Old Testament refers to Egypt as a "house of bondage." By now it is well known, however, that slavery was not more pervasive in ancient Egypt than elsewhere in the ancient Near East. It appears that in using this term, the biblical authors intended to express their indignation about Pharaonic Egypt as the epitome 'bad' society. The principal aspect of ancient Egypt that apparently offended the early Israelites was Egyptian land tenure institutions. In ancient Israel, private land ownership was considered the desired norm (in analogy to the vision of the American founding fathers). Moreover, in Israel as in rain-fed Syria and northern Mesopotamia, owner-operated farming was widespread. In contrast, in Egypt, as well as in southern Mesopotamia, essentially all farmers were serfs who tilled land that they did not own.

Intensive agriculture was adopted in the rain-fed northern highlands of northern Mesopotamia in the seventh millennium, well before it was taken up in the alluvial planes of southern Mesopotamia (Sumer) and in the Nile valley. Yet it was in Sumer were the first city-states were formed in the fourth millennium; and it was in Upper Egypt where the first central territorial state was formed (at about 3000 BCE). The rapid development of city states in Sumer has been described as a "takeoff" and as an "Urban Revolution" (Childe 1936, Liverani 2006). These critical developments raise two puzzles: why did the Urban Revolution take place in Sumer, even though it was 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 rapid development of a stable central state in Egypt – see Kemp (2006) and Wenke (2009) – is only one of the key differences between ancient Mesopotamia and ancient Egypt. In comparing the two ancient civilizations, Baines and Yoffee (1998:268) conclude: "the two civilizations are profoundly different." Other distinguishing features have been often noted by scholars (see Trigger 1993, 2003). Thus, even though in predynastic Egypt there were fortified city-states, after the early formation of the central state, Egyptian cities were not fortified and played a rather limited role as administrative centers. This led Wilson (1960) to famously characterize ancient Egypt as "a civilization without cities."[7] In contrast, up to the first millennium, Mesopotamia was ruled most

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

[7] Wilson contended (1960:135) that "Ancient Egypt carried on her life through dozens of moderate-sized owns and myriads of agricultural villages;" but was "without a single major city." Recent Egyptologists have emphasized the existence of walled cities in the predynastic period and of diverse population centers (see Kemp 2006), but these hardly contradict Wilson's perceptive generalization, particularly with the perspective of Mesopotamia (see Baines

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 trade, the sale and lease of land, and extensive lending were widespread in Mesopotamia and amply documented in hundred of thousands of surviving documents. Even though extensive records survive also from ancient Egypt, it appears that up to the mid-first millennium the extant of similar commercial activity in Egypt has been rather limited. As a result, as Wilson (1960) noted, whereas written law codes are known in Mesopotamia already in the late third Millennium, they were absent in ancient Egypt.

We shall argue that these and related institutional differences between ancient Egypt and Mesopotamia, as well as the major differences between southern and northern Mesopotamia, stem from the differences in the degree of transparency of farming activity. In particular, we contend that the principal unique feature of 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' of individual farmers throughout Egypt with high accuracy.[8] We argue that this feature enabled the different layers of government in Egypt to employ the 'stick' incentives scheme, rather than the 'carrot.' This stick scheme was adopted by local lords towards the individual peasants, explaining why Egypt did not resort to private land ownership. More importantly, the Nile's transparency enabled the Pharaohs to employ a similar stick 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 a significant share of the country's agricultural product with the use of a relatively lean bureaucracy and with provincial administrative centers that retained little independent power. We shall similarly argue that the differences between northern and southern Mesopotamia may be attributed to somewhat similar differences in the transparency of farming: the alluvial planes were more prone to a centralized governmental control than the mostly rain-fed highland in the north. Moreover, we shall apply a similar line of argument to explain why Egypt rapidly evolved into a central stable state with weak cities, while the cities in Mesopotamia were powerful and the successive attempts to unify Mesopotamia under a central state were prone to fail.

Given that our subject is wide-ranging, we will refer to the literature that pertains to our model and to the ancient world in the body of the paper.[9] At this stage, we will refer briefly

and Yoffee (1998:209). Whereas Van de Mieroop (1999:1) writes: "Mesopotamia was not only the oldest urban civilization, but also the most urbanized society of antiquity," Eyre (1999:35) summarized the situation in Egypt in Wilson's spirit: "Egypt was probably always a village society."

[8] See Cooper (1976) and the details in section 4 below.

[9] For a comprehensive survey of this literature as it applies to ancient Egypt see Allen (1997).

only to studies on endogenous institutions. The idea that social institutions are endogenous and determined to a major extent by environmental factors is an old one, and was adopted by many scholars across diverse social disciplines, among others by: Childe (1936), Braudel ([1949] 1976), Wittfogel (1957) and Diamond (1997).[10] Within economics, institutions and institutional change has been the subject of study, among others, by North (1990) and Greif (1993, 2006) – although not in an environmental context. Engerman and Sokoloff (1997) and Acemoglu Johnson and Robinson (2002) propose that European colonists adopted different institutions in the different environments that they colonized. By adopting institutions that protected property rights and fostered equality where there was no wealth to extract, and by maintaining inequality and adopting extractive institutions in the relatively wealthy societies, these colonists ultimately caused a "reversal of fortune."[11]

Unlike the cited economic literature that focuses on modern and pre-modern institutions that facilitate efficient market trade, the institutions that we study are those that prevailed in the ancient world over many millennia, such as land tenure and tenancy/slavery arrangements, property rights, law codes and the extant of the state. We seek to explain these early institutions by materialist considerations that incorporate environmental factors and farming technology. More specifically, we seek to apply an overall theme that the transparency of economic activity plays the lead role in explaining these social institutions. In addition, we incorporate Malthusian considerations that were not addressed in the above-cited contributions.

In Section 2 we first survey the literature on the emergence of the state, and then criticize it on the basis of the Malthusian argument that was noted above and offer the alternative focus on storage and appropriability. Section 3 is devoted to the formulation and solution of the principal-agent model on the impact of transparency. In section 4 we apply the results of this model to explain the institutional differences between ancient Egypt and Mesopotamia. Section 5 concludes by briefly applying the logic of our general theory to understanding the impact of the Industrial Revolution on the scale of the modern state.

[10] For much earlier theories on the environmental determinacy of social institutions see Meek (1976).

[11] An extensive literature studies the impact of (exogenous) institutions on economic performance, see Parente and Prescott (2000) and Acemoglu Johnson and Robinson (2005). The effect of geographical factors on economic performance was examined by Sachs and Warner (1995). Galor, Moav and Vollrath (2009) study the effect of geography on inequality, and thereby on human capital promoting institutions. Easterly and Levine (2003) offer an empirical attempt to distinguish between the institutional and environmental determinants of economic growth.

2 The Neolithic Revolution and the emergence of the state

2.1 The literature on the emergence of ranked society and the state

Anthropologists and archaeologists have long concluded that the bands of hunter-gatherers were very egalitarian and ostensibly leaderless – in sharp distinction to the hierarchical nature of apes (Woodburn 1982, Boehm 1999). We shall not touch upon the literature that seeks to explain this egalitarianism, nor shall we discuss proposed explanations for the transition of hunter-gatherers to agriculture (Richerson, Boyd and Bettinger, 2001). In the extensive literature that seeks to explain why the Neolithic Revolution led to the creation of hierarchical societies, it was customary to distinguish multiple stages of social organizations preceding the state, such as tribes and chiefdoms. The inherent difficulties in distinguishing between these stages, though, led more recent scholars to shift away from such typologies into a more general discussion of the causes for increased "inequality" "stratification" and "complexity," and for the creation of a "ranked society" (see Price and Feinman, 2010).

Many alternative theories have been offered to explain the emergence of ranked societies – Ames (2008:493-494) summarizes more than a dozen such theories. The causal factors that are identified in these theories include: creation of surplus; increased population density; a sedentary form of living; storage; specialization and exchange; environmental heterogeneity and resultant benefits to trade; increased importance of property; competition and warfare; aggrandizing ideology; religion, and more. The significance of each of these factors is disputed, and the mechanism by which each one contributed to increased complexity is often rather ambiguous. To us it seems that these theories often identify correlates of the transition to agriculture as causal. We elaborate on some of these proposed explanations below.

We do not attempt to define what a state is. Early theories on the emergence of the state are usefully summarized in the compendium by Cohen and Service (1978) and by Haas (1982). They distinguish between two classes of explanations: integration theories and conflict theories. The integration theories (starting with Plato and Hobbes) posit that the emergent state was designed to perform various functions to enhance the welfare of its subjects. The three primary functions that are often cited are: providing security; enforcing contracts to facilitate trade; and constructing and managing irrigation systems. The conflict theories (starting with Rousseau's essay on the origin of inequality and the ideas of Marx and Engels) posit the pre-existence of a class society with differential access to property. The state is then envisioned as formed by the elite in order to resolve class conflict and to protect the elite from the lower classes.

Childe (1936), who was among the first to analyze the political evolution of society on the basis of archeological findings, and who coined the terms "Neolithic Revolution" and "Urban Revolution," in effect synthesized the above channels. He argued that the transition to agriculture resulted in food surplus and that this enabled individuals to specialize in non-farming occupations. This was concomitant with resort to trade, with the creation of an elite, and with "political integration." This led subsequently to the formation of city-states and state bureaucracy. In Childe's theory, all these developments were preconditioned by the availability of surplus from the countryside to feed those who did not produce food.[12]

Another pertinent integrationist approach was the "Hydraulic Theory," associated mostly with Wittfogel (1957). Following others, Wittfogel noted that the early civilizations emerged in river basins in which extensive irrigation projects were undertaken. He posited that this was not a coincidence and that strong despotic states were formed in riverine environments where there was a potential for irrigation systems to raise agricultural productivity.[13] Wittfogel apparently envisioned irrigation projects, like major dams and mega canals, whose construction could not be undertaken without a despotic central authority due to major fixed costs. At the same time he presumed that a despotic central management was required to manage such irrigation systems, once they are constructed. His theory met strong criticism. The critics pointed out that the irrigation systems in the early civilizations were constructed by local communities, prior to the emergence of a strong central state; that even after the emergence of such states the management of these irrigation systems remained to the local elites; and that the few large-scale irrigation projects that were set up by some of the central states in antiquity were the consequence of central governments,

[12] Childe ([1942] 1952:24, 69) provided the following summary of the background for the Urban Revolution: "society persuaded or compelled the farmers to produce a surplus of foodstuff over and above their domestic requirements, and by concentrating this surplus used it to support the new urban population of specialized craftsmen, merchants, priests, officials and clerks." Halstead (1989) summarize early controversies on the concept of 'surplus' in this context and noted that Childe's argumentation suffered from circularity "in which the elite is a precondition of its own existence" (p. 69) – since an elite was required to generate the surplus to create itself. Halstead sought to resolve the problem by positing that early farmers on their own generated "normal surplus" above their subsistence in average years in order to buffer against expected future shortages. Given limitations for direct storage of this "normal surplus" at the household level, Halstead argued, an elite was formed as "social storage agent," to coordinate between surplus and deficit households. Our argument is quite different and pertains to significantly earlier phases of the rise of complexity. Unlike Halstead, the storage that we focus upon is not due to farmers' intention to buffer inter-annual fluctuations, but rather to the seasonality of grain, that necessitates storage.

[13] Wittfogel (1957:15-19) invoked a Hobbes-like consent scenario whereby farmers in these early civilizations elected to realize the potential of the productivity advantages of the "hydraulic potential," even though they cognized that it entailed that they had to "subordinate themselves to a direct authority" and to accept the "loss of personal and political independence."

rather than the cause for their immergence.[14]

Other theories ascribe a crucial role to "population pressure." It is argued (see Johnson and Earle, 2000) that the increased productivity of agriculture led to population growth, and then to overpopulation and to the deterioration of living conditions. This "population pressure" then led to a fierce competition over resources and to resort to violence and warfare. This, in turn, necessitated the reorganization of society into ever more complex social forms, leading ultimately to the formation of the central-state. Carneiro (1970) proposed an influential variant of this scenario with an additional geographical proviso. He, too, posited that population pressure led to intensified competition and warfare. Adopting a Marxian approach he further claimed that the state was created to serve the interest of the winners in extracting surplus from the losers. However, according to his "circumscription theory," this outcome pertains only to a circumscribed environment that is surrounded by regions with low productivity, such as deserts. The rationale for this proviso being that circumscription is required to bar the losers from escaping subjugation to the winners through migration to nearby areas. This environmental condition seems particularly applicable to Egypt. Indeed, Allen (1997) argued in detail that Egypt's extreme circumscription is the key factor that distinguished its institutions from those of Mesopotamia.[15] Carneiro's theory has been criticized on empirical grounds as inapplicable to various early states. As Allen (1997:138) admits, a major problem with the applicability of this theory to Egypt is that population density was low when the Egyptian state was formed at about 3000 BCE (see also Butzer 1976:82-85).

2.2 Our critique and the role of appropriability

As noted, the prevailing literature attributes the emergence of ranked society and the state to the increased productivity of agriculture. This increase in productivity is assumed to have generated food surplus (in the sense of abundance). In turn, surplus is assumed to have facilitated specializ-

[14] See Adams (1960:280, 1981) with respect to Mesopotamia and Wilson (1960:130-131) and Butzer (1976:110-111) with respect to Egypt.

[15] Mann (1986:75, 110-112) emphasized two additional environmental factors to explain Egypt's "uniqueness:" the ease of communication and central control due to the location of practically the entire population along the Nile, and the lack of "essential metals" that induced "foreign military expeditions." From our perspective, Egypt's circumscription may have helped its central state for a different reason. Rather than inhibiting its population from escape, Egypt's surrounding deserts provided natural insulation from external threats. In particular, once a central state was formed, its monopoly on the use of violence was not seriously challenged by nomad bandits from the desert margins of the Nile valley, since these deserts could not provide the nomads with adequate alternative living and refuge. Hence, unlike the case of Mesopotamia where nomadic bandits posed a perennial problem, there was no on-going need in Egypt for fortified regional centers (other than the temple precincts that were typically surrounded by walls).

ation in crafts, exchange, and the creation of elite. Alternative explanations assign a major role to population pressure that was induced by the increase in productivity. As argued above, without denying that an increase in productivity did occur, we contend that these proposed mechanisms are flawed in that they are inconsistent with the Malthusian framework of endogenous population that prevents surplus. The slow rate of technological improvements up to the Neolithic period, during the transition to agriculture and, subsequently, up to the modern age, implies that the size of the population could have easily kept pace with these advances to leave no surplus.[16]

In particular, once population is considered as endogenous, it cannot be employed as an explanatory variable. The non-feasibility of out-migration in Carneiro's variant of the 'population pressure' theory is similarly irrelevant since such migration could not have curtailed the winning elite from extracting surplus from any conquered land: if land is fertile enough, there would be people to populate it. In fact, the Malthusian argument implies that there would (almost) always be population pressure and in the long run people will live wherever they can.

Our counter proposal is that it was not surplus that created the state, but the reverse: it was the early state that generated surplus through expropriation. More specifically, we argue that it was a change in the ability to confiscate consumable output that explains the formation of social hierarchy and the ultimate emergence of the state. By expropriating output from the peasants, the elite and the state would have, in effect, reduced the peasant population. Similarly, from our point of view, what is called 'population pressure' did not 'necessitate' a state; but rather, increased 'population density' has reduced dispersion and rendered the tax technology more productive (even if the production technology did not change).[17]

As noted in our introduction, the aspect of the transition to dependency on cereals that we consider to have supplied the critical ingredient for the subsequent political changes is the necessity of storage.[18] Not only that grain is non-perishable and thus facilitates food storage, the seasonality

[16]Steckel and Rose (2002) report that skeletal evidence from the Western Hemisphere reveals that the transition to agriculture did not produce any gains in health (and in fact may have impacted health negatively). Ashraf and Galor (2011) provide supporting evidence for the Malthusian claims by showing that technological improvements before the industrial revolution have had a positive effect on the size of the population but no effect on income per capita. Aiyar, Dalgaard and Moav (2008) and Vollrath (2009) study technological progress and population dynamics in a Malthusian environment, while Galor (2005) supplies an overall framework. This literature, however, has not considered endogenous population in the context of the emergence of the state and of related social institutions, nor have existing theories of endogenous population considered the effect of the state on population (other than in the modern, non-Malthusian context).

[17]It is, indeed, often claimed that one of the major historical reasons for weak states in sub-Saharan Africa was high population dispersion (due to low land productivity) that hindered tax collection.

[18]The conflicting theories on storage or surplus as the prime causal elements in the emergence of proto-state institutions can be traced to Adam Smith and his contemporaries (see Meek 1976). In his conjectural theory of

of grain production makes substantial storage necessary. This inherent aspect of the new production technology in turn transformed the 'tax technology,' that is, the ability of non-farmers to expropriate food on an on-going basis.[19]

The argument that storage was the critical factor that explains the emergence of social inequality is not new; it was raised already by Testart (1982).[20] In his comparative study of early hunting-gathering societies in the northwestern coast of America, Testart concluded that hunter-gatherers who relied on seasonal and storable resources (such as preserved salmon or oak acorns) became more complex and acquired social features like those of the Neolithic societies that cultivated cereals. As his critics pointed out, however, Testart identified correlations and refrained from identifying the mechanism that related storage to social inequality.[21] Furthermore, his focus was on hunting-

human evolution, Smith posited four sequential "ages": of hunters, shepherds, agriculture and commerce. He argued that property, inequality and government could not have existed among hunters, but emerged with the transition to pastoralism: "Among nations of hunters, as there is scarce any property, or at least none that exceeds the value of two or three days' labour, so there is seldom any established magistrate or any regular administration of justice. ... Where there is no property ... civil government is not so necessary." (1976:709-710; see also 1978:201-8, 404-8). Smith argued that only the need to protect moveable property (flocks) from theft brought about government and the institutions for property protection (1978:16). But it was the subsequent transition to agriculture that generated surplus, division of labor and exchange, and that extended much further the role of the state (1978:409). It is evident, though, that the term 'surplus' was used by Smith in a dual manner. Following the Physiocrats, Smith took it for granted that agriculture generates surplus: "land, in almost any situation, produces a greater quantity of food than what is sufficient to maintain all the labour necessary for bringing it to market in the most liberal way in which that labour is ever maintained. The surplus, too, is always more than sufficient to replace the stock which employed that labour, together with its profits. Something, therefore, always remains for a rent to the landlord" (1978:162-3). But, at the same time, Smith often referred to 'surplus' also in a more limited sense of the produce of a commodity beyond one's planned consumption of that particular commodity, intended for exchange (e.g. 1978:31, 37, 180-181) – a usage that does not necessarily entail income above subsistence.

[19] To understand the role of the two features of non-perishablity and seasonality we suggest a thought experiment. Consider a band of foragers that relied solely on perishable berries that were available throughout the year. Suppose first that, due to some natural causes, berries gradually become more plentiful per unit of land. Would anything have happened beyond an increase in population density? We doubt it, though we consider below the possibility that sufficiently high population density could in its self enable on-going expropriation of a non-storable food. Next suppose that the supply of berries has not changed, but a technique was developed to preserve berries, enabling storage. Also in this case, if berries continued to be available throughout the year, we doubt that a ranked society would have evolved. Finally consider another scenario where along with the technique for storage, and with annual supply remaining unchanged, berries became available only during a brief season of the year. Only in this case, we contend, inequality would have developed – in spite of the assumed lack of improvement in food supply.

[20] Woodburn (1982) argued that the main feature of producing cereals was "delayed return," rather than the "immediate return" of hunting-gathering. It is indeed the case that in agricultural production the yield to one's labor is delayed. But this distinction, we contend, fails to identify the critical role another time gap: between the crop and its consumption, that underlies Testart's emphasis on storage.

[21] More accurately, Testart referred to all the "usual suspects;" referring to sedentary living, high population density, possible trading, prestige, and altered ideology. Among these, though, he referred also to the mechanism that we emphasize here: "stored food is the primary object of raids, and it may be stolen, monopolized by men of high status, or made the subject of rent or tribute" (p. 527).

gathering societies that became complex rather than on the mechanism that connects agriculture to complexity.

Testart's evidence, we argue, is complemented by that of Carneiro (1970). Carneiro's circumscription theory is based on evidence from the Amazons, where the inhabitants employed slash and burn farming techniques that require extensive rotation of land. He deduced from the observation that such an environment did not induce social complexity because it was not circumscribed. However, the tuber crops that were cultivated in the Amazon where highly perishable (once harvested) and rather unlike the seasonal cereals (or the maize that was raised in the Andeans) that require year-round storage. That is, we interpret the evidence by Testart and Carneiro as demonstrating that agriculture was neither necessary nor sufficient to explain social complexity, whereas the existence of storage and its related facets of transparency were.

The main mechanism by which food storage creates an inherent role for leadership and inequality was mentioned already. Stores of food can be stolen and thus invite robbers, or as Olson (1993) called them, roving bandits. Banditry existed also among hunter-gatherers, and probably existed wherever it was profitable. The availability of stores of food made robbery more profitable; thus making the Neolithic Revolution different from the previous technological improvements (like the invention of the bow) throughout the preceding millennia of human hunters-gatherers. The returns to scale that exist in violent conflict imply a need for cooperation and leadership in order to protect stores of food. As is often argued, it is likely that this was the reason for protective walls around early agricultural settlements (including a massive wall in Jericho in the seventh millennium).[22] In

[22] Communal storage managed by a leader may have been prompted by two additional considerations. Since there are economies to scale in storage (the volume of a sphere increases less than proportionally with its surface area), it was more economical for villagers to cooperate in protecting their stored grain from natural elements. In addition, the need to ration grain-use throughout the year, particularly in years of famine, would have entailed a serious problem of self-control that could be alleviated by subjecting oneself to a more farsighted leader. A dominant strand in the anthropological literature maintains that central storage played a major causal role in the emergence of complex society (see Johnson and Earle 2000:251-256, 301-302 and Halstead 1989). Under the influence of Polanyi (1944), it is commonly argued that the early agricultural societies were "redistributive," where produce was first (voluntarily) transferred to a central authority, the pooled produce was then mostly "redistributed," but in part also stored at the center overtime as a buffer to the community against future shortfalls. Applying an integrationist approach, this theory posits that the key role that is served by the central authority is that of insurance. Pooling and redistributing provide mutual insurance against idiosyncratic fluctuations in the produce of members of the pool. On the other hand, central storage provides the community with an essential means for contending with the aggregative risks of food shortages. Needless to say, our proposal for the prime role of storage is diametrically different. The storage that we emphasize takes place by the farmers, rather than at the center, and it is conducted primarily inter-annually as a technical necessity due to the annual seasonality of cereals, rather than for long-term insurance purposes. Furthermore, what we emphasize is the potential of stores of food to be expropriated (by the center) rather than as designed by the center for the benefit of the periphery. On a practical side, we conjecture that idiosyncratic shortfalls to individuals were taken care of within a rather small kin group (as was the case among hunter-gatherers) and

turn, the submission to a leader undoubtedly created inequality.

Olson's theory was somewhat different. He noted that in a state of anarchy, roving bandits are likely to be replaced by stationary bandits, simply because the latter are forward looking and thus internalize some of the benefits that farmers have from securing their produce from total confiscation.[23] From our standpoint, though, there is no need to distinguish between Olson's idea of an outside roving bandit that turned stationary and the alternative idea that a consensual leader from within arose to protect the villagers from external roving bandits. The main point that we seek to make, and that had not been properly acknowledged in the relevant literature, is that either way, the main aspect of the Neolithic Revolution that gave rise to social hierarchy was not the increase in productivity per se, but rather the induced change in appropriability that accompanied the switch to cereals.

Seen from this perspective, one may acknowledge that not only storage, but also the high population density that resulted in some areas from the transition to agriculture may have made appropriability easier. Note, however, that according to this channel, it is not that population density created a need for government (as is typically claimed), but rather that higher density may have made the early farmers more vulnerable to exploitation by creating an opportunity for potential exploiters.

As Olson emphasized, the extractive local leaders had also private incentives to contribute to their subjects' well being. In particular, it was the joint interest of the farmers and their leader that the latter would monopolize the use of violence and fend off roving bandits. In addition, Olson argued that stationary bandits like the Egyptian Pharaohs would find it profitable to invest in irrigation projects, in order to expand the productive capacity of the territory under their control.[24] In this important case, however, we contend that the prime direction of causality may have worked

did not need proto-state institutions. On the other hand, we doubt the significance of central stores in serving the community at large (rather than the leaders at the center) as a buffer against aggregative risk.

[23] Olson's theory may be classified within the conflict theories, but it differs significantly from the conflict theories that Haas (1982) describes. To begin with, Olson's theory does not assume a pre-existing elite and does not resort to the idea of population pressure. Furthermore, both the integration theories and the above discussed conflict theories posit that the state was formed in order to serve a purpose: either to enhance the common good (in the integration theories, in which Haas included Hobbes' theory of providing security) or the interest of the existing elite (in the conflict theories). Olson's theory, on the other hand, is non-teleological: the state evolving according to Darwinian rules of natural selection. For a survey of earlier proponents of this non-teleological variant of conflict theory see Mann (1986).

[24] Olson (1993:568) dismisses in this context also the role of ideology and religion: "These violent entrepreneurs naturally do not call themselves bandits but, on the contrary, give themselves and their descendants exalted titles. They sometimes even claim to rule by divine right. Since history is written by the winners, the origins of the ruling dynasties are, of course conventionally explained in terms of lofty motives, rather than by self-interest."

differently. It is quite plausible, as we elaborate below, that in both ancient Egypt and ancient Sumer of the fourth millennium irrigation was initiated on a small scale by entrepreneurial patriarchs and were the main source of power of the early city-states. This reconstruction is in line with Wittfogel's critics who argued that the essential irrigation system in both countries preceded the central state. Our approach, though, goes one step further; in effect turning Wittfogel's argument on its head. Wittfogel's critics convincingly contradicted his claims that the need for irrigation led to a despotic state; but they failed to offer an explanation for the apparent correlation between riverine irrigation systems and strong states, leaving the impression that these two phenomena were only spuriously correlated. The perspective offered here clarifies that the issue is not that riverine irrigation systems require construction, maintenance and management (of course they do), but rather that such systems provide transparency and thus enable control and expropriation of the crop by the elite – in analogy to our interpretation that (required) food storage facilitates confiscation. As illustrated below, the control afforded by irrigation is nuanced, and admits alternative state institutions in different environments. Furthermore, from the present perspective there is also nothing unique about irrigation per se in fostering strong states, other than that it is a prime facet of agricultural technology that facilitates control.[25]

3 A model of Transparency and Land Tenure

3.1 Preliminaries

In this section we seek to extend our analysis by formulating the idea of how transparency of the production technology affects the scale of government and related social institutions, using a simple principal-agent model. Greif (1993) is the closest study to ours. He employs a game-theoretic model to examine alternative institutional arrangements among medieval traders, where merchants provide agents with incentives schemes that combine direct remuneration with a threat of potential dismissal. Greif (2005) applies several game theoretic models to study the institutional foundations for the development of modern impersonal market trade. Our model, in contrast, is a simple principal-agent framework that suits, we believe, the issues at hand. In particular, the historical setting is entirely different, with the concomitant shift of emphasis from trade to farming, land tenure and government control.

[25] In Adams' (1966:66-69) critique of Wittfogel's thesis we detect awareness of the differences between the issue of whether a strong central authority was required for the construction and management of the early irrigation systems (as Wittfogel argued), or whether even small scale irrigation systems facilitated administrative control (as is argued here). Adams convincing refutes the first possibility, but does not provide evidence to contradict the second one.

The prototypical principal agent problem has been often applied to study sharecropping contracts in agriculture. Ackerberg and Botticini (2000) provide a detailed survey of the literature in an historical context and test three hypotheses concerning the advantage of share tenancy over fixed rent contracts. These are: (1) risk-sharing: Sharecropping is typically claimed to offer a balance between incentives and risk sharing; (2) moral-hazard: Sharecropping is often claimed to better protect land from abuse; (3) imperfect capital markets: Sharecropping is claimed to allow poor tenants who lack resources and face borrowing constraints to be profitably employed in agriculture. On the basis of a sample of landlords and tenants from medieval Tuscany, Ackerberg and Botticini find support only for the last two hypotheses.[26] As is well known, compared to fixed rent contracts, sharecropping has the major disadvantage that it distorts incentives, since tenants receive only a fraction of the return to their marginal effort. It has long been suggested that sharecropping could be made more efficient if landowners provide additional incentives to their tenants by dynamically reallocating land (see Cheung 1969). This proposal is captured in our model by the threat of dismissal.

Another pertinent study is that by Domar (1970) of slavery in farming, which is also placed in the historical setting of pre-industrial economies. Domar argued that coercive labor arrangements (slavery or serfdom) are more likely to emerge where labor is scarce relative to land.[27] His approach, though, treats population as exogenous. This assumption can be maintained for a relatively short span of a few generations. Our model concerns the very long-run, where Malthusian forces are presumed to prevail, and it offers rather different insights on the practices of slavery and on related land tenancy across regions and over time. Our model is also related to the literature on "efficiency wages" and on unemployment as a discipline device. Increasing transparency in Shapiro and Stiglitz (1984) will lead to lower wages, as in our model, but also to more employment, in contrast to our long-term predictions below.

[26] In our model, we do not allow fixed rental contracts, which would have resolved the incentive problem in this context. We also assume that agents are risk neutral. Both features can be justified by appeal to Ackerberg and Botticini's finding, but our main justification for avoiding fixed contracts is that they in effect ignore the feasibility constraint in cases of low output, when the agent cannot possibly pay the fixed rent and survive. In this case, the overdue rent has to be postponed, turning the contract into something akin to the one that we analyze. We also do not consider the possibility of sequential sanctions against a supposed-delinquent agent, like gradually increasing conditional probability of costly dismissal. This would not have altered the gist of the model of identifying between regimes with or without dismissal. Given that our model has a stationary environment and a single agent type, it involves no learning, becoming a simple form of the branch of literature on career concerns, such as Gibbons and Murphy (1992) and Holmström (1999), and also of Doepke and Townsend (2006) and Neeman and Moav (2010).

[27] Along this line, Lagerlöf (2009) offers an explanation for the rise and decline of slavery, based on population dynamics.

3.2 A Principal-Agent Model with differential transparency

The principal (the state) determines the size of land-plots and then allocates these plots to the agents (the state's subjects, or tenants) and designs a contract so as to maximize her expected income.[28] We thus perceive the state as an absentee landlord. Each agent chooses his effort level so as to maximize his expected income net of effort cost. We start by analyzing the case where plot size is given exogenously, and focus on the effect of the transparency of economic activity on the type of contract and on the division of income between the state and its subjects. We then extend the model to incorporate endogenous plot size, which in effect determines the magnitude of the population as a function of transparency.

The principal employs an agent who chooses how much effort to exert. The payoff to the principal is the output produced by all agents, less the payments to the agents and other expenses. The payoff to the agent is the payment he receives from the principal, less his cost of effort. We assume that the principal and the agent are both risk neutral.[29]

We simplify further by assuming that the level of output 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 can be either good or bad: $\theta \in \{G, B\}$. The ex-ante probability that the state of nature is good is denoted by: $p \in (0, 1)$. Output is a function of the effort and the state of nature; in particular output is

[28] This model has only one tier of hierarchy. Hierarchy in reality is of course multi-tiered, and this aspect will in fact assume prominence below, when we compare Egypt and Lower Mesopotamia. Appendix C examines a case of two tiers of hierarchy and demonstrates that the logic of the model is extendable to the multi-tier case.

[29] That pertinence of uncertainty in this context should be clarified. Adams (1981:244, see also 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, and Halstead (1989) with respect to ancient Greece. Given that in the former two regions much of the uncertainty was aggregative, these scholars envision that the prime method for insuring the farmers against shortages was through inter-annual storage at the state level. There is indeed ample evidence for storage at the state level in these two societies. However, as we argue later on, we doubt that storage at the state level served as an important mechanism to sustain farmers' in years of famine. Farmer's probably lacked political clout, making it doubtful that they could expect support from the central government in years of general distress. It makes more sense that long term storage at the center served primarily the urban elite and its urban dependents, and that it was the latter group that was most vulnerable to famine. The role of exogenous uncertainty in our model is of a very different nature. We focus on moral hazard, rather than on insurance, under the perception that these two by-features of uncertainty operate in opposite directions. The ability to sustain shortages is a major source of advantage for land consolidation by the rich; moral hazard, on the other hand, is probably the main source of advantage for small land proprietors. The assumption that farmers are risk neutral enables us to focus on moral hazard. We admit that this assumption is problematic in the context of subsistence farming, but this assumption is mitigated here in that we explicitly posit that farmers' subsistence is guaranteed, even in years of low output. This last assumption reflects our presumption that farmers would have desisted in years of famine from surrendering to the absentee landlord food to the point of jeopardizing their own survival.

high if and only if the state of nature is good and the agent exerts high effort:

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

The pertinent timing of the effort and the state of nature is not evident.[30] We assume here that the agent chooses the level of effort before he learns the state of nature.[31] We assume further 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] This signal is accurate with probability $q \in [1/2, 1]$.That is, we assume 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 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 $m > 0$. We assume that this maintenance with high effort is $m + \gamma$, where $\gamma > 0$ is the periodic cost of exerting 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 for the agent to exert effort: $p(H - L) > \gamma$. We presume that the agent has no alternative sources of income or wealth, does not save from one period to the next (not even by storage) and cannot borrow. The agent's only alternative employment is perceived to be as a domestic servant outside of farming, 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 effort, this

[30] The state of nature that is relevant for agriculture is in fact 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, but particularly the spring flooding. In rain-fed Upper Mesopotamia the main source of uncertainty is the more idiosyncratic magnitude and the timing of rains. In addition, there are of course all the other sources of natural uncertainties, like blight or locust.

[31] Appendix A 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 will circumvent this complexity by assuming that the principal cannot condition the payout to any agent on the output of nearby plots, and that the signal about the state of nature of each plot already incorporates what could be learned from other plots nearby.

periodic utility is: $U = I - (m + \gamma)$.[33] We denote the present value of the agent's utility from being employed in agriculture by $V \geq 0$, and denote by $\delta \in (0, 1]$ his discount factor.

We assume that the principal relies on the following incentive scheme. If output is high, then the principal retains the agent with certainty and pays the agent $\omega + b$, 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 good ($\sigma = \tilde{G}$), then the principal dismisses the agent with probability $d \in [0, 1]$, and replaces the agent with another one.[35] If the signal indicates that the state of nature was bad ($\sigma = \tilde{B}$), the principal retains the agent with certainty. 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. We assume that this cost is large enough to ensure that it will not be desirable to dismiss the agent when the output is low ($Y = L$) and the signal is bad ($\sigma = \tilde{B}$). This requires: $x > \delta p\gamma/(1 - p)$.[36]

We find it instructive to consider the bonus b as a "carrot" and the dismissal probability s as a "stick." The optimal employment contract in this environment can then be recast as balancing between the provision of incentives through a "carrot" or through a "stick." This balance turns out to depend on the precision of the public signal q.

[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 complicates the analysis without adding any significant insight.

[34] By the nature of farming, this payment is conducted after the harvest and feeds 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 access 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] 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 agents that the principal employs as replacement are presumed to be men who have come of age and who (unless taking over from their deceased fathers) would otherwise have to become domestic servants. Our presumption that the principal pays farming agent their long term subsistence, regardless of the output produced, implies that the farming population will not shrink. Our complementary assumption is thus that any additional population will have to be absorbed outside agriculture, where its long term survival is not guaranteed.

[36] In evaluating the extent of loss due to dismissal, from the principal's point of view one has to consider also its timing. The agent-tenant was probably dismissed after the harvest, and unlikely gave up his retained share of that harvest. This meant that the landowner had to come up right away with an advance to feed the new tenant for his labor, until the next harvest arrived.

The above principal-agent problem is thus 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 below, but for now it is taken as if it was exogenous. The principal's objective is to solve for the employment contract that maximizes her periodic expected payoff, denoted by π:

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

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

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

and subject to the subsistence constraint:

$$\omega \geq m + \gamma.$$

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 and d, the carrot and the stick.

We can then rewrite the principal's objective function (OF) and the agent's incentive constraint (IC) more succinctly as follows:[37]

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

s.t.

$$pb + pqd\delta V \geq \gamma \qquad (IC)$$

3.2.1 Two Types of Contracts: 'Pure Carrot' or 'Maximum Stick'

The linearity of the principal's objective function and the agent's incentive compatibility constraint in b and d implies that in the optimal solution, either $b \geq 0$ or $d \in [0, 1]$ must be pushed into a

[37] The (IC) constraint implies that the agent benefits from exerting effort. To see that the left hand side indeed equal 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. In particular, $Pr(Dismiss|e = l) - Pr(Dismiss|e = h) = s[Pr(\sigma = g) - Pr(B)Pr(g|B)] = s[[Pr(G)Pr(g|G) + Pr(B)Pr(g|B)] - Pr(B)Pr(g|B)] = pqs$. 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 (given V).

corner solution. We obtain that the principal will design the following two contracts, depending on the degree of transparency. Defining the threshold transparency:[38]

$$\hat{q}(V) = \frac{(1-p)x}{p\delta V + (1-p)x}, \tag{1}$$

then we obtain:

a 'pure carrot' contract, if $q < \hat{q}(V)$,

$$d_c = 0,\ b_c = \frac{\gamma}{p}, \tag{2}$$

or a 'maximum stick' contract if $q > \hat{q}(V)$,

$$d_s = 1,\ b_s = \frac{\gamma}{p} - q\delta V_s. \tag{3}$$

Our labels (and subscripts) for these two types of contracts are self-explanatory. In the 'pure-carrot' contract, the agent is never dismissed and is incentivized only by the bonus. On the other hand, in the 'maximum-stick' contract, the agent is dismissed with certainty whenever the output is low yet the signal is good ($Y = L, \sigma = \tilde{G}$).

The reason that the optimal contract switches from carrot to stick when the quality of information improves is as follows. A principal relying on a 'stick' to incentivize the agent has to incur the fixed cost of dismissal, x, whenever a dismissal takes place. When the agent exerts effort, dismissal occurs with probability $(1-p)(1-q)$ – when the state of nature is bad and the signal is deceptive: $\theta = B$ and $\sigma = \tilde{G}$. Hence the probability of dismissal, and thereby the expected cost of dismissal, decrease with the quality of information, q. Only when q is large enough is it worthwhile for the principal to incentivize the agent with the stick. The threshold $\hat{q}$ balances between this expected cost of dismissal $(1-p)(1-q)x$ and the expected savings to the principal in reduced expected bonus payments $p(b_c - b_s)$ due to the threat of dismissal, which by (2) and (3) equals $pq\delta V_s$.

The optimal contract that was derived thus far still depends on the endogenous value V. In order to solve for the optimal contract explicitly, in terms of its parameters only, we need to substitute away for the value of V. 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:

[38] The threshold is identified by continuity of the objective function. If $q = \hat{q}(V)$, then any pair b and s that satisfies the agent's incentive compatibility constraint is optimal. Observe that the 'maximum stick' contract cannot be such that b = 0 because in this case the agent's utility if he exerts high effort is zero, and he may as well attain a positive utility by shirking.

$$V = [\omega + pb - (m + \gamma)] + [1 - Pr(Dismiss|e = h)]\delta V$$

Given that $\omega = m + \gamma$, and that the probability of dismissal upon high effort is $d(1-p)(1-q)$, V becomes:

$$V = \frac{pb}{1 - \delta[1 - d\,(1-p)\,(1-q)]}. \tag{4}$$

This implies that under the 'pure-carrot' regime (2), V is:

$$V = \frac{\gamma}{1-\delta} \qquad if\ q < \hat{q}. \tag{5}$$

Solving from (3) and (4) for b and V in the 'maximum stick' regime yields:[39]

$$V_s = \frac{\gamma}{1 - \delta(p + q - 2pq)} \qquad if\ \ q > \hat{q}, \tag{6}$$

and

$$b_s = \frac{\gamma}{p} - \frac{q\gamma}{1 - \delta(p + q - 2pq)} \qquad if\ \ q > \hat{q}. \tag{7}$$

By checking the principal's objective function (see below), the threshold level of transparency $\hat{q}$ turns out to be determined by a quadratic function that is expressed by the implicit condition:

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

This quadratic equation 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.

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

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

and

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

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

And their combined expected income is:

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

It is readily observable that in 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. All these features are lost in the range of high transparency. At this range of 'maximum stick,' where $q > \hat{q}$:

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

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

and

$$I_s + \pi_s = p(H - L) + L - (1 - p)(1 - q)x. \tag{14}$$

As the expected total income indicates, the 'maximum stick' contract entails the inherent inefficiency of dismissing the agent – even though he works diligently. This efficiency loss equals the expected cost of dismissal, $(1 - p)(1 - q)x$, and declines as accuracy improves. When the signal is fully accurate ($q = 1$), the 'maximum stick' regime become efficient. By construction, the principal's payoff is continuous at the threshold 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 from 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 'maximum stick' regime: at the threshold accuracy $\hat{q}$ his expected income I (and his expected periodic utility) drops by the expected cost of dismissal $(1 - p)(1 - \hat{q})x$. After that threshold, his expected periodic income (and periodic utility) continues to decline. 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 (the carrot), while still maintaining the incentive constraint.

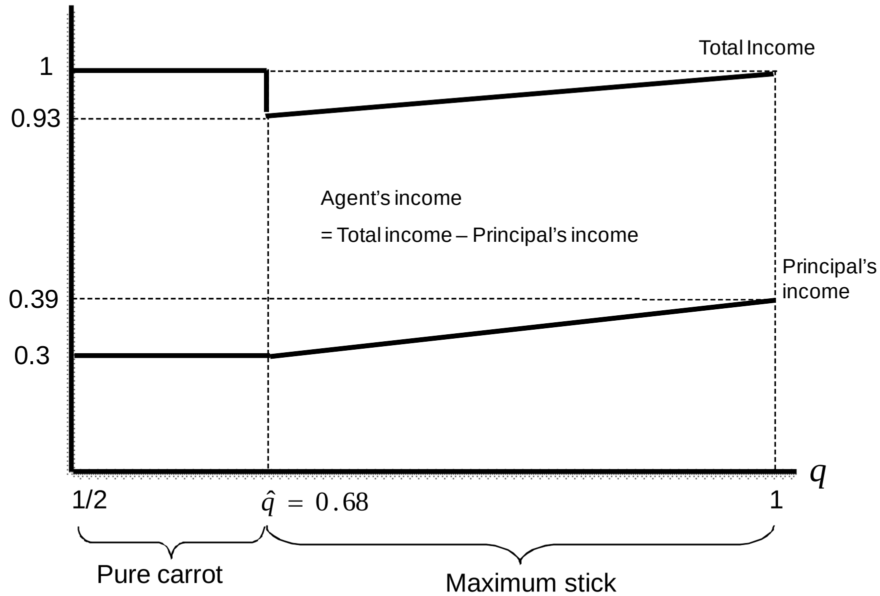

Figure 1 Periodic expected income as a function of signal accuracy

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 line; with the balance between the 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: $E(Y) = pH + (1-p)L = 1.$[40] 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. In addition, we set $x = 1$, so that the present value loss of replacing an agent is one expected crop.[41] To be consistent 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 $\gamma = 0.1$. 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, we set $\delta = 0.8$.

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, unlike the case in the range $\hat{q} \leq q < 1$.

[40] One should think of this unit as representing about 1.5 tons of grain of net output, in the sense that it is 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 model and to calibrate early Near Eastern farming see Hunt (1987), even though in his model the labor input is neglects.

[41] We think of this loss as the present value of the reduced output that may stretch over more than one year, due to the need for a new agent to acquire the specific skills of the dismissed one.

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 opaque signal range 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 would in fact dissipate entirely if he were fully patient ($\delta = 1$).

3.3 Transparency and Population Density

We now enable the principal to adjust the plot size that we denote by λ. The output from a plot of that size is assumed to be:

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

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. Thus, her problem becomes:

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

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:

$$d_c = 0; \qquad b_c = \gamma(\lambda)/p; \tag{15}$$

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:

$$m/2 = \gamma(\lambda_c)\eta(\lambda\gamma). \tag{16}$$

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

Similarly, in an informative environment where q is sufficiently high, the optimal contract is of the 'maximum stick', and according to (7):

$$d_s = 1; \qquad b_s(q,\lambda) = \frac{\gamma(\lambda)}{p} - \frac{q\delta\gamma(\lambda)}{1-\delta(p+q-2pq)} \tag{17}$$

The principal's problem, reduces then to selecting λ to minimize $T(m+\gamma(\lambda)+pb_s(q,\lambda)]/\lambda$.

Upon substituting from (17), this is equivalent to minimizing: $T[z(q)+\gamma(\lambda)]/\lambda$, where we define:

$$z(q) = \frac{m[1-\delta(p+q-2pq)]}{2[1-\delta(p+q-1.5pq)]} = \frac{m}{2}\frac{1+0.5pq\delta}{1-\delta(p+q-1.5pq)}. \tag{18}$$

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

$$z(q) = \lambda_s\gamma'(\lambda_s) - \gamma(\lambda_s) = \gamma(\lambda_s)\eta(\lambda_s). \tag{19}$$

Under our concavity assumption on $\gamma(\lambda)$, the right hand side of (19) increases with plot size λ. Since $z(q)$ increases in q, and since by (18) $z(q) > m/2$ for any q, conditions (16) and (19) imply that plot size λ_s in the 'maximum stick' regime will increase with transparency q, and will be larger than λ_c. This is due to the fact that under the stick regime 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 'maximum stick' 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 'pure stick' 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.[42]

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 may seem like confirmable predictions; however the presumption that all else remains the same is daunting. This presumption, undermines the possibility of comparing rain-fed upper Mesopotamia with irrigated lower Mesopotamia, and it also undermines the comparison of lower Mesopotamia, where land was typically fallowed every alternate year, with Egypt, where agricultural land was cultivated without fallowing.

3.4 The Urban Sector

We have considered the size of the farming population, but have not yet discussed the non-farming sector that we also interpret as the urban sector. To simplify, assume that m is the long-term maintenance cost of a household in the urban sector, and that a part W of the state's food revenue is 'wasted' (say, on sumptuary meals or on imports of prestige goods from abroad).[43] The expected non-farming population may then be estimated to be: $(\Pi - W)/m$. However, this crude calculation neglects to consider the dynamics of food supply in the urban sector.

In the pure carrot regime, the agent's income in good years is:$\omega + b_c = m + \gamma + \gamma/p$, and in bad years it is: $\omega = m + \gamma$. In this regime, the principal receives from this particular land plot

[42] Assuming $c(\lambda) = c\lambda^{1+\eta}$ and positing that $\lambda_c = 1$, implies that $\lambda_c^{1+\eta} = m/2c\eta$ and requires: $\eta = m/(2c)$. Under this specification, $\lambda_c^{1+\eta} = \theta(q)/c\eta = 2\theta(q)/m$. Under the above calibration, the threshold $\hat{q}$ falls from 0.68 to 0.65 and the plot size λ_c increases by 11 to 14.5 percent. Each tenant's expected utility pb increases beyond $\hat{q}$ from about 3.6-2.4 percent of expected output to about 5.3-3.8 percent. For reference, we also note that in an equilibrium without food expropriation by non-farmers, the Malthusian population size could be estimated at $N_m = T/\lambda_m$, with land size per household satisfying: $m + c(\lambda_m) = \lambda_m EY$. Under the above parameters, we obtain: $\lambda_m = 0.5045$, so that the potential population size in the absence absentee land ownership (and the state) would have been almost doubled.

[43] If farmers are employed (during the summer) in building monuments, and are paid for their extra effort by the state, as was customary Egypt, this would also be included in W, but in that case we would have to recalculate the food supply in the farming sector.

$H - (m + \gamma + \gamma/p)$ in a good year, and $L - (m + \gamma)$ in a bad year. Under our assumption that effort is efficient, $p(H - L) > \gamma$, both the agent and the principal suffer from a bad year. But their respective losses are not proportionately the same. In fact, in our illustrative calibration above, where we set: $L = m + \gamma$, the principal's payoff in a bad year is zero.[44] This dire outcome is mitigated however if the principal possesses very many land plots that are not perfectly correlated. In the unlikely case when the many land plots are statistically independent, the principal obtains the riskless total revenue $\Pi_c = T\pi_c$, where π_c is as in (10) above. But it is more likely that the correlation between the land plots will be substantial (as was the case in Egypt and in Southern Mesopotamia). Our analysis thus far leads us to posit that under such conditions, the food revenue that was available to the urban sector in bad years was particularly low.

The vulnerability of the urban sector of the ancient economies to aggregative downward shocks has two important, potentially testable, implications.[45] First, hunger and starvation are likely to be concentrated not in the farming sector, but rather among the lower strata of the urban sector: servants, small artisans and the like. This consideration 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. Secondly, this conclusion implies the bulk of the inter-annual storage of foodstuff is likely to be conducted in the urban sector by the elite; but that its prime purpose would be buffer the expected shortages in farming revenues in bad years. Under the conditions envisioned here, the urban inter-annual would not have served the farming sector.

4 Application

The main conclusions that we seek to draw from our model are qualitative. In particular, we obtained that even when the state has absolute power, an opaque farming technology implies land tenure institutions that are tantamount to farmers owning their land, under the provision that they pay a part of the output to the state. On the other hand, high transparency implies land tenure arrangements that are akin to serfdom, where the cultivator peasant has no title to the land and can be dismissed at the landowner's will. In addition, we obtained that the farmer's well-being (represented by utility) will be higher in the former environment. That is, from the agent's perspective, opacity of the production technology provides a shield from the coercive power of the

[44] More generally, the condition $p(H - L)/L > c/(m + c)$ is sufficient to guarantee that the output shares shift in favor of the agent in a bad year. We presume that this condition holds.

[45] Admittedly, this conclusion is based critically on our assumption that farmers' basic wage guarantees subsistence. As noted above, we believe that this simplifying assumption is defensible.

absentee landlord and the state, and thereby provides also some form of freedom.

In this section we seek to demonstrate that these results on the effects of transparency on appropriability shed significant new light on the institutions of the first state civilizations. Our assumption that the state can dictate the contract to farmers is based on a presumption that the state has already the power to coerce. That power, however, is mitigated here by informational asymmetries. According to our model a transparent production technology implies a 'stick' type relation, where the state employs the threat of dismissal to incentivize its agent; under an opaque technology, only a 'carrot' type of incentive will be used and the agent is freed from the threat of dismissal. Our goal here is to demonstrate that these insights are consistent with the different social institutions of the major civilizations of the ancient world. We will distinguish between three regions: Egypt, Upper Mesopotamia and Lower Mesopotamia. Our discussion of the social institutions in each of these regions is preceded by rudimentary summary of the nature of its farming technology.

Agriculture within the Nile valley is believed to have started only in the fourth millennium, yet the Egyptian state sprang from southern Upper Egypt already by 3000 BCE.[46] The farming technology that was applied at that time, known as "irrigation flood basin," prevailed with only minor modifications until the beginning of the twentieth century.[47] Figure 2 depicts an independent system of four connected flood basins. These 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 elevated natural levee and are separated one from another by man-made lateral dykes. Each of these basins would 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.[48]

[46] The term Upper Egypt refers to the Nile valley (south of Cairo). The role of the Nile's delta (north of Cairo) gained greater importance only subsequently.

[47] Our description of the flood basin method is based on Butzer (1976) and on the report by Willcocks (1899) of the system that operated in Upper Egypt in the late nineteenth century, prior to the construction of the (low) Aswan Dam. The main feature that distinguished the basin system of the 1880's from the one of the third millennium was the addition of extended longitudinal canals that enabled the cultivation, and sometimes even of perennial cultivation, of elevated areas in the desert's margin and along the Nile's banks.

[48] The width of the Nile valley varies from about two kilometer 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 1880's there were 212 basins in Upper Egypt, with an average size of about 27 square kilometers. Based on Hassan (1994:165), a basin of this size would have been populated in ancient times by about 1,400 families.

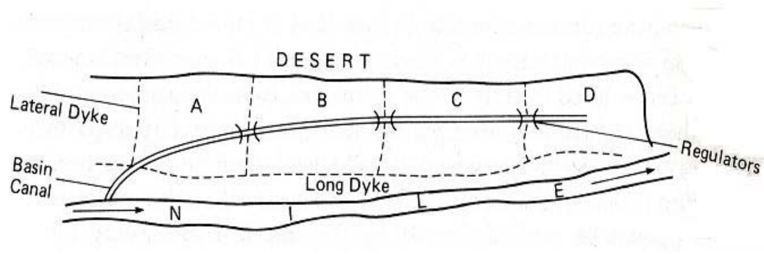

Figure 2: A schematic flood basin system typical for Upper Egypt. Source: Lloyd 1983:327, adapting Willcocks 1899:610.

The Nile's main source of water is the early-summer Monsoon rain in eastern Africa. When the Nile's water rose sufficiently during the period of inundation –in mid-August – the local inhabitants would breach an earthen dam at Nile's embankment to open up the basin canal (see Figure 2) to flood the basin system. After the lowest basin (D) was filled up, the inlet regulator was sealed and the higher basin (C) was filled up, and so on until the entire system was filled up. The feeding breach in the Nile's bank was then sealed, creating a terraced system of lakes, each one covered by about 1.5 meters of standing water. The muddy water would remain in the basins for about 40 days, depositing its rich mineral nutrients and soaking the soil. Then (in early October), after the water level of the Nile receded, the basin system would be gradually drained back into the Nile by opening up an escape channel leading from the lowest section D to the river.[49] After a short while, when the soil dried somewhat, the peasant farmers would plow and sow their fields (mostly with barley). The seeds would then grow on the wetness that was trapped in the soil, without any further irrigation. Four months later (in March) the farmers' bottleneck season arrived, when they had to reap the ripened crop before the hot winds of April and May could parch it and cause the grain seeds to disperse while still on the stalk.

The fluctuations in the Nile's inundation level, however, necessitated adjustments to the regular method of operation. With low inundation levels, some basins would not be filled up sufficiently and food production would falter. Occasional high inundation levels could be even more damaging, overflowing the river banks, destroying the dykes and occasionally flooding the villages and destroying the locally stored grain. Even in normal years, the mode irrigation could be seriously disrupted, for example if one of the lateral dykes collapsed. And still, the above description makes

[49] Willcocks explained (p.59) that holding the water in the basins for too long engendered worms that destroy the crops and delayed the ripening of the grain. The drainage of the water back to the Nile also flushed away the salt that is naturally created in irrigated soil.

it evident how rather homogeneous the conditions were within each basin in any given year.[50] This description also makes it evident how significant was the role of the village headmen who organized the villagers to perform the necessary tasks of opening, sealing and repairing the dykes during the inundation season, as well as cleaning the canals from silt and maintaining the dykes in the dry season. Particularly important was the role of the nobleman who supervised the entire basin system and timed the inflow of the water and its outflow. By the nature of these decisions, they had to be made at the local level, and the required earth-work had to be performed cooperatively, sometimes on an extremely short notice.

The local-public good nature of the basin irrigation system thus explains why a local leader was needed, with the power to extract corvée labor services from the local farmers, but not much beyond that. One may conjecture that the basin system was initiated by an entrepreneurial village head in southern Upper Egypt who resided near a small natural potential basin (like D). He probably realized that the cultivation of grain inside the valley was hampered by that the entire watering of the crops had to be completed in advance of sowing. To cope with this difficulty, after burning down the natural vegetation during the dry season, he sought to trap the inundated water inside the basin for sufficient long time by sealing its outlet to the Nile. This conjectured scenario on the origins of the basin system helps explain why the earliest city-states were all formed in the southern end of Upper Egypt, where the Nile valley is narrow and such a rudimentary basin irrigation system could have been set up with only modest investment.[51] That is, the simple basin system was probably operated from the start under local leadership, and, in turn, served these local leaders to amass power and to become the mayor-kings of incipient city-states in the late fourth millennium. This led within a rather short span of time to the early unification of Egypt under a single ruler. Only subsequently did some of the Pharaoh's embark on more substantial irrigation projects that extended intensive agriculture into additional regions of Egypt.

This description clarifies 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. Indeed, Eyre (1994:72-73) challenges altogether the "received view" of "a strong centralized bureaucracy exercising detailed and uniform control over the productive resources [of Egypt]." He argues that this view is based on propagandist official proclamations and on the misplaced assumption "that effective irrigation and agriculture required a dictatorial regime." Eyre argues that the Egyptian central state's ability to

[50] Homogeneity was not perfect even within each basin, since the ground was not entirely level; a phenomenon that enabled for isolated swamps to persist at the lower, outer margins.

[51] See Butzer (1976:100-103).

tax was in fact aligned with a decentralized organization of locally controlled 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 role of the impersonal 'state' was typically that of delegation rather than direct control." (1994:74).[52]

The description above explains also why, in spite of the ex-ante non-predictability of the Nile's inundation, the farming activity in the Nile valley was, in effect, transparent to the central government. Indeed, evidence for the recording of the Niles' annual level of inundation exists already from the mid-third millennium (see Kemp 2006:64). The most detailed description of how taxes were levied in Upper Egypt derives, though, form much later periods. Cooper (1976:366) 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."[53]

This description clarifies not only how the Pharaoh could efficiently extract tax revenue from the provinces, but also how each governor could, in turn, determine how much tax was due from the noblemen who supervised individual irrigation basins. The homogeneity within basins likewise enabled the local noblemen to have effective control over the revenue received from the village headmen. It was the village community as a whole that was responsible for paying taxes. For that purpose, the village headman exercised tight control of the village land and in effect determined the land assignment of the individual farmer.[54] By custom the same fields was assigned to the same

[52]See also Butzer (1976:43). Taking it to extreme, Eyre (1997:375) quotes approvingly an assessment of Egypt's modern water system: "The central state pretends to regulate everything and in fact regulates nothing."

[53]Eyre (1999) argues 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. Baines and Yoffee (1998:206) write similarly: "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 also notes (1999:45) that the government often sought to gain more direct micro-information on individual plots and that such intervention was "characteristically resisted" by the villages: "The balance between the government attempts to control the taxation of land and the government's willingness to resign that control to local middlemen is a recurring theme." Indeed, the transparency of Egyptian farming to the center was due not only to the ease to observe the height of the inundation (in advance of the cultivation season), but also due to the relative ease of monitoring farming activity upon traveling by boat up the Nile. That this was done in practice is evidenced by the minutes of a monitoring expedition from about 1140 BCE, (the Wilbour Papyrus, see Kemp 2006:256) that (apparently) recoded rent assessments that were due to temple institutions from more than 2,000 plots of land in Middle Egypt that were leased to various agents.

[54]We formally address a setup with multi-tiered tax farming in appendix C where we show that our model captures the essential element that distinguished Egypt, namely that what happened at the local level was highly transparent all the way to the top of the hierarchy.

farmer, as long as he worked diligently; but farmers did not have any secure tenure and the village head or the estate manager could reassign fields as he saw fit.[55] This system at the village level, combined with the practice that the middlemen positions in the hierarchy were also dependant on revocable appointments from above, fits very well the 'pure stick' regime of our model when transparency is high.

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 (Baines and Yoffee 1998:206). But at the same time, it has been established that significant fraction of the land in Egypt 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 also by powerful individual nobles.[56] From our perspective, though, the more significant feature of the land tenure system in Upper Egypt is that even when the land was "private," it was worked by peasants who did not own the fields they tilled. This was the case, we argue, because the transparency of the farming robbed these peasants of the main protection that is otherwise offered by the existence of moral hazard. Or differently stated (see Figure 1), the high transparency of farming in Egypt enabled the state to obtain a larger share of the output without incurring a significant efficiency loss.[57]

[55]See Eyre (1997: 378; 1999:51-52). The transparency of Upper Egyptian farming is further manifested in that even in the cases were private land was leased in the mid first millennium (see Hughes 1952), the leases were for one year only and were negotiated after the inundation level became known.

[56]I do not refer here to orchard gardens that were apparently considered 'private' all along. For a survey of the land tenure regime in Upper Egypt see Manning (2003:65-98) whose main focus is on the Ptolemaic period. K. Baer (1962:25) claims that "private individuals could own farm land at all periods of ancient Egyptian history," but was able to collect only limited evidence (at least compared to Mesopotamia) concerning land prices from the New Kingdom period and later. Hughes (1952:1) states more cautiously: "in theory all the land belonged to Pharaoh throughout Egyptian history;" and then notes that at various periods estates were assigned to court officials and to military officers; and that much land belonged to the temples. Hughes states that at least in the first two millennia of the historic period (up to the late New Kingdom) there was never "a large body of small landholders holders who managed and worked their plots themselves ... the lowest classes were largely serfs on the domains of Pharaoh, the wealthy and the temples" (pp. 1-2).

[57]Eyre (1997) contends that the divorce between land-ownership and actual farming was endemic to Egypt and persisted essentially until the mid-twentieth century. Indeed, in his extensive studies on land tenure in Egypt and the Middle East since the nineteenth century, G. Baer noted how different Egypt was from the remaining parts of the Middle East. Among others, Baer noted that at the early part of the nineteenth century "Most of the land [in Egypt] virtually belonged to the state," but was effectively governed by tax farmers who obtained land right "by public tender in exchange for a sum of money" and for paying a land tax (1962:1). In the agrarian reforms of the nineteenth century, the state in Egypt abolished tax farming and introduced land ownership and direct collection of taxes by the state (1969:62). Parallel Ottoman agrarian reforms in abolishing feudal privileges in the Fertile Crescent met much stronger resistance and the state reverted effectively to a tax farming system (1969:64-65). In considering the factors that contributed to these differential developments Baer listed first: "1. The government in Egypt was of a centralist character, stemming from its importance in the economy as a regulator of irrigation by the waters of

The hierarchical nature of land management in Egypt explains in turn the limited extent of real-estate transaction in ancient Egypt. The absence of legal title to land implied also that land could not be used to secure loans, and explains the paucity of early records for loans and related commercial activity. It was the village leadership who resolved local legal disputes within the village, applying traditional common law (Eyre 1999:44). Disputes higher up in the hierarchy, were similarly resolved by the authority just above the disputants. This lean legal, economic and social hierarchical structure explains also why the Egyptian cities functioned primarily as administrative and religious centers, rather than as metropolitan hubs. These features, and the lack of inheritable titles to land, explain also the absence of written law codes in Egypt – the earliest codes extant are from the second half of the first millennium.[58]

The nature agrarian and related social institutions in Mesopotamia was significantly different than in Egypt. Mesopotamia, however, was also much more heterogeneous than Egypt. In particular, rain-fed Upper Mesopotamia (or Northern Mesopotamia, later known as Assyria) can be considered from our perspective as closer to the alternative polar case of opacity. Irrigated Lower Mesopotamia (or Southern Mesopotamia, south of modern Baghdad; also known as Babylonia), on the other hand, presents an intermediate case, though in many respect much closer to the transparency of Upper Egypt. Wilson (1960:128) provided a valuable overarching comparison of the irrigation systems in Egypt and Lower Mesopotamia. He described the basins irrigation system along the Nile as leaves that grow out directly out of a single "stalk of grass or bamboo;" while (Lower) Mesopotamian settlement pattern "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 rain-fed Upper and irrigated Lower Mesopotamia.

the Nile. . . . " He then emphasized that navigation in the Nile enabled the government better communication and control over its territory than was feasible elsewhere in the Ottoman Empire (1969:63). Baer (1966:79) listed several additional geographical factors that made the agrarian structure in Egypt different. These include also: irrigation in contrast to rain-fed agriculture; the relative regularity of the Nile as compared to the Tigris and Euphrates; the flatness of Egypt; the utter barrenness of the Egyptian deserts that diminished Bedouin threat; the greater defense capacity due to its isolation. Our identification of "transparency" as the key factor to explain the uniqueness of Egypt, does not preclude the significance of these other geographical factors, but provides, we believe, a more general explanation for the long-standing centralist tendencies in Egypt and for the relative weakness of the centrifugal forces there.

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

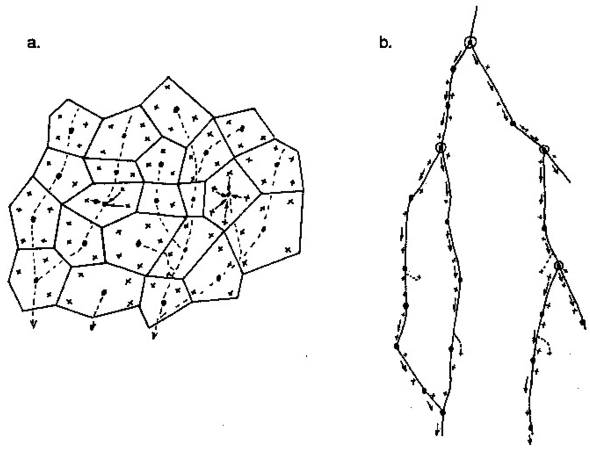

Figure 3: Schematic Settlement patterns in Upper Mesopotamia (a) and Lower Mesopotamia (b); Solid circles represent central settlements; crosses represent minor settlements (villages). Source: Wilkinson 2003:211.

It was at the highlands of Upper Mesopotamia that agriculture was first introduced into the country.[59] By the nature of rain-fed highlands, however, spatial weather conditions are rather idiosyncratic, and hence our characterization of farming in that region as opaque. Archaeologists have identified substantial ancient mound (Tells) in Upper Mesopotamia from as early as the late fifth millennium.[60] Wilkinson (1994, 2003:211) argued that the settlement pattern in Northern Mesopotamia was characterized by a scatter of a large number of roughly equivalent, nucleated administrative units, with a radius of about five kilometers. Each such unit was administered by a central settlement that relied on agricultural "surplus" 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" (1994:503). Wilkinson (2003:211) attributes this nucleated settlement pattern to the fact that no site had an "overwhelming situational or demographic advantage." Without disputing this observation, we contest the logic of this argument. By the increasing returns nature of violent conflicts, aggressive leaders are typically able to subjugate their neighbors. As in Egypt and in Lower Mesopotamia, an overwhelming a-priori advantage is not a prerequisite for the establishment of larger states. From the perspective advanced here, we suggest that that the key to the nucleated pattern of semi-autonomous administrative units in early Upper Mesopotamia was not the absence of a-priori advantage to any particular site, but the restricted

[59] We restrict our attention to Upper Mesopotamia, but it is widely recognized that the agrarian and social institutions that prevailed in this region were similar to those that prevailed in Syria and in rain-fed adjacent areas.

[60] See Wilkinson (1994) and Ur (2010).

ability to control and to extract on-going tax revenue, even after winning several territorial battles, due to the opacity of this region's faming activity.

The early tendency towards increased urbanization in that Upper Mesopotamia was puzzlingly aborted, however, in the later part of the fourth millennium – just when the first city-states started to flourish in Sumer. Furthermore, due to the lateness of written records from that region, rather little is known about its rural sector in the early millennia. On the basis of cuneiform evidence from the mid-first millennium it has been determined that the kings and elite owned large estates, that (unlike lower Mesopotamia) the temples did not possess much economic clout, and that much land was also owned by nuclear families who worked their patrimonial land. It seems, though, that land ownership in Upper Mesopotamia was in a constant flux, with small land holders regularly losing their ancestral fields to rich families through debt and sale under duress (see Zaccagnini 1999 and Jas 2000).[61]

Within Lower Mesopotamia (or Babylonia), Scholars typically distinguish between northern Babylonia (also identified as Akkad) and southern Babylonia (Sumer). Agriculture in the arid alluvial plains of Lower Mesopotamia depended entirely on irrigation from the Euphrates or the Tigris. Unlike the Nile that feeds from Monsoonal summer rains and enables standard autumn sowing, the rivers in Mesopotamia are fed by winter rains and by spring melting snow in the mountains to the north and east. As a result, the rivers in Babylonia are low on water in October-November and swell only in April-May. This mismatch of shortage and excess required the resolution of two major problems.[62] First, the cultivation (in winter) when the rivers where low required canals (unlike the simple breaches in the levee that sufficed in Egypt), and the shortage of water implied that not all the potentially cultivable land in Babylonia could be cultivated.[63] Secondly, to avoid the danger of

[61] The persistence of owner-occupied farming in Upper Mesopotamia indicates that the process of land consolidation due to indentured poor farmers losing their land must have been matched by an opposite process of the gradual dissolution of large, presumably less efficient, estates. Interestingly, the record from Nuzi from the mid-first millennium indicates that rules were enacted in an attempt to preserve patrimonial land. 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 and became a tenant on his former land (Zaccagnini 1984). This pattern of land tenure was typical throughout the rain-fed regions of the ancient Near East. Jas quotes Warriner (1948:21, 104) who claimed that "the subject of Iraq land tenure is one which has led many observers to fear the loss of their sanity," but noted that the different 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. . . "

[62] See Adams (1981:3-6) and Postgate (1994:178). Harvesting had to be conducted by April-May, due to the same hot parching winds that affected Egypt.

[63] Adams (1981:6) estimates that only 8,000-12,000 square kilometers could be cultivated, out of a potential estimated by Wilkinson (2003:76) at about 50,000 square kilometers. Thus, unlike rain-fed upper Mesopotamia, a piece

flooding exactly when the harvest was due, required diverting the excess spring water away from the fields into swamps at the far end of the cultivation zone. It was the difficulty in overcoming these two problems that delayed irrigation agriculture in Babylonia until long after rain-fed agriculture flourished in Upper Mesopotamia and irrigation systems were established in southwest Iran (see Wilkinson 2003:72-76). But once irrigation agriculture was introduced into Babylonia, it led to a rapid development of civilization. More than thirty major city-states have been identified in Babylonia in the fourth and third millennia; the largest of these cities was Uruk, after which the entire period is typically named. It was at Uruk that writing presumably originated at about 3200 BCE, when its population reached about 20,000 (Yoffee 2005:43).

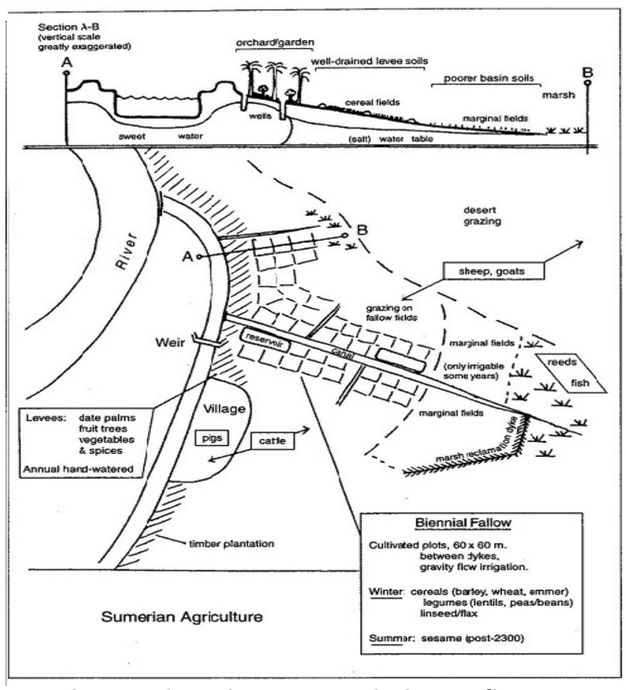

Figure 4 A schematic irrigated agricultural area in Babylonia; Source: Wilkinson (2003:92), following Postgate (1994:175).

Figure 4 presents a schematic reconstruction of an agricultural cell in Lower Mesopotamia (Postgate, 1994:175).[64] This depiction reveals the diversity of local farming conditions in Babylonia,

of land would have been worthless for cultivation purposes, without an attached water allocation. This serves to emphasize the extent of power of those who could deny water. The relative abundance of land may explain also the prevalence of the rotating fallow in Southern Mesopotamia. Wilkinson (2003:71, 80) alludes to the possibility that the subsequent extension of irrigation in northern Babylonia contributed to the ultimate decline of Sumer.

[64] As Postgate notes, the gradient in the cross sectional diagram is grossly exaggerated. According to Wilkinson (2003:77) the levees rise by only 2-3 meters above the flood basins. The square fields indicate that what is reconstructed here is northern Babylonia, since in Sumer the fields were very long, stretching all the way from the canal to the marshland (see below).

even within a compact zone. The individual land plots were irrigated by an intricate hierarchy of branching canals that diverted the water from the rivers. Date orchards and vegetables were grown on the raised levees of the canals, while cereals were cultivated on the levees' outer slopes. These fields varied widely in their quality, due mostly to the poor drainage of the lower fields that were affected by the saline water table from the adjacent marsh. The accumulation of salt in the topsoil as a result of irrigation was handled by the fallowing of all fields in every alternate year – just as in Upper Mesopotamia, but unlike Upper Egypt, where the water table was lower and where the salt in the topsoil was naturally flushed away. Another perennial difficulty was posed by the pastoralist inhabitants of marginal grazing land and the swamps, who threatened farmers' security.[65]

Several scholars have sought to explain why the "urban revolution" took place in Lower Mesopotamia, and even more specifically in Sumer. Liverani (2006) insists that the urban revolution was in fact transformational, and attributes it to a "secondary agricultural revolution" of locally administered irrigation agriculture (p. 28). More specifically, though, Liverani and Steinkeller (1999) place particular emphasis, not on the irrigation itself, but on subsequent technical innovations that consist of the cultivation of elongated fields with a deep plowing technique, aided by a seeding plow pulled by oxen.[66] The introduction of these farming techniques, they argue, was the source of economies of scale that contributed not only to higher productivity, but also to the centralization of farming.[67]

[65] Given the susceptibility of farmers to banditry, it is likely that the existence of pastoralists in their vicinity was a mixed blessing to farmers. It may have contributed through exchange and through peak-season work to farmers' welfare when their security was guaranteed, but also posed a potential threat to them. As Olson's (1993) proposed, this threat probably contributed to the farmers' subordination to the urban elite that purported to provide them with protection. This consideration may explain the phenomenon that Adams (1981:137-145) identified of the "implosion" (see Yoffee 1995:54) of the rural settlements in the late fourth millennium. The shift of the rural population into the walled cities may have resulted from weakening, rather than the strengthening, of the Sumerian city-states relative to the pastoralists. This hypothesis gains some credence in that the rural sector apparently recovered during the stronger reign of the Akkadian and Ur III dynasties (Steinkeller 2007).

[66] See Liverani (2006:14-19). The lesser emphasis on the introduction of irrigation may reflect Liverani's apparent aversion to Wittfogel's ideas, which he describes as "notorious" and "abnormal" (2006:15). Potts (1997:70-89) 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 fields by means of 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 front next to the 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 not only saved seed, but more importantly enabled the efficient utilization of scarce water during the cultivation season, since only the furrows had to be watered. The deep furrow implied also that salinity of the topsoil was concentrated mostly away from the seeds.

[67] 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, it seems to us that it was the introduction of irrigation farming, and its great

Several scholars have sought to explain why the "urban revolution" took place in Lower Mesopotamia, and even more specifically in Sumer. Liverani (2006) insists that the urban revolution was in fact transformational, and attributes it to a "secondary agricultural revolution" of locally administered irrigation agriculture (p. 28). More specifically, though, Liverani and Steinkeller (1999) place particular emphasis, not on the irrigation itself, but on subsequent technical innovations that consist of the cultivation of elongated fields with a deep plowing technique, aided by a seeding plow pulled by oxen.[68] The introduction of these farming techniques, they argue, was the source of 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, it seems to us that it was the introduction of irrigation farming, and its great transparency and amenability to control, that provides the key clue to the rapid rise of the early city-states in Sumer.[69]

During the seventh to the fifth millennia the Persian-Arab Gulf extended more than two hun-

---

transparency and amenability to control, that provides the key clue to the rapid rise of the early city-states in Sumer.

[68] See Liverani (2006:14-19). The lesser emphasis on the introduction of irrigation may reflect Liverani's apparent aversion to Wittfogel's ideas, which he describes as "notorious" and "abnormal" (2006:15). Potts (1997:70-89) 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 fields by means of 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 front next to the 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 not only saved seed, but more importantly enabled the efficient utilization of scarce water during the cultivation season, since only the furrows had to be watered. The deep furrow implied also that salinity of the topsoil was concentrated mostly away from the seeds.

[69] Algaze (2001, 2005, 2008) emphasizes the relatively greater productivity of agriculture in Sumer (in terms of grain yield per unit of cultivated land – even though according to the above, the constraining factor was water, rather than land). But, in effect, he downplays the significance of irrigation and agriculture by attributing the "takeoff" of the city-states in Sumer mostly to trade. Algaze (2001) notes that at its heyday in the late fourth millennium, the city-state of Uruk engaged in extensive long term trade with its trading colonies up the Euphrates, from which it imported scarce timber, metals and precious stones. Algaze (2005, 2008) emphasizes short-distance trade and attempts to apply to Sumer theories on the growth of pre-modern trading centers. He attributes this trade to the diversity and richness of the ecosystems within Sumer, related to the transgression of the gulf's littoral zone and to the relative ease of riverine transportation. Along with Adams (2005) and others, we are not swayed by his arguments. The long-distance imports of Uruk served the local elite and have taken place 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 the local elite; yet Algaze's arguments fail to explain the source of the elite's ability to extract surplus. The trading hypothesis fails also to explain the timing and location of the Uruk takeoff. Given that the gulf was extended already during the sixth and fifth millennia, if littoral resources were an essential contributing factor for that takeoff, why did it occur only in the fourth millennium, after the gulf started to recede? And what was then the comparative advantage of Uruk within Sumer, given that it was situated relatively further away from the wetlands?

dred 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 (1988:55-56), 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.[70] One may surmise that winter farming in Sumer was critically facilitated by the high level of the Euphrates' river bed there, relative to the nearby alluvial plain with its higher water table (see Figure 4). The combination of these factors meant that the outer slopes of the river's levees (and of the levees of abandoned river channels) could be cultivated in winter by using short diverting ditches with only minor earthwork.[71] This method of farming was highly productive, and, at the same time, made the Sumerian peasants utterly dependent on the rationed water. The water supply was managed by the elite that controlled the water direction at various canal junctures upstream. This control, in turn, enabled the elite to extract surplus from these peasants. At the same time, the operation of this complex irrigation system required skilled managers with "through knowledge of local conditions on a day-to-day basis" (Hunt 1987:172). This implied that, unlike the case of Egypt, the local managing elite were, in effect, irreplaceable. In terms of transparency, we interpret these two considerations as meaning that the farming activity was highly transparent to the local elite, but that the local elite was rather opaque to remote potential authority. The first of these two feature, we contend, explains why powerful early city-states were able to form in Sumer, why these cities were often run by a coalition of families, why these families in effect owned the land, and why independent owner-occupied farming was practically nonexistent in Sumer – in analogy to Upper Egypt, but unlike Upper Mesopotamia. The second of the above two features explains, in turn, why the local elite was extremely powerful and resilient, and why, as a result, repeated aggressive attempts to unify Babylonia under one of the rival city

[70] The presumption of increased aridity is often invoked. Some scholars maintain that the northward shift of the Monsoon rain belt during the mid-Holocene reached southern Sumer and then retracted; but, unlike the case of Upper Egypt, this is not easily verified.

[71] See Adams (1981:245) and Wilkinson (2003:89). This may explain not only the priority of the Euphrates over the Tigris, but also the priority of Sumer over Akkad. As Nissen notes (1988:144-145), "the Tigris cuts deeply into the land, while "the river [Euphrates] was especially low-lying in northern Babylonia," flowing there "through a relatively narrow plain." Given that the prime problem of irrigation in Lower Mesopotamia was that the crops had to be irrigated in October-January when the rivers' water was low, the elevated river bed of the Euphrates in Sumer enabled irrigation of the levee's outer slope during the winter months by means of short irrigation ditches. The cultivated area could then be protected from the spring floods by opening up alternative channels to divert the excess water directly towards the marshes in the alluvial plain. In contrast, irrigation from the Tigris, and from the Euphrates in upper and middle Mesopotamia, required long canals to capture the river water upstream, in order to channel the water to the cultivable areas in a gradient lower than the river's. These irrigation projects were thus more elaborate, and probably required organized public initiative, rather than simpler local entrepreneurship.

states ended up in failure – unlike the case of Egypt.

Indeed, scholars have long maintained the bulk of the agricultural land of the Sumerian city-states was owned, at least nominally, by each city's temple; and temple affairs where managed by dominant elite families. The actual farming in Sumer was conducted by peasant families of sharecroppers who were controlled by a hierarchy of managers, under the ultimate control of these dominant elite families. This system has been described as "a pyramid of individual families" (Steinkeller 1999: 293) or as an "institutional household" (Renger 1995).[72] In comparison, the tenure system in northern Babylonia (Akkad) was a mixture of the one that prevailed in Sumer with the one in Upper Mesopotamia, in which the temples did not play a dominant role in agriculture. The significant extent of private land ownership in northern Babylonia is attested by the records of numerous land transactions from the second and first millennia.[73]

The rival city states of Sumer coexisted, and periodically fought each other, for about a millennium before they were first consolidated by Sargon of Akkad at about 2350 BCE. But Sargon's central state lasted 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 was similarly dissolved. The next territorial state was established by Hammurabi of Babylon (at about 1800 BCE), but lasted only a short time. Thus, until the first millennium, Mesopotamia was ruled almost always by competing, independent city-states, with only brief intermittent periods of a central territorial state.

This brief survey of the history of the Mesopotamian states raises the obvious puzzle: how to account for the apparent non-viability of the central territorial states in Babylonia? Our model accounts for this phenomenon that sets ancient Mesopotamia so clearly apart from ancient Egypt. Unlike Egypt, where knowledge of the inundation magnitude provided the Pharaoh with control over all the districts along the Nile, there was no such lever of control in Babylonia. The construction, maintenance and intricate management of irrigation were conducted at the level of the city-state. This required highly specific and up-to-date knowledge that was held by the local elites of these cities and made them indispensible. Yoffee (2006:37) states that Sargon of Akkad was well aware of this intermediation problem when he ascended to power and thus founded a new capital city

[72] There is extensive literature on the extant of land ownership in Sumer by private ('non-institutional) people (see Renger 1995). But from our perspective the important observation is that these private lands were owned by elite city families and not by the actual cultivators.

[73] For the land tenure system in Northern Babylonia in the third and early second millennium see Renger (1995), Steinkeller (1999) and Goddeeris (2002).

of his own, in attempt "to disenfranchise the old landed aristocracy." But, after conquering the diverse city states in Babylonia, he ruled them by appointing "royal officials, who served alongside the traditional rulers of the conquered city-states;" even though this "uneasy sharing of power ... led to a power struggle" and to the demise of these territorial states (Yoffee 1995:292-293).

Our perspective clarifies that it was not due to oversight that the sporadic central rulers in Babylonia did not disenfranchise the local elites in the city-states that they conquered. One of the most revealing aspects of the continued power of these city-states was that even when they were subjected to a central territorial authority, they retained near-autonomy in conducting their internal affairs. The local elites in these cities participated in assemblies that elected local officials and decided lawsuits. In fact, this form of proto-democracy (first identified by Jacobsen 1943) started already in the early period of Sumer and continued, in effect, throughout the Assyrian, Babylonian and Persian empires in the first millennium.

The evidence adduced here conforms to the predictions of our simple model that reduces the exogenous environmental features to a single dimension of 'transparency.' This model thus helps to account for the complex and diverse institutions that prevailed in ancient Egypt, in Upper Mesopotamia and in Lower Mesopotamia. It places Egypt at one end of the spectrum, with the greatest local and global transparency; next to it is Sumer (southern Babylonia) with high local but not global transparency; at the other end of the spectrum is Upper Mesopotamia with limited local and global transparency and with Akkad (northern Babylonia) in between the last two regions.[74]

The striking differences between the three major regions of ancient civilizations can thus be accounted for within our model. Egypt's uniqueness, we argue, was due to the relatively high transparency of its farming to the central government. This explains: (a) why Egyptian peasant farmers were tenant-serfs, without title to the land that they cultivated, while much of the land in Upper Mesopotamia was in patrimonial ownership and was cultivated by its direct owners. (b) Why records for private real estate transaction and loan contracts (that would be typically secured by land) are abundant in ancient Mesopotamia and hardly exist in ancient Egypt (up to the mid-first millennium). (c) Why the local nobles and the regional governors in Egypt were in effect agents of the Pharaoh, subject to dismissal if found delinquent, and why the cities in Egypt were in effect mere administrative centers; whereas the cities in Mesopotamia retained all along much power, and were effectively autonomous, run by the resilient local elite. (d) Why legal disputes were resolved

[74] While Egypt may indeed be almost at the one extreme of the conceivable spectrum of transparency, Upper Mesopotamia was clearly not at the opposite end. Mediterranean economies that are based entirely on rain and on irregular terrain with idiosyncratic micro-climate (like that of ancient Israel) are clearly closer to that end of the spectrum, with the implication that one should expect there owner occupied farming and weaker states.

in Egypt (up to the mid first millennium) by local elders on the basis of common law and without resort to any legal codes imposed from above; while such disputes, as well as commercial disputes or disputes over inheritance (that hardly existed in Egypt) where resolved in Mesopotamia by due court process, guided by law codes that were issued by the central government. (e) Why the central state in Egypt rose so much more quickly than in Mesopotamia, why it was so much more stable, and why it could siphon off a major share of the countries produce – that enabled the erection of the great pyramids already in the mid-third millennium; (g) Why the Pharaohs from early on were considered as incarnations of the gods who regulated the cosmos, while even the mighty kings of Mesopotamia (with a single exception in early Akkad) were only considered as envoys of the gods (Baines and Yoffee 1998).

5 Concluding thoughts

In the spirit of Hicks (1969) and North (1981), we sought to apply here economic theory to the analysis of the early phases of state societies. Our overall theme is that the viability of the state and related social institutions are conditioned on the extractive technology; this technology, in turn, is a by-product of the transparency of the production technology. This theme is applied to two critical phases in the evolution of civilization. First, we sought to reexamine the causal mechanism that ties the Neolithic Revolution to the emergence of hierarchical society. We argue that the transition away from near-egalitarian, ostensibly leaderless foraging bands to stratified farming societies occurred primarily because the production of cereals (and the assembly of herds) required storage. These stores of food were easy targets for roving bandits and forced the early farmers to subject themselves to hierarchical leaders, whether from within or from outside. We employ Malthusian considerations to challenge the role of the two dominant alternative explanations that ascribe the emergence of ranked society to the greater productivity of agriculture, working itself through either the creation of "surplus" or to "population pressure."

We further apply our overarching theme to explore the source of the fundamental institutional differences between the early civilizations of Mesopotamia and Egypt. We interpret the transparency of farming technology in this historical context as determined principally by the direct control of water supply or knowledge thereof. In particular, we argue that the rapid rise of the powerful central state and its subsequent stability in Egypt, the weakness of its cities, the lack of land-owning peasantry and the lack of written law codes, reflect the fact that Egyptian farming was highly transparent, both at the local level and at the state-wide level. We argue further that the same paradigm is applicable to explain also the key differences between Upper Mesopotamia, where transparency was low and Lower Mesopotamia, where it was high at the local level of the city-states, but much less so at the level of the central state.

We would like to conclude by placing our contribution in a wider perspective: first in reference to the prevailing paradigm on the economies of antiquity, and then in reference to more recent history. Among anthropologists, archaeologists and economic historian of antiquity it has become by and large a truism that the ancient world was fundamentally different from the modern world, and that economic theory is inapplicable to the study of antiquity. In these disciplines it has become standard since the 1950's to replace the ideas of Adam Smith with the ideas and terminology of Karl Polanyi (1944). In a nutshell, Polanyi posited: (a) that the basic tenet of economists about exchange in impersonal and self-regulating markets is inapplicable to the ancient world; (b) that the social exchange that prevailed in pre-modern societies was based either on gift reciprocity (without immediate quid-pro-quo), or, at a more complex stage, on "redistribution." By "redistribution" Polanyi meant that much of the agricultural produce flows from the periphery to the center, and is then partly redistributed back to the periphery, even if a major part is skimmed at the top.[75] Both component of Polanyi's paradigm find ethnographic support in small tribal societies or in chiefdoms (see Johnson and Earle 2000). However, Polanyi's paradigm was adopted, even if somewhat adapted, by many if not most of the leading historians of early civilizations. Thus, one often encounters the characterization of ancient Egypt (see Kemp 2006), Mesopotamia (see Renger 2007, Liverani 2005) and Greece (see Finley 1999) as essentially "redistributive economies."

Some historians of antiquity, however, challenge the applicability of Polanyi's paradigm to these early states (see Nakassis, Parkinson and Galaty 2011). This challenge was partially reflected already in earlier controversies over the existence or inexistence of "private" land in these ancient societies (see Renger 1995, Steinkeller 1999 and Goddeeris 2002, 2007). By now it seems, though, that most Assyriologists acknowledge the existence of markets and market-determined prices in ancient Mesopotamia.[76] Van de Mieroop (1999, 2004) places this acknowledgment in perspective

[75] Polanyi ([1944] 1957:49-55) romantically envisioned a system of redistributive social exchange as prevailing among early hunters ("The members of a hunting tribe usually deliver the game to the headman for redistribution." p. 49); among the South Pacific islanders ("A substantial part of all produce of the island is delivered by the village headmen to the chief who keeps it in storage. . . ." P. 48); in ancient Egypt ("Redistribution was present on a gigantic scale in the civilization of the pyramids. [. . .] The household of the patriarchal family was reproduced here on an enormously enlarged scale, while its 'communistic' distribution was graded, involving sharply differentiated rations." pp. 30, 31), and all the way to pre-modern Western societies.

[76] Prominent is Algaze's (2008) insistence on the centrality of trade for the emergence of the Sumerian civilization.

by downplaying 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" (1999:117-118). With regard to the second component of Polanyi's paradigm, Van de Mieroop states unreservedly: "Polanyi's most lasting contribution lies in the demonstration of alternative means of exchange, reciprocity and redistribution. The existence of the latter in the Mesopotamian economy is obvious: the third-millennium public institutions of temples and palaces supported numerous families through a system of rations ... " (1999:118).

It seems that most scholars of antiquity generally adhere to this, affirmative, more central, component of Polanyi's paradigm. A somewhat similar evaluation of Polanyi was expressed also by North (1977:706): "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 contended that "Karl Polanyi's challenge must be met head on if economic history is to provide us with improved insights about our ancient past" (p. 707), and suggested (p.709) that "transaction cost analysis is a promising analytical framework to explore non-market forms of economic Organization."

Our foray into this critical debate addresses this central second component of Polanyi's paradigm, and in a sense attempts to meet the agenda that was set by North (1977). We apply economic theory to the analysis of non-market relations in antiquity, without ever referring to markets. The theory on asymmetry of information and moral hazard that we apply assumes a central place in the micro-economic theory of recent decades. But this theory of non-market, hierarchical relations featured prominently already in Adam Smith's analysis of land tenure institutions. We show here, in effect, that Polanyi's celebrated claim that the economy is "embedded" in social institutions (Polanyi 1944:57, Dalton 1968:xliv, 74, 148) does not have to mean that social institutions ought to be taken as exogenous. In particular, economic theory can explain some of the most salient institutional features of ancient societies and account for the fundamental social differences between ancient Egypt and Upper and Lower Mesopotamia.[77]

[77] Beyond exposing the dated and narrow view of economic theory that prevails among the scholars who adopt Polanyi's views about the inapplicability of economics to antiquity, our address of the nature of ancient social institutions casts doubt on the essential fruitfulness of Polanyi's anti-economics approach. We find the terminology of Polanyi's "redistribution" paradigm similarly misleading. The above quote from Van de Mieroop illustrates the ambiguous use of the term 'redistribution.' Scholars who apply this term to the hierarchical economies of ancient Egypt and Mesopotamia typically do not provide evidence for any redistribution to the peasantry; in accordance with Polanyi's grand vision of "the patriarchal family ... on an enormously enlarged scale" (see above). Rather, they rely on evidence for the prevalence of a system of "rations" that were in effect remittances to state functionaries or wages-in-kind to various workers and craftsmen employed by the state or by the temples. Similarly, when these scholars refer to how the ancient state obtained its revenue, they often resort to terms like "tribute" that connotes a voluntary

In conclusion, we would like to shift the scope on the ancient world and to express our belief that our overarching theme is applicable also to more recent phases of human history.[78] It seems particularly worthwhile to compare the mechanism by which the Neolithic Revolution impacted the formation of the early state to how the Industrial Revolution affected the modern state. The integrationist and conflict arguments that anthropologists utilize to explain the emergence of the state are analogous to the main theories that have been employed to explain the modern growth of government (Mueller 1987). Mainstream public finance literature (following Wagner [1883] 1967) focuses on the demand for public goods (the integrationist approach) that is claimed to rise more than proportionately when per capita income increases. The political economy literature (Meltzer and Richard 1981) emphasizes the redistributive nature of government spending, arguing that dominant interest groups manage to increase the scale of government to serve their own interests (presumably redistributing towards the dominant poor, in a reverse direction to that of the early conflict theory).[79]

The Malthusian consideration that we employ to challenge the applicability of these two mechanisms to the early state are clearly inapplicable to the modern state, since the modern increase in the scale of government happened to coincide with the escape from the Malthusian trap. Nevertheless, the comparison of the two Revolutions leads us to challenge the applicability of these two explanations also with regard to the modern age. The appetite of autocratic governments, from the Pharaohs of ancient Egypt to the Bourbons and the Romanovs, could not have been smaller than that of the modern democratic government. If one presumes that an absolutist government (Le-

contribution and avoid almost invariably the simple term 'tax.' Steinkeller (1987:28) examined the extensive system of transfers between the Ur III state and its provinces, known as bala, and claimed in Polanyi's spirit: "the bala institution functioned as a central redistribution system, integrating all the provinces into one interdependent whole." In reexamining the terminology and the evidence on the bala system, Sharlach (2004) argued: "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 concluded by stating: "there is no evidence known to me to imply the provinces sent their regional specialties to the crown for centralized redistribution," (p. 160); "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).

[78] The relatively stable share of government since the emergence of the ancient civilizations until the Industrial Revolution, in spite of continued technological improvements, should be recognized as key historical puzzle. We conjecture that these technological changes (like the technological improvements prior to the Neolithic Revolution) did not alter substantially the nature of the induced tax technology.

[79] Both of these explanations are premised on some form of democratic government that is attentive to its subjects. This reveals a striking difference between these two technological revolutions. As noted above, the Neolithic Revolution was accompanied by a transition from egalitarianism to stratification and to absolutist governments. The Industrial Revolution, in contrast, was accompanied by the end of absolutism and a transition to democratization. It also led (with some delay) to significant leveling of wealth and income disparities.

viathan) aims at maximizing its long-term net tax revenue, it follows that the shift from autocracy to democracy in itself could not have raised the scale of government any further. We deduce from this simple logic that the sharp rise of the modern state should not be attributed to an increase in the relative craving for government spending. Rather, it ought to be explained, once again, by the transformation of the tax technology, along similar lines to those that we employ to examine the evolution of early state and the subsequent differences between to two most ancient civilizations. In tune with Parkinson's (1960:3) 'Second Law' that "[State] expenditure rise to meet income," with Olson's non-teleological theory of the state, and with Kau and Rubin (1981), we thus posit that the modern growth of the state should be ascribed to a fundamental 'improvement' in the tax technology that was induced by the Industrial Revolution. The shift to mass production by hired labor entailed a massive accounting paper trail. This increased the state's ability to tax; in part by making private companies into tax collection agents, and in part by introducing the income tax.[80]

Our analysis may prove relevant also to future political developments. The prevailing thought among economists is that asymmetry of information is essentially a hindrance for efficiency. Our proposed framework reveals another side to this coin. From agents' point of view, the lack of transparency may protect their freedom and possibly also their well-being.

Appendix A

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, $\theta = G$, since otherwise it is inefficient to exert effort. The principal's objective function remains unchanged and the incentive constraint becomes:

$$b + \delta V + \omega - (m + \gamma) \geq ((1 - q) + q(1 - d))\delta V + \omega - m.$$

This implies that the principal sets the bonus so that $b = \gamma - qd\delta V$. Thus, once more, the objective function is linear in d, and admits corner solutions, with a 'pure carrot' contract when q is below some threshold and a 'maximum stick' contract above that threshold. When q is low and

[80] See also Kleven, Kreiner and Saez (2009). Peacock and Wiseman (1961) suggest a ratchet model whereby a temporary major fiscal need, like financing war, creates a long-term 'improvement' in the tax technology that leads to increased public spending. Their theory is reminiscent of the interpretations by Schumpeter ([1918], 1991), Tilly (1990) and Gennaioli and Voth (2010) of how the state's capacity to tax evolved since the middle ages in response to the exigency of wars. Besley and Persson (2009, 2010) have studied how investment in legal infrastructure affects the state's fiscal capacity.

the contract is pure carrot, we obtain: $b_c = \gamma$; $V_c = p\gamma/(1-\delta)$, and when q is above the threshold, $b_s = \gamma - pq\delta\gamma/[(1-\delta(p+q-2pq)]$; $V_s = p\gamma/[(1-\delta(p+q-2pq)]$. All these variables are lower than the respective ones under the assumption in the model presented in the main text. The threshold transparency level is such that the principal is indifferent at that level between the two contracts, satisfying: $\hat{q}/(1-\hat{q}) = (1-p)x/p\delta V_s$.

Appendix B

In this Appendix we consider alternative modeling of the basic moral hazard problem, where the agent has the option to hide output and effort is costless. In particular, the agent can report that output was low even when it is high. The principal provides the agent with a bonus b if reported output is high, and dismisses the agent with probability d if the reported output is low and the signal indicates that the state of nature is good. The basic wage here covers subsistence: $\omega = m$.

The principal's objective function is:

$$\pi = \max_{b,s} p(H-b) + (1-p)(L-(1-q)dx) - m$$

The incentive for the agent to accurately report that output is high:

$$b + \delta V \geq (H-L) + [(q(1-d) + (1-q)]\delta V$$

A binding incentive compatibility constraint thus implies: $b = (H-L) - qd\delta V$. Substituting in the objective function:

$$\pi = \max_{d} x(L-m) + pqd\delta V - (1-p)(1-q)dx.$$

This is linear in s, and generates once again the two types of contracts, where the threshold transparency level satisfies:

$$\frac{\hat{q}}{1-\hat{q}} = \frac{(1-p)x}{p\delta V_s}.$$

A pure carrot contract has: $d_c = 0, b_c = H-L$, and $V_c = p(H-L)/(1-\delta)$. A maximum stick contract, for q above the threshold, satisfies: $d_s = 1$ and $b_s = (H-L) - q\delta V_s$, with the value function:

$$V_s = \frac{p(H-L)}{1-\delta(p+q-2qp)}.$$

These results confirm that the analysis of the main model is qualitatively robust to alternative scenarios of the moral hazard problem.

Appendix C

The aim of this appendix is to extend our model to two tiers. We elect to simplify things by extending the model from Appendix B, with hiding output instead of effort. For that purpose we attach a subscript 1 to all the variables in the preceding appendix, except for the probability p. Denote by $\pi_1(q,\theta)$ the expected revenue to the lower level, intermediate principal, conditional on the true state of nature and on the degree of transparency. This function is of the following form:

if $q < \hat{q}_1$,

$$\pi_1(q_1, B) = \pi_1(q_1, G) = L_1 - m_1,$$

and for $q > \hat{q}_1$:

$$\pi_1(q_1, B) = L_1 - m_1 - (1-q_1)x_1; \qquad \pi_1(q_1, G) = L_1 - m_1 + \frac{pq_1\delta_1(H_1 - L_1)}{1 - \delta_1(p + q_1 - 2pq_1)}$$

Now, we denote by q_2 the degree of transparency of the state of nature to the higher level principal. The case of interest is where $q_2 \leq q_1$ and where $q_1 > \hat{q}_1$; that is, when the transparency declines when one climbs up the hierarchy, and when it is high enough at the lower level to employ a 'maximum stick' type contract. Define now: $L_2 = \pi_1(q_1, B); H_2 = \pi_1(q_1, G)$. From the point of view of the higher tiered principal, the intermediary receives L_2 when the state of nature is bad, $\theta = B$ (with probability $1-p$) and receives H_2 when $\theta = G$ (with probability p). Once again we assume that the agent (=intermediary) may underreport his revenue. And, as in the basic model, the higher principal receives a signal for the state of nature, whose accuracy is q_2. She then uses an analogous two-edged incentive scheme: a bonus b_2 and a threat to dismiss the intermediary with probability d_2, and at a cost of x_2, when the reported revenue is L_2 but the signal indicates a good state of nature. Let m_2 be the maintenance cost of the lower principal, and assume $\omega_2 = m_2$. The maximization for the higher principal is, just as in the preceding appendix:

$$\pi_2 = \underset{b_2, d_2}{max} p(H_2 - b_2) + (1-p)(L_2 - (1-q_2)d_2x_2) - m_2.$$

The incentive constraint to report revenue truthfully will also be analogous to the one above, and under the presumption that it is binding it requires: $b_2 = (H_2 - L_2) - q_2d_2\delta_2V_{s2}$. The resultant maximand is, once again, linear in the dismissal probability d_2, and hence admits corner solutions. Thus if q_2 is low enough, the intermediary will never be dismissed, and will be incentivized only by a bonus. On the other hand, when the transparency q_2 is high enough, the intermediary's tenure will not be secure, and he will be subject to dismissal.

References

[1] Acemoglu, Daron, Simon Johnson and James A. Robinson, (2002), "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 117, 1231-1294.

[2] Acemoglu, Daron, Simon Johnson and James A. Robinson (2005) "Institutions as the Fundamental Cause of Long-Run Growth," *Handbook of Economic Growth*, Edited by Philippe Aghion and Steven N. Durlauf, Elsevier, 386-472.

[3] Ackerberg, D. A. and M. 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.

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

[5] Adams, Robert McC. (1966) *The Evolution of Urban Society*, Aldine Atherton.

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

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

[8] Aiyar, S., C. Dalgaard and O. Moav (2008) "Technological Progress and Regress in Preindustrial Times," *Journal of Economic Growth*, 13, 125-144.

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

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

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

[12] Allen, D. and J. Lueck (1992) "Contract Choice in Modern Agriculture: Cash Rent versus Cropshare," *Journal of Law and Economics,* 35 (2) 397-426.

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

[14] Ames, Kenneth M. (2008) "The Archaeology of Rank," in *Handbook of Archaeological Theories*, edited by R. Alexander Bentley, Herbert D.G. Maschner and Christopher Chippindale, Altamira Press, 487-513.

[15] Ashraf, Quamrul and Oded Galor (2011) "Dynamics and Stagnation in the Malthusian Epoch," *American Economic Review*. 101, 2003–41.

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

[17] Baer, Gabriel (1966) "Land Tenure in Egypt and the Fertile Crescent, 1800-1950," in *The Economic History of the Middle-East 1800-1914*, edited by Charles Issawi, University of Chicago Press, 79-90.

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

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

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

[21] Bard, Kathryn A. (2008) "Royal Cities and Cult Centers, Administrative Towns, and Workmen's settlements in Ancient Egypt," in *The Ancient City: New Perspectives on Urbanism in the Old and New World*, edited by Joyce Marcus and Jeremy A. Sabloff, School of Advanced Research, 165-182.

[22] Bellwood, Peter (2006) *First Farmers: The Origins of Agricultural Societies,* Blackwell Publishing.

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

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

[25] Boehm, Christopher (1999) *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*, Harvard University Press.

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

[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, R. (1970) "*A Theory of the Origin of the State*," Science 169, 733–738.

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

[30] Childe, V. Gordon ([1942], 1952) *What Happened in History*, Penguin Books.

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

[32] Cohen, Ronald and Elman R. Service, editors (1978) *Origins of the State: The Anthropology of Political Evolution,* Institute for the Study of Human Issues.

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

[34] Dalton, George, editor (1968) *Primitive, Archaic and Modern Economies: Essays of Karl Polanyi*, Beacon Press.

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

[36] Doepke M. and R. Townsend (2006) "Dynamic Mechanisms Design with Hidden Income and Hidden Actions," *Journal of Economic Theory*, 106, 235-285.

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

[38] Easterly, William and Ross Levine (2003) "Tropics, Germs, and Crops: How Endowments Influence Economic Development," *Journal of Monetary Economics*, 50, 3–39.

[39] Engerman, Stanley L. and Kenneth L. Sokoloff (1997), "Factor Endowment, Institutions, and Differential Paths of Growth Among New World Economies" in *How Latin America Fell Behind*, edited by Stephen Haber, Stanford University Press.

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

[41] 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.

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

[43] Fenske, James (2011) "Ecology, Trade and State in Pre-Colonial Africa," MPRA working paper.

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

[45] Galor, Oded (2005) "From Stagnation to Growth: Unified Growth Theory", *Handbook of Economic Growth,* Edited by Aghion, P. and Durlauf, S. Elsevier.

[46] Galor, Oded, Omer Moav and Dietrich Vollrath (2009) "Land Inequality and the Emergence of Human Capital Promoting Institutions" *Review of Economic Studies*, 76, 143–179.

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

[48] Gibbons R. and K.J. Murphy (1992) "Optimal incentive contracts in the presence of career concerns: Theory and evidence," *Journal of Political Economy*, 100, 468-505.

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

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

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

[52] Goddeeris, Anne (2007) "The Old Babylonian Economy," in *The Babylonian World*, edited by Gwendolyn Leick, Routledge, 198-209.

[53] Haas, Jonathan (1982) *The Evolution of the Prehistoric State*, Columbia University Press.

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

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

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

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

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

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

[60] Jacobsen, Thorkild (1943) "Primitive Democracy in Ancient Mesopotamia," *Journal of Near Eastern Studies*, 2, 159-172.

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

[62] Johnson, A. W. and T. Earle (2000) *The Evolution of Human Societies From Foraging Group to Agrarian State*, Stanford University Press.

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

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

[65] 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.

[66] Kuijt Ian (2009) "What do We Really Know about Food Storage, Surplus, and Feasting in Preagricultural Communities?" *Current Anthropology,* 50, 641-644.

[67] Kuijt, Ian and Bill Finlayson (2009) Evidence for Food Storage and Predomestication Granaries 11,000 Years Ago in the Jordan valley," *Proceedings of the National Academy of Science*, 106, 10966-10970.

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

[69] Liverani, Mario (2005) "the Near East: The Bronze Age," in *The Ancient Economy: Evidence and Models*, edited by J.G. Manning and Ian Morris, Stanford University Press, 47-57.

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

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

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

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

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

[75] Meek, Ronal L. (1976) *Social Science and the Ignoble Savage*, Cambridge University Press.

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

[77] Moav, O. and Z. Neeman (2010) “The Quality of Information and Incentives for Effort,” *The Journal of Industrial Economics,* 58, 642-660.

[78] Mueller, D.C., (1987) “The Growth of Government; A public Choice Perspective,” *IMF Staff Papers,* 34, 115-149.

[79] Nakassis, Dimitri, William A. Parkinson and Michael L. Galaty (2001) “Redistribution Economies from Theoretical and Cross-Cultural; Perspective,” *American Journal of Archaeology*, 115, 177-184.

[80] 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.

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

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

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

[84] Parente, S., and E.C. Prescott (2000) *Barriers to Riches*, MIT Press.

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

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

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

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

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

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

[91] Price, T. Douglas and Gary Feinman, editors (2010) *Pathways to Power: New Perspectives on the Emergence of Social Inequality*, Springer.

[92] Renger, Johannes M. (1995-1996) “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.

[93] Renger, Johannes (2007) "Economy of Ancient Mesopotamia: A General Outline," in *The Babylonian World*, edited by Gwendolyn Leick, Routledge, 187-197.

[94] Richerson, Peter J., Robert Boyd and Robert L. Bettinger (2001) "Was Agriculture Impossible During the Pleistocene but Mandatory During the Holocene? A Climate Change Hypothesis," *American Antiquity*, 66, 387-411.

[95] Sachs, J. and A.M. Warner (1995) "Natural Resource Abundance and Economic Growth," *NBER* Working Paper 5398.

[96] Schumpeter, Joseph A. ([1918], 1991) "The Crisis of the Tax State," in *The Economics and Sociology of Capitalism*, edited by Richard Swedberg, Princeton University Press, 99-140.

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

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

[99] Silver, Morris (1995) *Economic Structures of Antiquity*, Greenwood Press.

[100] Slemrod, Joel B. and Shlomo Yitzhaki, (2002) "Tax Avoidance, Evasion, and Administration" in *Handbook of Public Economics*, Volume 3, edited by A. Auerbach, A. and M. Feldstein, Elsevier.

[101] Smith, Adam (1976) *An Inquiry into the Nature and Causes of the Wealth of Nations*, edited by R.H. Campbell, A.S. Skinner and W.B. Todd, Oxford University Press.

[102] Smith, Adam (1978) *Lectures on Jurisprudence*, edited by R.L. Meek, D.D. Raphael and P.G. Stein, Oxford University Press.

[103] Steckel, Richard H. and Jerome C. Rose, editors, (2002) *The Backbone of History: Health and Nutrition in the Western Hemisphere*, Cambridge University Press.

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

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

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

[107] Testart, Alain (1982) "The Significance of Food Storage among Hunter-Gatherers: Residence patterns, Population densities, and Social Inequalities," *Current Anthropology*, 23, 523-537.

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

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

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

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

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

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

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

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

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

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

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

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

[120] Willcocks, W. (1899) *Egyptian Irrigation*, Second edition, London.

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

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

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

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

[125] Woodburn, James (1982) “Egalitarian Societies,” *Man*, 17, 43-51.

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

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

[128] Zaccagnini, Carlo (1984) “Land Tenure and Transfer of Land at Nuzi (XV –XIV Century B.C.)” in *Land Tenure and Social Transformation in the Middle East*, edited by Tarif Khalidi , American University of Beirut, 79-94.

[129] 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 *Urbanization and Land Ownership in the Ancient Near East*, edited by Michael Hudson and Baruch A. Levine, Peabody Museum, Harvard University, 331-352.