Geography, Transparency and Institutions*

Joram Mayshar† Omer Moav‡ Zvika Neeman§

August 14, 2014

Abstract

We propose a theory by which geographic variations explain cross-regional institutional differences in: (1) the scale of the state, (2) the distribution of power in the state hierarchy, and (3) farmers' property rights over land. The mechanism underlying our theory is based on the effect of geography on transparency of farming, which in turn determines the state's extractive capacity. We apply our theory to explain differences in the institutions of Egypt, Southern Mesopotamia and Northern Mesopotamia in antiquity.

KEYWORDS: *Geography, Transparency, Institutions, Land Tenure, State Capacity, State Concentration*

JEL CLASSIFICATION NUMBERS: *D02, D82, H10, O43*

*We would like to thank the Editor and four anonymous referees for excellent comments and guidance on the revision of the manuscript. We have also benefited from comments from Daron Acemoglu, Bob Allen, Josh Angrist, Eddie Dekel, Diana Egerton-Warburton, Christopher Eyre, James Fenske, Oded Galor, Maitreesh Ghatack, Jeremy Greenwood, James Malcomson, Andrea Matranga, Jacob Metzer, Stelios Michalopoulos, Motty Perry, Torsten Persson, Herakles Polemarchakis, Louis Putterman, Debraj Ray, Ariel Rubinstein, Yona Rubinstein, Larry Samuelson, Matthew Spigelman, Yannay Spitzer, Nathan Sussman, Juuso Valimaki, Joachim Voth, David Weil, and from comments from participants in various seminars and conferences.

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

‡Department of Economics University of Warwick, School of Economics Interdisciplinary Center (IDC) Herzliya, CAGE 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 protection of property rights assumes a paramount role in recent theories on the success of nations (North 1981). In that vein, Acemoglu and Robinson (2012) argue that extractive institutions that compromise property rights are the most detrimental factor for economic prosperity. Besley and Persson (2009, 2010) hold a seemingly opposite view – that the state's capacity to tax is a precondition for the state's ability to supply public goods and thereby to promote economic growth. But both theories share the implicit belief that extracting resources by the state can only be beneficial if the taxes are not excessive, and if the revenue is used to provide public goods rather than for consumption by the elite.

In this paper we offer a different perspective. In ancient times, while states supplied the basic public good of defense, there is little evidence that they were structured for anything other than to maximize the extraction of resources from the agricultural hinterland for consumption by the elite. As such, these ancient states appear to have lacked the prerequisite conditions for success according to both of the above-mentioned theories. Yet, Ancient Egypt, for example, had a prosperous civilization that built the great pyramids and was stable over several millennia – in spite of its extractive government and in spite of the absence of land property rights for its peasant farmers.

Rather than challenge the theories on the institutional preconditions for modern prosperity, we offer an explanation for the emergence of different institutions in early states, where land was the predominant capital asset. Our basic argument is that what distinguished between the nations of antiquity was the state's ability to appropriate revenue from agriculture. This ability was facilitated primarily by geographical and technological conditions that have been overlooked by recent studies on extractive institutions and state capacity. We suggest that geographic differences in the ability to appropriate revenue led to salient institutional differences across regions. The specific phenomena that we seek to explain are: the scale of the state; state concentration (the relative power of the center verses the periphery), and land tenure regimes.

In particular, we attribute the power and resilience of Ancient Egypt's central government, the relative weakness of its regional centers, as well as the lack of title to land by the peasantry, to the fact that its farming activity was highly transparent and thus appropriable. Although we focus on early state societies, we posit that our theory is applicable also to the modern expansion of the state, in that an increase in the transparency of production can explain the recent increase in state capacity to tax (see section 4). Furthermore, since our theory offers an explanation for the emergence of different institutions across regions, we believe that it is also relevant for understanding

deep rooted factors that play a key role in recent studies on comparative development.[1]

We employ a conventional principal-agent model, focusing on the implications of variation in the extent of informational asymmetry.[2] In the basic model, the agent represents a tenant farmer and the principal is an absentee land-owner representing government. The incentive scheme that the principal can use consists of a 'carrot' in the form of a bonus payment to the agent upon delivering high output, and a 'stick' in the form of possible dismissal as punishment for suspected shirking. To this end, we embed the model in a multi-period setting. Dismissal is painful for the agent who is presumed to be no longer employable as a farming tenant and is thus forced to relocate to the urban sector, where he enjoys no rents. Dismissal is also costly for the principal because she will have to replace an experienced agent with an inexperienced one and possibly also forgo the output retained by the dismissed tenant. Our main exogenous variable, representing the degree of transparency of farming, is the accuracy of a signal that the principal observes regarding the state of nature, from which she may infer (with some error) whether the agent worked diligently or not.

The model's results are that the more accurate the signal, the smaller is the role of the carrot, the larger the role of the stick, and the larger is the state's revenue. Our main contribution stems from our interpretation of these results: greater transparency induces a form of servitude as the tenant is evicted upon suspected shirking. On the other hand, opacity results in the state allowing the agent to retain a larger share of the output, without threat of dismissal. Thus, low transparency secures farmers' property rights over the land that they cultivate. The farmers in this case own the land de facto, even though the extractive state has absolute power to expropriate both their output and their land.[3]

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

[2] In employing a formal game theoretic model for explaining historical institutions, we follow the lead of Greif (1993, 2006).

[3] In our model, the principal is assumed to observe output but not the state of nature or the agent's effort. In online Appendix A we present an alternative framework that delivers similar qualitative results, in which the principal does not observe output and the moral hazard problem pertains to hiding (or misreporting) output by the agent. In online Appendix B we examine an alternative modeling strategy to demonstrate that when the principal can elect costly monitoring to obtain a signal on the agent's effort, the principal will choose to monitor and to punish the agent upon suspected shirking, only if the accuracy of the signal is sufficiently high and the cost of monitoring sufficiently low. Thus, as in the main model, opacity leads to property rights, whereas transparency of effort at a low cost leads to a form of servitude. We note that the existing literature on endogenous slavery typically focuses on labor shortage (Domar, 1970, Lagerlöf, 2009), but Dari-Mattiacci (2013) provides a recent exception that focuses on information asymmetry as we do.

Thus, consistent with North (1981), who offers an explanation for the development of western societies since the Middle Ages, property rights do not rise spontaneously in our framework, but are rather granted by an authoritarian government that seeks solely to maximize its revenue, with no direct concern for achieving efficiency.[4] In North's framework, securing the property rights of the non-elite serves the state as a commitment device, in order to overcome the hold-up problem of ex-post expropriation and thus to incentivize private investment by the non-elite. However, the issue of irreversible private capital investment is probably much less significant in agricultural societies, where land is the principal resource. In the context of ancient agricultural societies, we argue that it is the extent of asymmetric information that plays a key role in explaining the granting of property rights. Thus, when transparency is high enough, it is the threat of dismissal – an evident indication of the *lack* of full property rights – that serves to incentivize the agent. It is only in the case of sufficient opacity – when the cost of erroneous dismissal and replacement of the agent outweighs the benefits in incentives – that the absolutist state gives up the threat of dismissal and grants de facto property rights.[5]

In a two-layered extension of our basic model, designed to explain differences in state centralization, we examine the role of differential transparency at different levels of governmental hierarchy. We show that when farming activity is sufficiently transparent, not only locally to the intermediary (governor) but also globally to the upper level of the hierarchy (king), the intermediary retains a smaller share of the revenue and is subject to dismissal. On the other hand, if farming activity is sufficiently opaque to the king, the governor retains autonomy and a larger share of revenue. We maintain that it was the ability of the central authority to extract revenue from the subordinated lords that was the key for the success of early central states, and of Egypt in particular.[6]

[4] Demsetz (1967) explains the emergence of private property rights (out of a state of open access) as resulting from some form of an invisible hand, or communal agreement to resolve the inefficiency due to externalities.

[5] Besley and Ghatak (2009) survey the literature on the protection of property rights. They do not evaluate the possibility that an all-powerful authority may commit under some circumstances of opacity not to dismiss agents, and in effect grant agents with property rights to land, but avoid such a commitment under sufficient transparency. Dow and Reed (2013) propose that after the adoption of agriculture, coalitions of individuals who gained control over fertile land managed to exclude outsiders from sharing land's rent, thereby obtaining property rights. Focusing on ancient Egypt, Allen (1997) suggests that the lack of property rights was due to the relative abundance of land. We note, however, that absence of owner-occupied farming persisted in Egypt until the nineteenth century (G. Baer 1962), when there was no scarcity of labor.

[6] According to Ma (2011), the long-term success of Imperial China was similarly due to its ability to restrain the power of local officials. This was accomplished by the replacement of a hereditary feudal system with one based on rotating meritocratic bureaucracy. The effective denial of tenure to provincial bureaucrats served to overcome the local informational advantage that would otherwise enable them to gain independent power. Thus, whereas in ancient Egypt, the lack of informational advantage to provincial officials was essentially exogenous, due to the signals available directly to the Pharaoh, the denial of informational advantage to local Chinese officials was by design,

In section 3 we apply our theoretical insights to three regions of the ancient near east: Northern (upper) Mesopotamia, Southern Mesopotamia (Babylonia), and Egypt. We argue that distinctive geographical and climatic conditions, leading to different systems of irrigation and different patterns of transparency, can account for the different institutions that persisted for several millennia in these regions. In particular, farming in hilly Northern Mesopotamia was mainly rain fed, whereas in both Southern Mesopotamia and in Egypt farming relied entirely on riverine irrigation. Differences in seasonal patterns and in the terrain resulted also in major differences between the latter two regions. The Nile receives its water mainly from the early-summer monsoon rains in eastern Africa. As a result its flow peaks in summer, when it floods the narrow river valley, allowing for an irrigation system in which flood water was retained in pool-like basins, soaking the fields and depositing nutrients. The water was subsequently drained back to the Nile, in time for the sowing season of the staple cereals. The comparatively homogenous fields and the critical observability of the Nile's peak flood level revealed the 'state of nature' faced by individual farmers throughout Egypt with high accuracy, turning Egypt into a highly transparent economy at both the local and the state levels.

The Tigris and Euphrates, on the other hand, are fed by the winter rains in the mountains of modern Turkey and Iran and by melting snow in the spring. Thus, in Southern Mesopotamia these rivers are in low water in October-December when the fields require irrigation, and swell in the late spring, in the harvest season. This seasonal pattern posed a major mismatch problem that prevented irrigation by flooding. An extended canals system was required in alluvial Southern Mesopotamia to direct water to the fields in the cultivating winter season, as well as to divert the rising water away from the fields in the spring. The elaborate canal system provided the local elite with control leverage and information on the local state of agriculture, but kept local farming conditions highly opaque to any distant central government due to its high local heterogeneity. On the other hand, rain dependent dry-farming in the highlands of Northern Mesopotamia created idiosyncratic farming conditions that we characterize as opaque, both locally and even more so from a distance.

We claim that these differential transparencies explain institutional differences between these civilizations of antiquity. The high transparency of Egyptian farming explains how the Pharaohs could amass power by siphoning off a lot of tax revenue with a relatively lean state bureaucracy.

through fundamental administrative innovations. Unlike the existing literature on governance and hierarchy (see Tirole, 1986, Melumad, Mookherjee and Reichelstein, 1995, and Garicano, 2000, among others), our multi-tier model focuses on the allocation of power within different tiers of the hierarchy.

It further explains why the positions of district governors and state bureaucrats in Egypt were revocable and non-hereditary (at least in the Old Kingdom), as well as the weakness of its provincial cities. At the same time, the differential transparencies can explain why only weak local states existed in Northern Mesopotamia, as well as owner-operated farming, while in Southern Mesopotamia peasants were - as a rule - tenants, land was owned by the local urban elite, and city-states flourished and retained autonomy even in periods when a central state arose. These differences, could explain, in turn, why urban civilization first flourished in Southern Mesopotamia, even though farming started in Northern Mesopotamia and was adopted in the south only several millennia later. Furthermore, even though farming was adopted in Egypt more than a millennium after Southern Mesopotamia, the greater transparency of farming in the Nile valley accounts for the much faster rise of a powerful central state there, and for its greater subsequent stability.[7]

Our theory sheds new light on the long debated coincidence between irrigated riverine environments and early central states. According to Wittfogel's (1957) influential "Hydraulic Theory", large-scale irrigation infrastructure was necessary to realize the agricultural potential in such environments, and strong centralized states were thus a prerequisite for development. Wittfogel's many critics pointed out, however, that the irrigation systems in ancient Egypt and Mesopotamia (as well as in China and Peru) were initiated as local ventures and constructed communally, prior to the emergence of a strong central state. Moreover, even after a central state emerged, the management of these irrigation systems was not central, but rather remained with the local elites. Due to the cogency of these arguments, Wittfogel's theory is now considered defunct. Our theory provides an alternative explanation for the correlation between riverine environments and ancient states that in a sense reverses the causality of Wittfogel's theory. We suggest that irrigation systems provide transparency and means of control, and thereby facilitate on-going revenue extraction that is essential for the viability of any state. Thus, it is not that a despotic state was required to operate irrigation systems, but rather that such systems facilitated state control.

This explanation is indirectly supported by Bentzen et al. (2012). Using geographical indicators for the potential yield of irrigation as an instrumental variable for actual irrigation, they show that irrigation-based societies have been less likely to adopt democracy. Additional evidence comes from the Moche valley in the arid northern coast of Peru. Billman (2002) reports that an early irrigation

[7]Further details and evidence are presented in section 3. We do not claim that our theory provides the only explanation for these differences between Egypt and Mesopotamia. Following Carneiro (1970), Allen (1997) attributes the success of the Pharaohs to another geographical feature that enabled the elite to extract a surplus from farmers. In particular, he argues that the deserts surrounding the Nile valley confined the population and inhibited the peasants from avoiding taxation via out-migration.

system in that valley in 400 BCE-800 CE created an opportunity for leaders "to control land and the flow of water" and enabled them "to finance the creation of centralized, hierarchical political organizations" – thus leading to the formation of an early territorial state.

The distinctive feature of our approach is a non-teleological understanding of pre-modern government, in the spirit of Olson (1993), which focuses on the available tax technology.[8] In our model the state exists already, and possesses a monopoly on the exercise of power. Nevertheless, our theory sheds light also on the emergence of states. In a related paper (Mayshar, Moav and Neeman, 2011), we argue that the transformative facet of the Neolithic Revolution that gave rise to social hierarchy was not the surplus created by the increase in productivity, as is typically assumed, but rather the induced radical change in the tax technology that enabled the elite to appropriate, and thus create surplus. Consistently with this claim, de la Sierra (2013) employs evidence from the mining regions of the Democratic Republic of Congo to show that a rise in the price of the metallic substance coltan – produced from relatively bulky and hence transparent ores – led to the cessation of rival armed groups in the coltan rich regions and to the monopolization of violence; whereas an increase in the price of gold, which is easier to conceal and is hence less transparent, did not. Somewhat similarly, Buonanno et al. (2012) show that the spread of the Sicilian mafia in the 19th century was correlated with an increase in the price of sulphur.

Beyond our present focus on transparency in agriculture, another aspect of food production that is a prerequisite for the emergence of hierarchy and state capacity, and is emphasized in our companion paper, is the requirement of storage. Consider, for instance, Diamond's (1997) attribution of the economic underdevelopment of New Guinea to its inability to adopt the productive agricultural innovations that benefited Asia and Europe. We suggest that the economic underdevelopment of New Guinea, and of similar tropical countries, can be explained by the limits that tropical agricultural farming places on the capacity to appropriate. New Guinea's main agricultural produce was based on roots and tubers whose non-seasonal nature and greater perishability upon harvest imply that storage was both inefficient and not strictly required. As a result, food output was less amenable for appropriation. In the temperate regions, in contrast, where cereals became the staple food adopted by early farmers, storage was not only feasible due to low-perishability, but also mandatory due to the crops' seasonality. Yet storage made farmers vulnerable to expropriation, both by robber bandits and by the state: in a single inspection, the expropriator could impound a large fraction of a farmer's annual output.

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

The principal-agent model that we adopt here focuses on the role of differential transparency. The model adapts a standard framework that has often been applied to sharecropping contracts in agriculture (Stiglitz 1974). Ackerberg and Botticini (2000, 2002) provide medieval evidence in support of various hypotheses concerning the advantages and disadvantages of share tenancy. One supported hypothesis is that sharecropping allows poor tenants who face borrowing constraints to be profitably employed in agriculture. This consideration justifies the assumption of limited liability of agents in our theory, as it does in standard tenancy models. In particular, due to borrowing constraints, remuneration of agents below subsistence cannot serve as an alternative 'stick' to incentivize agents. Our assumption that the 'stick' is in the form of a threat of eviction upon suspected shirking is consistent both with practice (Cheung, 1969; Banerjee et al., 2002) and with the literature on tenancy contracts (e.g. Banerjee and Ghatak, 2004).[9] Indeed, land reforms are often designed to limit evictions (Chattopadhyay, 1979), and security of tenure is typically one of the main components of tenancy laws (Hossain, 1982). Our related presumption, that the dismissal and replacement of an agent is costly to the principal, is partly based on the idea that agents acquire specific local knowledge which is lost upon dismissal. This consideration likely implies a positive correlation between transparency and dismissal costs: homogenous terrain and climatic conditions that offer high transparency are probably associated with lower dismissal costs. This correlation reinforces our results: both high transparency and low replacement costs reduce the cost of including the threat of eviction in the contract.

The details of tenancy arrangements in antiquity are not typically recoverable by archeological evidence. As a result, historians of antiquity often employ evidence from the more recent past. It is on such a basis that Eyre (1997, p. 378; 1999, pp. 51-52) maintains that in ancient Egypt the village community as a whole was responsible for paying taxes, with the village headman exercising tight control over village land. In particular, Eyre claims that farmers did not have secure tenure and that the village head or the estate manager could reassign fields as he saw fit, even if by custom the same fields were annually assigned to the same farmer, or to his heir. This description supports

[9] Dismissal is the key element in Shapiro and Stiglitz's (1984) "efficiency wages" theory, although their contribution doesn't examine how employment contracts may be affected by differences in transparency. Hammurabi's law code (ca 1750 BCE, see Roth, 1997) provides evidence that tenant eviction was practiced in the ancient world. One law states that if a housing tenant has paid the annual rent in advance, "but the owner of the house then orders the tenant to leave before the expiration of the full term of his lease, the owner of the house ... shall forfeit the silver that the tenant gave him." Another law stipulates: "If a man rents a field in tenancy but does not plant any grain ... he shall give to the owner of the field grain in accordance with his neighbor's yield." Both laws indicate that the legal relation between tenant and landlord was as between free individuals of equal legal standing. While the possibility of dismissing the farming tenant is not mentioned in the first case, the second case indicates the viability of this option.

our assumptions that the threat of dismissal (or relocation) of individual farmers, implying lack of peasant property rights, was indeed used as an incentive device.

Before turning to the specifics of our model, we summarize its contribution to the theory of tenancy and institutions. An important feature of our application of the standard principal-agent framework is its focus on the impact that the degree of informational asymmetry has on the optimal combination of stick and carrot. This focus delivers insights that have been previously overlooked in the context of the study of institutions. The main insights concern the effect of transparency on property rights, as well as on the autonomy of intermediaries in the hierarchy. Our result that the efficacy of punishment increases with transparency appears to be generally applicable, to the point of seeming obvious: a threat of punishment, even when it is costly to the principal, becomes a more effective incentive device as the correlation between the agent's undesirable behavior and the exercise of punishment increases.

In section 2 we present our model, in section 3 we discuss the application of our theory to ancient Egypt and Mesopotamia, section 4 offers a brief discussion of the relevance of our 'transparency theory' to the rise of the scale of the state during the recent century and a half, and section 4 concludes.

2 An agency model with differential transparency

Consider a state with a given area of arable land, which is divided into plots. Each plot is allocated to one risk neutral agent-tenant. We model the principal (the state) as an absentee landlord who designs a contract that maximizes her expected periodic income, which is given by the total output produced by all the agents, less any payments made to the agents and the costs of replacing dismissed agents.[10] Each agent decides how much effort to exert. His payoff is the payment received from the principal, less his cost of effort.

We first characterize the optimal contract between the principal and a single agent in a simple basic model. We then extend the basic model to a two-tiered hierarchy with multiple agents.[11]

[10] We thus follow Olson (1993) and model the state as engaged in the expropriation of its subjects, without explicit consideration of the use of the revenue for the provision of public goods. This abstraction is not crucial for our arguments. Besides, beyond the provision of security, which we take for granted because it also serves the interests of the state, it seems to us to be a rather reasonable simplification. Besley and Ghatak (2009, p. 4560) claim that starting at the 14th century CE, "Expropriations by government are a fact of historical experience"; Ma (2011) offers a similar perception of the state in imperial China.

[11] Throughout our analysis we assume that plot size and therefore also the size of the population are given. In online Appendix F, we generalize the model to include an endogenous population size and an endogenous plot size that is determined by the principal to maximize its income. The main qualitative results are unchanged.

2.1 The basic model

We consider a Principal-Agent model that has the following characteristics. Both the output that is produced by the agent and the agent's choice of effort can be either low or high: $Y \in \{L, H\}$, and $e \in \{l, h\}$, respectively. The state of nature is also binary and can be either good or bad: $\theta \in \{G, B\}$. The annual output is a function of the effort exerted by the agent and the state of nature. We assume that it is high if and only if both 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 agent chooses the level of effort before he learns the state of nature.[12] The ex-ante probability that the state of nature is good is denoted by: $p \in (0, 1)$. After the agent chooses the level of effort, both the agent and the principal observe a public signal about the state of nature: $\sigma \in \{\tilde{G}, \tilde{B}\}$. The accuracy of this signal, $q \in [1/2, 1]$ is such that:

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

The level of accuracy q represents the degree of transparency of production in our model. If $q = 1$ then the signal perfectly reveals the state of the world; if $q = 1/2$ then the signal is uninformative.

We denote the annual cost (in units of output) of providing for the agent (and his family) until the next harvest period by $m + \gamma$, where $m \geq 0$ is the cost of subsistence in case the agent exerts (costless) low effort, and $\gamma > 0$ is the annual cost of exerting high effort. We assume that output is in any case larger than the cost of providing for an agent who exerts high effort: $L \geq m + \gamma$.

The agent's only alternative mode of employment outside agriculture is work as a domestic servant. We normalize his utility in this case to zero. The agent's annual utility as a tenant farmer equals his expected income, denoted by I, less the cost of subsistence and effort. Thus, the agent's annual utility if he exerts high and low effort is given by $I - (m + \gamma)$ and $I - m$, respectively. We assume that the agent has no other sources of income or wealth, that he cannot save, and that he cannot borrow. We denote the agent's intertemporal discount factor by $\delta \in [0, 1)$.

The principal employs the following incentive scheme. If output is low, the principal pays the agent a basic wage ω. If output is high, then the principal pays the agent $\omega + b$, where $b \geq 0$ is a bonus payment. The basic wage ω has to sustain an agent who exerts effort until the next harvest: $\omega \geq m + \gamma$.

[12] In practice, both the agent's effort and the relevant state of nature for agriculture are vectors whose components are distributed over the agricultural seasons. In online Appendix C we show that if the agent learns the state of nature before exerting effort, then the payoff to the Principal is higher, and no qualitative changes are implied.

When output is high the principal retains the agent. The agent is also retained when output is low but the signal indicates that the state of nature is bad ($\sigma = \tilde{B}$). But if output is low and the signal indicates that the state of nature is good ($\sigma = \tilde{G}$), then the principal may dismiss the agent and replace him with another.

We denote the probability with which the agent is dismissed when output is low and the signal indicates that the state of nature was good by d. For simplicity, we assume that the principal employs a pure strategy, namely $d \in \{0, 1\}$.[13] If the agent is dismissed, then the principal incurs a fixed cost $x > 0$ that represents the cost of dismissal and the present value of lost output during the training of a new agent.

We refer to the bonus payment as a 'carrot' and to the possibility of dismissal as a 'stick.' Thus, the solution of the Principal-Agent problem described here strikes an optimal balance between the use of a carrot and a stick as incentive devices. The fact that the principal is restricted to set d equal to either zero or one implies that only two types of contracts may be optimal. We refer to the contract where $d = 0$ as the 'pure carrot' contract, and to the contract where $d = 1$ as the 'stick and carrot' contract. We denote this pair of contracts with subscripts c and s respectively. Under the 'pure-carrot' contract, the agent is never dismissed and is incentivized only through bonuses. Under the 'stick and carrot' contract the agent is dismissed whenever output is low but the signal is good ($Y = L$, $\sigma = \tilde{G}$).[14]

The optimal balance between the carrot and the stick depends on the transparency of production, or the precision of the public signal q, as described in the following proposition.

Proposition. If $x > p\delta\gamma/\left(1 - \delta/2\right)\left(1 - p\right)$, then the optimal contract that is selected by the principal has the following properties:

1. the agent's basic wage is set at its lowest possible value, or $\omega = m + \gamma$.
2. There exists a threshold $\hat{q} \in (1/2, 1)$ such that:

   if $q < \hat{q}$, then the optimal contract is a 'pure carrot' contract: $d_c = 0$, and $b_c = \gamma/p$;

   if $q > \hat{q}$, then the optimal contract is a 'stick and carrot' contract: $d_s = 1$, and

[13] In online Appendix D we extend the model to the case where $d \in [0, 1]$. In Appendix E we consider an alternative extension, where the principal may warn the agent when he suspects him of shirking, and dismiss the agent only after an endogenously determined number of warnings. The qualitative results of the model regarding the effect of transparency q on the optimal contract are unchanged in both extensions.

[14] One may argue that the principal may have an incentive to renege on the contract chosen, and to avoid paying the bonus to the agent, or to not dismiss the agent when this is called for by by the contact. This is not a concern, however, if the principal is patient and faces many agents simultanuously who are likely to believe that once the principal reneges, she will continue to do so in the future.

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

if $q = \hat{q}$, then both contracts above are optimal.

If $x \leq p\delta\gamma/\left(1-\delta/2\right)\left(1-p\right)$, then either the 'stick and carrot' contract or dismissal of the agent upon observation of low output are optimal.

Proof. Denote by V the present value of the agent's utility from being employed in agriculture in a stationary equilibrium where he exerts high effort every period. The fact that the agent's utility upon dismissal is zero implies that:

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

Denote the probability of a bad harvest and a good signal for an agent who exerts high effort by $\mu = (1-p)(1-q)$. The probability of dismissal of an agent who exerts high effort is $d\mu$. It follows from (1) that:

$$V(b,d) = \frac{\omega + pb - m - \gamma}{1 - \delta\left(1 - d\mu\right)}. \tag{2}$$

The principal's objective function (OF) is to maximize her per-period expected payoff, denoted by π,

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

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

$$\begin{array}{l} p[b+\delta V] + (1-p)[q + (1-q)(1-d)]\delta V + \omega - m - \gamma \geq \\ p[q(1-d) + (1-q)]\delta V + (1-p)[q + (1-q)(1-d)]\delta V + \omega - m, \end{array} \tag{3}$$

where $V = V(b,d)$ as in (2). Part (1) of the proposition follows from the fact that since ω cancels out from (3) it is optimally set to $\omega = m + \gamma$. Plugging (2) into constraint (3) and simplifying yields the incentive constraint:

$$pb\left(1 + \frac{pqd\delta}{1-\delta\left(1-d\mu\right)}\right) \geq \gamma. \tag{IC}$$

Part (2) follows from the maximization of (OF) subject to (IC). Because the Principal sets b as low as possible, the incentive constraint is binding in the optimal solution. The threshold $\hat{q}$, is given by the unique solution in the interval $[0,1]$ of the quadratic equation that equates the values of the objective function with $d = 0$ and $d = 1$. To see that $\hat{q} > 1/2$ if $x > p\delta\gamma/\left(1-\delta/2\right)\left(1-p\right)$, rearrange this quadratic equation as:

$$\hat{q}/(1-\hat{q}) = (1-p)\,x[1 - \delta(p + \hat{q}\ - 2p\hat{q})]/p\delta\gamma, \tag{4}$$

and note that while the left-hand-side of (4) is convex and increasing from zero to infinity as $\hat{q}$ increases from zero to one, the right-hand-side is positive and linear in $\hat{q}$. This implies that there exists a unique intersection between the two curves in the interval $[0,1]$. The condition on x is obtained by requiring that for $\hat{q}=1/2$ the right-hand-side is larger than the left-hand-side, which is equivalent to $\hat{q}>1/2$.

Finally, in the analysis above we only considered two pure strategies, the third pure strategy of dismissal of the agent upon observation of low output regardless of the signal is dominated by the 'pure carrot' contract if $x > \delta p\gamma/(1-p)$. Thus, it is never optimal in the range where $x > p\delta\gamma/\left(1-\delta/2\right)(1-p)$. □

The switch of the optimal contract from 'pure carrot' to 'stick and carrot' when the quality of information improves captures the essence of our claims. The logic behind it is simple. A principal relying on a 'stick' to incentivize the agent has to incur the cost x whenever a dismissal takes place. But given that the agent is incentivized to exert effort, dismissal is in fact a "wasteful mistake" that occurs with probability $\mu=(1-p)(1-q)$. The probability of dismissal, and so also the expected cost of dismissal, μx, decreases when the quality of information q improves. When dismissal is sufficiently costly to the principal, incentivizing the agent through a stick is beneficial for the principal only when q is large enough. The threshold $\hat{q}$ is determined so that it exactly balances the expected cost of dismissal μx with the expected savings to the principal due to a smaller bonus.

The effect of transparency on income and its allocation

If the economy is less transparent ($q<\hat{q}$), the principal optimally refrains from ever dismissing the agent. In this case, the contract is socially efficient and the expected income of both the principal and the agent is independent of q. Under this 'pure-carrot' regime the expected income of the agent, I_c, and the principal, π_c, are:

$$I_c = m + 2\gamma \text{ and } \pi_c = p(H-L) + L - 2\gamma - m,$$

and their combined expected income is:

$$I_c + \pi_c = p(H-L) + L.$$

In contrast, if the economy is sufficiently transparent ($q>\hat{q}$), then the optimal contract is a 'stick and carrot:'

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

and

$$I_s + \pi_s = p(H - L) + L - \mu x.$$

The expected total income reveals that the 'stick and carrot' contract is socially inefficient because the agent is sometimes dismissed even though he works diligently. This efficiency loss, namely the expected cost of dismissal μx, declines as accuracy improves. In the limit, when the signal is accurate ($q = 1$), then the 'stick and carrot' regime becomes socially efficient.

The principal's payoff is continuous at the threshold of transparency $\hat{q}$ 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 and from a decline in the agent's income. Indeed, it is the agent who bears the entire burden of the 'stick and carrot' regime: at the threshold accuracy, $\hat{q}$, his expected income I drops discreetly by the expected cost of dismissal: $(1 - p)(1 - \hat{q})x$. Past that threshold, his expected per-period income continues to decline with q. Within this range, the benefit that the agent obtains due to the reduced probability of dismissal enables the principal to reduce the bonus payment b, while still maintaining the incentive constraint. These features are summarized by Figure 1 below. The principal's expected income π as a function of accuracy q is depicted by the lower solid line. Total expected income $I + \pi$ is depicted by the upper solid line; and the difference between these two lines represents the agent's expected income.

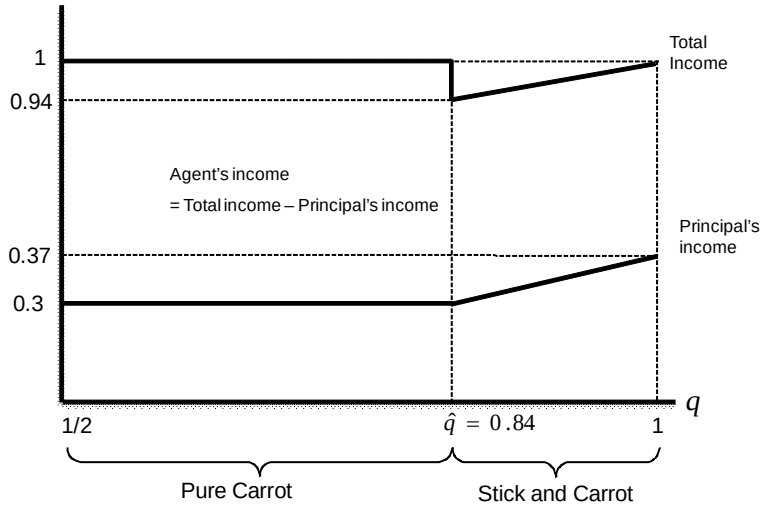


Figure 1: Periodic expected income as a function of signal accuracy

Figure 1 adopts a simple illustrative calibration. We set: $H = 1.1$, $L = 0.6$ and $p = 0.8$, so that a bad harvest with a significantly lower crop occurs once in about every five years, and so that the

expected crop size of each plot is set to one: $E(Y) = pH + (1-p)L = 1$.[15] 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 set the subsistence cost to $m = 0.5$ and the effort cost to $\gamma = 0.1$, and thus the basic wage is $\omega = 0.6$. Given an interest rate (in grain) of one third or more in the ancient world, we set $\delta = 0.75$. Finally, we set $x = 2$, so that the present value cost of dismissing and replacing an agent is two expected crops.[16]

It is instructive to compare the outcome when the signal fully reveales the state of the world ($q = 1$) with the outcome when the signal is highly inaccurate ($q < \hat{q}$). In both cases the diligent agent is never dismissed and the economy is efficient (As seen in Figure 1). However, the distribution of income is quite different. The agent's (gross) income falls from $I_c = m + 2\gamma$ in the range of the opaque signal to $I_s = m + 2\gamma - p\delta\gamma/[1-\delta(1-p)]$ when $q = 1$, since the bonus that the agent requires in order not to shirk is reduced to a minimum. The agent's utility from being employed in agriculture, namely his income net of effort, is entirely dissipated in this case if he is very patient ($\delta = 1$).

Discussion

Our findings thus far imply that when transparency is sufficiently low, the agent-tenant is in a 'pure carrot' regime in which he is never dismissed and could be considered a de facto owner of the land that he cultivates. This effective ownership is not due to the benevolence of the principal (the state), nor to any impediments that prevent dismissal. Rather, the agent has rights to the land because, given low transparency, it is optimal for the principal to refrain from dismissal.

In contrast, when transparency is sufficiently high, the agent-tenant is in a 'stick and carrot' regime, in which, when low output is observed, the farmer could be evicted, and thus cannot be considered to have ownership rights to the land that he cultivates. In this range, as transparency increases, the probability of an unjustified dismissal of an agent who exerts high effort is smaller, and so is the probability of retaining a tenant who exerts low effort. Thus, an increase in transparency

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

[16] With these parameters $\hat{q} > 1/2$ is achieved already with $x = 0.48$, however, in the version of the model in which d is continuous, (online Appendix D), a higher x is required for obtaining a range of $q > 1/2$ in which $d = 0$ is optimal. Thus, for consistency, we set $x = 2$. A number of additional elements that could be added to our model also serve to render the stick less attractive and help guarantee that $\hat{q} > 1/2$ with a much lower value of x. These include a cost to obtain a signal and effort exerted in land maintenance (since the threat of eviction might reduce this investment, as shown empirically by Deininger and Jin, 2006).

enables the principal to rely more on the 'stick' and less on the 'carrot'. That is, under the optimal 'stick and carrot' contract, a more accurate signal implies more efficacy for the threat of dismissal, and, correspondingly, a smaller share of output for the tenant and an increase in the revenue appropriated by the state.

The general effect of transparency on the optimal combination of the 'stick' and 'carrot' is robust and does not depend on our specific modeling assumptions. The credible threat of using the 'stick' reduces the cost of incentivizing the agent with the 'carrot.' However, to maintain credibility of the threat, punishment must be used (even if unjustifiably) whenever output is low and the signal is good. Since the probability of punishment declines with transparency, the expected cost of including a 'stick' in the contract declines with transparency.

The key ingredients that underline these results are standard in the literature. First, we assume that punishment takes the form of a threat of dismissal.[17] Second, we assume that the agent faces limited liability, in the form of a lower bound on material remuneration. The specific assumption that the minimal remuneration, ω, is equal or greater than the cost of subsistence when exerting high effort, $m + \gamma$, is made here in order to simplify the exposition. A strictly positive ω is crucial, though, for the effectiveness of the stick.

Our main departure from the existing literature is in including a public signal upon which the principal conditions the contract. This allows us to perform comparative statics with respect to the accuracy of this signal and thereby obtain new insights into the link between geography and institutions.

Finally, we would like to address the issue of the closure of the model as far as population and income is concerned. Our assumption that farmers' expected income exceeds the Malthusian

[17] One might question why we do not allow for corporal punishment as an incentive device, as was common with slaves, since this is presumably painful for the agent but plausibly imposes only a small cost on the principal. As Chwe (1990) notes, the rare use of corporal punishment in labor relations is altogether perplexing, given that it may seem to be Pareto improving in comparison to costly dismissal. Perhaps corporal punishment renders an agent less productive because of physical or emotional repercussions. We do not attempt to resolve this puzzle here. We note, however, that in ancient Egypt and Mesopotamia, the peasants were typically tenants, rather than slaves. The Laws of Hammurabi, cited in note 9 above, reveal that even if subject to eviction, tenants were treated as free individuals, and disputes between landlord and tenant were considered as civil cases, where corporal punishments were not practiced – unlike criminal cases where such punishments were common (such as cutting off a thief's hand – law 253). Moreover, slaves were not usually employed in agriculture in these societies. Dandamaev (1984, p. 277) concludes with respect to Southern Mesopotamia: "slave labor did not play a decisive role in agriculture . . . and proved to be unprofitable." We surmise that this may be due, at least in part, to the absence of the threat of dismissal: unlike tenants, slaves require close ongoing supervision. In addition, once heterogeneity of skills is taken into account, it is apparent that the dismissal of an agent (say, when his output is low but the signal is good) may serve not only as an ex-ante incentive device, but also as an ex-post selection device to weed out those with the lowest skills, implying that one should expect to observe dismissal when transparency is high, even if corporal punishment could be used.

threshold for stable population seems to imply a steadily growing farming population. As discussed further in Appendix G, we propose to close the model by assuming that any excess workers from the rural sector, including dismissed agents, are employed outside of farming, where the wage is low (particularly in famine time) and does not guarantee reproduction.[18]

2.2 A Two-Level Hierarchy Model

We now extend our model to include two tiers of government. Extension of the model further to n tiers of hierarchy along the same lines is straightforward. For the relations between the governor and the farmers in the village under her control we keep the basic model. For the relations between the upper echelon (the king) and lower level of hierarchy (the village governor), we employ a variant of our basic model where the governor may hide output rather than exert low effort, because this seems more consistent to us with the historical record.

We attach a subscript of 1 or 2 to variables at each level of the hierarchy, from the bottom up. We assume that there are two independent state variables that determine the state of nature in each plot of land: $\theta_1 \in \{G, B\}$ is plot specific, and $\theta_2 \in \{G, B\}$ is village specific. The plot-specific states are assumed to be independent across plots, conditional on the village's specific state, and the village specific states are assumed to be independent across villages. We denote by $p_1 \in (0, 1)$ the probability that each plot of land is in a plot-specific good state, and by $p_2 \in (0, 1)$ the corresponding probability for the entire village.

As in the basic model, output in each plot can be either low or high: $Y_1 \in \{L_1, H_1\}$ and the agent's effort can be either low or high: $e \in \{l, h\}$. Plot output is assumed to be high if and only if the agent exerts high effort and both the plot's and village's states of nature are good ($\theta_1 = \theta_2 = G$). Thus, the state of nature in a specific plot is good with probability $p_1 p_2$, otherwise it is bad.

We assume that the village specific state of nature, θ_2, is revealed to both the farmer and the governor after the farmer's effort decision is made. In addition, if the village specific state is good ($\theta_2 = G$), then the governor receives plot-specific signals σ_1 for each plot in the village. These signals are accurate with probability $q_1 \in [1/2, 1]$ and are (conditionally) independent across plots.

At the higher level of the hierarchy, between the village governor and the king, we assume an analogous information structure. The king does not know the specific states θ_1 of individual plots, nor the states θ_2 for any of the villages. But he receives an independent signal σ_2 about each of

[18] Wu (2012) presents historical evidence that cities were indeed a population sink, drawing population from the countryside yet with negative natural population growth. Clark and Hamilton (2006) show that net fertility before industrialization was significantly lower in urban regions of England in comparison to rural areas.

the latter, whose accuracy is denoted by the probability $q_2 \in [1/2, 1]$.

The contract selected by the governor will have the same structure as in the basic model. It specifies a basic wage $\omega_1 = m+\gamma$, a bonus b_1 if output is high, and a dismissal probability $d_1 \in \{0, 1\}$ at a cost of x_1 to the governor, if output is low ($Y_1 = L_1$) but both the village's state and the plot's signal are good ($\theta_2 = G, \sigma_1 = \tilde{G}_1$). Thus, subject to the farmer exerting effort, he is dismissed with probability: $\Pr(dismiss|e = h) = (1 - p_1)\, p_2\, (1 - q_1)\, d_1$.

The governor's maximization problem is a variant of the principal's problem in the basic model, in which $p_1 p_2$ substitutes for p as the probability of high output, and in which the probability of dismissal is $p_2\,(1 - p_1)\,(1 - q_1)\,d_1$ instead of $(1 - p)(1 - q)d$. Thus, the governor chooses a 'pure carrot' contract ($d_{1c} = 0$) if transparency is below some threshold, $q_1 < \hat{q}_1$, and a 'stick and carrot' contract if $q_1 > \hat{q}_1$. Above $\hat{q}_1$, the expected income of the governor is increasing with q_1.[19]

We now turn to study the king's problem. We assume that the number of plots in each village is sufficiently large so that the total revenue obtained by the village governor can be substituted by their expected value. The governor's revenue is then limited to two possible outcomes, depending on the village-specific state of nature θ_2. We denote by L_2 and H_2 the governor's income in a bad year ($\theta_2 = B$) and a good year ($\theta_2 = G$) respectively. If N_1 is the number of plots in a village, then:

$$\begin{aligned} L_2 &= N_1\left[L_1 - (m+\gamma)\right], \\ H_2 &= H_2(q_1) = N_1\left[p_1(H_1 - L_1 - b_1) + L_1 - (1-p_1)(1-q_1)d_1x_1 - (m+\gamma)\right]. \end{aligned}$$

The parameters b_1 and d_1 are those selected by the governor (as a function of q_1). Beyond the threshold $\hat{q}_1$, the good-year revenue H_2 is increasing in q_1.

Recall that the king receives a signal σ_2 regarding the village state θ_2, whose accuracy is denoted by q_2. As mentioned above, we assume that the governor may under-report the revenue collected to the king. That is, she may report collecting L_2, even though she in fact collected H_2. The king is assumed to employ an analogous two-edged incentive scheme to the one above: a bonus b_2 if the governor reports collecting H_2, and a threat of dismissal at a cost of x_2 to the king, if the governor's report is L_2, but the signal σ_2 indicates that a large village harvest was to be expected.

The king maximizes:

$$\pi_2 = \max_{b_2 \geq 0, d_2 \in \{0,1\}} p_2(H_2 - b_2) + (1 - p_2)[L_2 - (1 - q_2)d_2x_2].$$

[19] The corresponding bonus payments are: $b_{1c} = \gamma/p_1p_2$ under 'pure carrot' and $b_{1s} = (\gamma/p_1p_2)\left[1 - p_1p_2q_1\delta_1/\left(1 - \delta_1(1-p_2) - \delta_1p_2(p_1 + q_1 - 2p_1q_1)\right)\right]$ under 'stick and carrot'. If $p_2 = 1$, this is identical to the analogous expressions under the basic model.

The incentive constraint for the governor is:

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

where $\delta_2 V_2$ is the discounted value of the governor from keeping her position. Under the optimal contract the incentive constraint is binding. Setting the governor's utility of unemployment to zero, we obtain, in analogy to (1) in the basic model:

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

from which it is possible to solve for $V_2(b_2, d_2)$ as in (2), and then solve explicitly for the optimal incentive scheme b_2 and d_2 selected by the king.

Thus, subject to additional parameter restrictions on x_2 and δ_2 that are analogous to those in the Proposition, there exists a threshold $\hat{q}_2 > 1/2$ such that if village farming is sufficiently opaque to the king ($q_2 < \hat{q}_2$) the governor enjoys a carrot regime, in which she is autonomous in the sense that she is never dismissed, namely $d_{2c} = 0$. In this regime, the per-period revenue to the king is independent of the state of nature and is given by $\pi_{2c} = L_2$; the governor retains for herself the difference $b_{2c} = H_2 - L_2$ whenever the village state of nature is good, and zero otherwise.

On the other hand, when village farming is sufficiently transparent to the king ($q_2 > \hat{q}_2$), a stick and carrot regime prevails. Under this regime, the governor is dismissed whenever the king is led to expect high revenue based on his observed signal, but the governor reports that collected revenue is low. This occurs with probability $(1 - p_2)(1 - q_2)$. In this regime, following a similar derivation to the one in the basic model $d_{2s} = 1$ and $b_{2s} = (H_2 - L_2) - q_2 \delta_2 V_{2s}$, where:

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

and the king's expected revenue is:

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

The threshold transparency level $\hat{q}_2$ is determined by the implicit condition $\pi_{2s} = \pi_{2c}$. As in the basic model, the transparency threshold $\hat{q}_2$ increases with the cost of dismissal x_2 and decreases with the governor's discount factor δ_2.

As in the basic model above, the balance of power between the king and a provincial governor depends on the transparency of the provincial economy to the king. When local conditions are sufficiently opaque to the king, the intermediary governor enjoys substantial autonomy in that she pays a (relatively low) fixed tribute and always retains her position. But if the transparency of the

local provincial economy to the king is sufficiently high, then the governor is subject to dismissal and retains a relatively lower share of the revenue collected. If transparency is very high and the governor is very patient, she retains little beyond her minimal maintenance costs.

3 Application: ancient Egypt and Mesopotamia

In this section we demonstrate that the theoretical insights obtained from our model are consistent with the institutions of the major civilizations of the ancient near east during the fourth to the second millennia BCE.

The vast majority of the population in antiquity was rural.[20] However, archaeological and textual sources about antiquity pertain almost exclusively to the urban centers and to the elite, biasing the typical reconstruction of society towards the perspective of the urban elite. As argued by Hicks (1969), theory may be required in this case to offer a more balanced reconstruction of history. In that spirit, we believe that our model provides a framework within which one may better understand key social institutions that persisted over many centuries, even millennia, and that distinguished between the different ancient civilizations.

Our theory provides three main predictions that link transparency to institutions.

Predictions.

According to our basic model:

(1) When farming is locally transparent, farmers do not own the land they cultivate,

(2) The more transparent farming conditions are, the higher state capacity is.

And according to the hierarchical extended model:

(3) When farming is less transparent to the central state, local lords retain autonomy and higher income.

To recall, in the introduction, we suggest that ancient Egypt can be characterized at one extreme, with farming being highly transparent to both the local elite and the state's center. Northern Mesopotamia is representative of the other extreme, with low transparency at both the local and the central levels. Southern Mesopotamia, we suggest, was comparatively transparent at the local level, but quite opaque to the central state. In this section we substantiate these crude characterizations, and examine the consistency of our three main predictions with the evidence on the institutions that emerged in these regions.

[20] In 1961, ninety four percent of the population in Bangladesh was rural; and as late as 1980 – the earliest year for which FAO statistics (http://faostat.fao.org) provides with data on agricultural population – eighty nine percent of the population in Ethiopia was engaged in agriculture. A similar pattern must have prevailed in antiquity.

Intensive agriculture was first adopted in the highlands of Anatolia and northern Mesopotamia in the seventh millennium BCE. This occurred two and three millennia respectively before agriculture was taken up in the alluvial planes of Southern Mesopotamia (Sumer) and in the Nile valley. It was in Sumer, however, that the first major city-states were formed in the fourth millennium – an advance that some scholars describe as a "takeoff" or as an "Urban Revolution" (Liverani 2006, Childe 1951). Still, the first central territorial state was formed in Egypt, in about 3000 BCE, starting from a core in Upper (southern) Egypt (Kemp 2006, Wenke 2009). The rapidity of the formation of a central state and its subsequent stability are among the key features that distinguish between ancient Egypt and Mesopotamia, leading Baines and Yoffee (1998, p. 268) to conclude: "the two civilizations are profoundly different."

Scholars have often noted major additional distinguishing features (Trigger 1993, 2003). Thus, fortified city-states existed in predynastic Egypt, but Egyptian cities ceased to be fortified after the formation of the central state and played a fairly limited role as administrative centers. This led Wilson (1960) to famously characterize Ancient Egypt as "a civilization without cities." In contrast, up to the first millennium, southern Mesopotamia was ruled most of the time by rival and independent city-states, leading Adams (1981) to characterize southern Mesopotamia as "the Heartland of Cities." Unlike the cities in Egypt, those in southern Mesopotamia retained their power, and as a result, successive attempts to unify Mesopotamia under a central state were prone to failure. At the same time, the mostly rain-fed highlands of Northern Mesopotamia gave rise to more limited city-states than in the alluvial plains of Southern Mesopotamia.

3.1 Egypt

Farming in the Nile valley originated in the late fifth millennium BCE, towards the southern tip of Upper Egypt, from whence the Egyptian central state subsequently emerged.[21] The distinctive technology that underlined Egyptian agriculture in the Nile valley is known as "flood basin irrigation." This farming technology prevailed with relatively minor modifications for more than five millennia, until the construction of the first Aswan Dam at the beginning of the twentieth century.[22] Natural and man-made lateral dykes across the narrow valley connected the river banks

[21] The term Upper Egypt refers to the Nile valley south of Cairo. For brevity we shall avoid reference to the Nile's delta and to the Fayum depression.

[22] Our description of the flood basin method is based on Willcocks (1899) and Butzer (1976). We avoid mention of the cultivation of raised land in the Nile's riverbanks and along the desert margins, which required extended low gradient canals in parallel to the river. These were developed by far-sighted Pharaohs only in the second and first millennia BCE, and facilitated the cultivation of orchard gardens and two annual crops.

and the desert margin to create an extensive system of large basins. When the Nile swelled in mid-August, the local inhabitants would breach its embankment to fill each basin system. This created a system of terraced pools, each covered by about 1.5 meters of standing muddy water. The water was retained in the basins for about 40 days, depositing its rich mineral nutrients and soaking the soil. In early October, after the Nile receded somewhat, the water in the basin system was gradually drained back into the Nile, leaving the fields to be sowed, mostly with barley.[23] The moisture trapped in the soil served as the sole source of water during the growing season. In late March, the farmers would harvest the ripened crops, before the hot winds of April and May could parch the grain stalks and cause the seeds to disperse to the ground.

The unpredictable fluctuations in the Nile's inundation level caused annual variations in crop output. Particularly high inundation that could break the dikes and flood the villages and the grain stockpiles posed as much of a threat as a particularly low inundation level. The timing, the length and the severity of the hot spring winds contributed to the uncertainty. However, in any given year the conditions that farmers faced were fairly homogeneous within each basin, and even across basin systems. As a result, farming activity was highly transparent not only locally, but also to the central government. Scholars have long noted evidence that records of the Nile's peak level of inundation were already kept in the third millennium BCE (Kemp 2006, p. 64). The peak level was recorded apparently because this information served the Pharaohs as a control device. Cooper (1976, p. 366) describes the taxation of Egyptian agriculture in the middle ages: "Agriculture was so well regulated in Egypt that, on the basis of the Nile flood recorded by the Nilometer, the government knew in advance what revenue to anticipate." In particular, "The height of the Nile flood determined how much and in what manner the tax assignments were made in each district." We conjecture that this was generally the case also in antiquity, for which no similar detailed records survive.[24]

G. Baer (1969, p. 17) describes village life since the medieval period up to modern (nineteenth century) Egypt. He contends that it was characterized by three phenomena: (a) village land was periodically redistributed by the village-head to the peasants; (b) the village inhabitants were collectively responsible for tax payments; (c) the village as a whole was held responsible for

[23] The seed, scattered by hand in a broad-throw, was then buried in the wet soil by animals' hooves or by light plowing.

[24] The transparency of Egyptian farming was also due to the relative ease of monitoring farming activity in real time by traveling by boat along the Nile. Evidence that this was done in practice is provided by the Wilbour Papyrus that contains the minutes of a monitoring expedition from about 1140 BCE that recorded rent assessments for more than 2,000 large plots of land in Middle Egypt that were owned by temple institutions (Kemp 2006, pp. 254-6). Abundant pictorial evidence reveals also that local officials measured the extent of fields prior to the harvest.

maintaining irrigation work and for providing labor for any required public works. This characterization implies that the village headman exercised tight control over village land. Even though it was customary to assign the same fields to the same farmer annually, or to his heir, Baer's first characteristic implies that farmers did not have secure tenure and that village heads (or the estate managers, in the case of temple land) could reassign fields as they saw fit. This is consistent with our first prediction, as well as with the long-standing phenomenon, whereby land in Egypt was not owned by the people who cultivated it, but rather by absentee landowners.[25] Indeed, the prevailing notion in ancient Egypt was that the entire land belonged to the Pharaoh (Baines and Yoffee 1998, p. 206); or as Hughes states: "in theory all the land belonged to Pharaoh throughout Egyptian history" (1952 p.1). However, as all scholars recognize, this notion coexisted in various periods with a practice by which much land was owned by institutions and by prominent individuals. It has been established, for example, that during the Hellenistic period a significant fraction of the land was de facto "owned" by the temples, by various lay organizations and by powerful individuals such as court officials and military officers (for the Ptolemaic period see Manning 2003, pp. 65-98).

Nevertheless, from our perspective, the more significant feature of the land tenure system in Upper Egypt is that even when land was privately held, it was owned by absentee landlords. Consistently with prediction 1, Hughes (1952, pp. 1-2) summarizes that in the first two millennia of the historic period there was never "a large body of small landholders who managed and worked their plots themselves ... the lowest classes were largely serfs on the domains of Pharaoh, the wealthy and the temples."[26] This was the case, we posit, because the high transparency of farming eliminated the main disadvantage to absentee ownership, leaving the peasants vulnerable by denying them the shield that is otherwise provided by their inherent informational advantage. Significantly, in the few known cases from the mid-first millennium BCE where private land leases survived from antiquity, the lease contracts were for one year only (Hughes 1952), providing further support for our proposed mechanism that tenants were constantly under the threat of not having their contract renewed.[27]

[25] It was apparently indignation over the fact that Egyptian farmers did not own the land that they cultivated that led the authors of the Old Testament to refer to Egypt as a "house of bondage," even though slavery was no more pervasive in ancient Egypt than elsewhere in the Ancient Near East.

[26] Eyre (1997) contends that the divorce between land-ownership and actual farming was endemic to Egypt and persisted essentially until the mid-twentieth century. According to G. Baer (1962, pp. 1-70; 1969, pp. 62-78), even the major agrarian reforms during the nineteenth century that gave land title to the cultivating peasants (with the intention of increasing tax collection by eliminating intermediary tax collectors), ended up with much of the land reverting to large absentee landlords.

[27] Another feature of Egyptian farming that reduced the advantages to owner-occupied farming or to long-term leases was that land could not in effect be over-exploited by the cultivating tenants, since land fertility was sustained

From the perspective of the local lords the transparency of Egyptian farming was instrumental in raising revenue, by efficiently incentivizing the peasants with minimal material reward and without the need for close supervision. The Nile's global transparency, in turn, enabled the Pharaohs to employ a similar incentive scheme towards the district governors and down the chain of middlemen who remitted taxes from the periphery to the center. That is, consistently with predictions 2 and 3 above, we argue that the high transparency of the state of nature in every district (Nome), and in every flood basin system within each district, explains why the Pharaohs were so powerful without engaging in direct control, and why the provincial centers retained so little independent power. This is consistent with Eyre's (1994, p. 74) summary: "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 credible threat of dismissal enabled the pharaohs to run a highly lean state bureaucracy and to siphon off a substantial share of the tax revenue. Indeed, at least in the early Old Kingdom period, the positions of governors and state bureaucrats were by a revocable appointment, and non-hereditary.[28] The leanness of the intermediating bureaucracy is closely related to the relative weakness of the cities in the different districts. The provincial cities remained essentially administrative centers, without amassing substantial independent wealth to threaten the predominance of the center, in accordance with Wilson's above cited observation that Egypt was "a civilization without cities." The high transparency at all levels of the governmental hierarchy can also explain the rapidity of the formation of a strong central state in Egypt and for its remarkable subsequent stability.

Our theory can also account for other striking phenomena that distinguished ancient Egypt from ancient Mesopotamia. Consistent with the absolutist power of the Egyptian monarchs, the Pharaoh was considered as god incarnate, unlike their counterparts in Mesopotamia.[29] In addition, the lack of legal title to land and the hierarchical nature of land management help explain the relatively low number of real-estate transactions in ancient Egypt (in comparison to Mesopotamia).[30] The absence

by the Nile's deposits and agrarian capital investment was by way of dikes and local canals that had to be undertaken communally.

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

[29] The Mesopotamian kings (with a single exception of Sargon in early Akkad) were only considered as envoys of the gods (Baines and Yoffee 1998).

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

of legal title to land also implies that land could not be used to secure loans, thus helping to explain the paucity of early records of loans, once again in comparison to their abundance in Mesopotamia. These factors, together with the homogeneity of the country and the relatively small surplus that remained in the countryside, help explain the relatively limited extent of commercial activity in ancient Egypt (again, in comparison to Mesopotamia). Another remarkable difference between these two ancient civilizations was noted by Wilson (1960): whereas law codes were promulgated in Mesopotamia as early as the late third millennium BCE, the earliest known legal code from ancient Egypt is from the second half of the first millennium BCE.[31]

3.2 Southern Mesopotamia

Cereals were cultivated in arid Southern Mesopotamia on the outer slopes of the rivers' levees, and on the levees of abandoned courses of the rivers, and depended entirely on irrigation from the Euphrates or the Tigris. As noted already, two major problems were created by the fact that these rivers are fed by winter rains from the far-away mountains to the north and north-east and by melting snow in the spring (Adams (1981, pp. 3-6) and Postgate (1994, p. 178)). First, since the rivers are relatively low in the cultivation season, agriculture depended on canal irrigation, and on mechanisms for the distribution of scarce water that could not irrigate all the potentially arable land.[32] Second, the swelling of the rivers in the spring endangered the harvest and required the diversion of the excess water away from the sloped fields and into the plain at the marshy lower end of the cultivation zone.

It was the need to overcome these two major problems that apparently delayed the adoption of extensive agriculture in Babylonia until long after agriculture flourished in Northern Mesopotamia and irrigation systems were established in southwest Iran (Wilkinson 2003, pp. 72-76). An intricate system of canals was eventually constructed to direct water from the elevated river to the fields and channel excess water to the swampy alluvial plain below the cultivated fields (Adams (1981,

pales in comparison to the abundant data on land transactions from Mesopotamia – and this cannot be attributed solely to perishability of the papyrus records, given the abundance of other textual evidence from Egypt.

[31] The village leadership in ancient Egypt thus resolved local disputes within the village, applying traditional common law (Eyre 1999, p. 44), while disputes among those higher up in the hierarchy were resolved by the authority just above the disputants. For surveys of the legal institutions of ancient Egypt and Mesopotamia see Westbrook (2003).

[32] Adams (1981, p. 6) estimates that due to the shortage of water, only 8,000-12,000 square kilometers could be cultivated out of a potential estimated by Wilkinson (2003, p. 76) to be about 50,000 square kilometers. The shortage of water at the critical cultivation season is evidenced by the use of irrigation fees, already in the late third millennium BCE (Ouyang 2010). This underscores the power available to those who could deny water.

p. 245) and Wilkinson (2003, p. 89)).[33]

Unlike the homogeneity of the cultivated land in Egypt, farming conditions in Southern Mesopotamia were thus quite complex. Even fields within the same zone could vary in quality, depending mostly on how high they were above the saline water table in the adjacent marsh. The overriding factor however was the absolute dependency of cultivation on rationed water that was controlled upstream, and that could have been directed elsewhere. Therefore, farmers were completely dependent on the local elite who controlled the flow of water at various canal junctures. This control provided the elite families with relatively efficient means to extract surplus from the cultivating peasants. We thus categorize farming activity in Southern Mesopotamia as highly transparent to the local elite.

Consistent with prediction 1, we contend that this transparency explains why owner-cultivated farming was practically nonexistent in Southern Mesopotamia. As in Egypt, cultivation was conducted by peasant sharecroppers who were managed by a hierarchy of intermediaries, under the ultimate control of dominant elite families who resided in the urban centers (and who controlled each city's temple). This system has been described as "a pyramid of individual families" (Steinkeller 1999, p. 293) or as an "institutional household" (Renger 1995). In accord with prediction 2, this high local transparency explains why powerful early city-states, controlled by elite families, were able to form in Southern Mesopotamia. Indeed, once irrigation agriculture was introduced, it led to relatively rapid development of civilization. More than thirty major city-states have been identified in Southern Mesopotamia in the fourth and third millennia BCE. Writing is believed to have originated in about 3200-3100 BCE in the largest of these cities, Uruk, when its population was about twenty thousand (Yoffee 2005, p. 43).

The operation of the complex irrigation system in Southern Mesopotamia required skilled local managers with a "thorough knowledge of local conditions on a day-to-day basis" (Hunt 1987, p. 172). Unlike the case of Egypt, the local managing elite in Southern Mesopotamia were thus indispensable and irreplaceable. In other words, we interpret the farming activity in Southern

[33] Additional innovations that contributed to the success of farming in Southern Mesopotamia included the cultivation of long deep furrows in narrow sloped fields, plowed by oxen. The narrow fields, which could be more than a kilometer long, sloped down from the feeding canal towards the marshy plain. The deep furrows enabled seeding and watering only within the furrows, and also helped to divert the saline topsoil mostly away from the plants. Liverani (2006, p. 28) refers to the local administration of agriculture in Southern Mesopotamia as the "secondary agricultural revolution," and attributes the "urban revolution" to these agricultural innovations, and in particular to the use of oxen, which he views as a source of substantial economies of scale that contributed not only to higher productivity but also to the centralization of farming. The attribution of the urban revolution to these developments, rather than to the adoption of irrigation, may however reflect his aversion to Wittfogel's ideas.

Mesopotamia as comparatively opaque to remote authority. According prediction 3, we contend that this opacity explains why the local elite in Southern Mesopotamia were extremely resilient and why strong cities were one of the most distinctive features of the Mesopotamian civilization. Our hierarchical model implies that even when early city-sates in Southern Mesopotamia managed to conquer competing city-states, they still needed the cooperation of the existing local elites in order to obtain on-going tax revenue from the conquered countryside.

Consistent with our reversed-Witffogelian explanation of the association between irrigation and state power, we note that in Egypt and Southern Mesopotamia irrigation was maintained and managed at the local level. It was the specific knowledge possessed by the local elites in the Mesopotamian cities that assured their essential autonomy.

This feature helps explain why several aggressive attempts to unify Southern Mesopotamia under one of the rival city-states ended in failure after a relatively short period – in marked distinction to the quick and durable unification of Egypt. The rival city states of Southern Mesopotamia fought each other periodically for a millennium, before they were first consolidated under Sargon of Akkad in about 2350 BCE. However, 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 too collapsed. The next territorial state was established by Hammurabi of Babylon in 1790-1760 BCE, but it weakened substantially under his heirs and collapsed altogether at about 1600 BCE. Thus, until the first millennium, Mesopotamia was ruled most of the time by rival city-states, with only brief intermittent periods of a central territorial state.[34] This analysis is consistent with Yoffee's (2005) description of the fate of Sargon's earliest central state, according to which Sargon of Akkad was well aware of the intermediation problem when he ascended to power, and sought "to disenfranchise the old landed aristocracy" (p. 37). But after conquering the diverse city states in Southern Mesopotamia, he ruled them through appointed "royal officials, who served alongside the traditional rulers of the conquered city-states" (p. 142). It was this "uneasy sharing of power ... [that] led to a power struggle" and to the ultimate demise of Sargon's territorial states (Yoffee 1995, pp. 292-293; 2005, p. 143).

[34] Starting in the first millennium BCE, the successive empires of Assyria, Babylonia and then Persia, like the subsequent Greeks and Romans, developed administrative methods that enabled them to subject formerly independent city-states; yet, even under these empires, the cities in southern Mesopotamia typically retained much of their former autonomy (Van de Mieroop 1997, pp. 128-139).

3.3 Northern Mesopotamia

Due to the uncertain and idiosyncratic nature of rainfall and the relative unevenness of the terrain, farming in the highlands of Northern Mesopotamia was compartaivley opaque.[35] Early urbanization occurred in Northern Mesopotamia during the late fifth and early fourth millennia BCE, but ceased already in the later part of the fourth millennium – when the first city-states started to flourish in Southern Mesopotamia.[36]

Wilkinson (1994; 2003, p. 211) concludes that the settlement pattern in Northern Mesopotamia was generally characterized by a scatter of a large number of roughly equivalent, nucleated units, with each unit administered by a central settlement with a radius of control of about five kilometers, determined by the "constraining effect of land transport and the convenience of being within one day's round trip of the center" (1994, p. 503). Wilkinson (2003, p. 211) attributes this nucleated settlement pattern to the fact that no site had an "overwhelming situational or demographic advantage." Without disputing the factual basis of this observation, we differ on the logic behind this argument. By the winner-takes-all (increasing returns to scale) nature of violent conflicts, a priori advantage is not a prerequisite for the formation of larger central states under city leaders who happen to militarily defeat their neighbors. From our perspective, the key to the nucleated pattern of semi-autonomous administrative units in early Northern Mesopotamia was the inability of the winner of any such territorial conflict to extract on-going revenue from distant conquered territories. In a more pronounced version of the situation in Southern Mesopotamia, we thus propose that the localized nature of the early states in this region is consistent with our third prediction: the significant opacity of farming activity in Northern Mesopotamia limited the span of control of its early city states.[37]

Cuneiform documents about farming in Northern Mesopotamia are available mostly from the mid-second millennium BCE, from the vicinity of the town of Nuzi. This evidence reveals that the local kings and the elite owned large estates, but that unlike Southern Mesopotamia, the temples did not possess much economic power. Much land was owned also by nuclear families who worked their

[35] See Wilkinson (1994) and Jas (2000). Agriculture in Northern Mesopotamia was, however, significantly less opaque than that in more arid regions of the Ancient Near East. Noy-Meir (1973) demonstrates how extreme the effects of spatial variations in micro-climate and terrain quality can be on desert plant populations.

[36] The large size of these early cities and the architectural remains of the dwellings suggest that these cities were inhabited not only by the elite, but also by the farming peasants (Ur 2010). This pattern of inhabitance is in fact consistent with the limited span of control and with the elite's inability to raise the needed resources to secure the countryside from banditry, forcing the peasants to seek protection within the walls of the central city.

[37] The requirement of transporting the crop tribute to the center over land (rather than by water) was another contributing factor for the limited span of control of early potential states in Northern Mesopotamia.

patrimonial land. The evidence suggests that land ownership in Northern Mesopotamia was in fact in a constant state of flux, with small landholders regularly losing their land to rich families through debt and sale under duress (Zaccagnini 1999; Jas 2000). The persistence of owner-occupied farming in Northern Mesopotamia indicates that the process of land consolidation must have been matched by an opposing process of the gradual dissolution of large, presumably less efficient, estates.[38] The prevalence of owner-cultivated private farming in Northern Mesopotamia is consistent with prediction 1, given the presumed low transparency of farming in that region.

4 Application to the modern growth of the state

Our proposal concerning the impact of transparency on the technology for extracting taxes and on the size of the state is applicable not only to antiquity, but also to the substantial growth in the share of the state in aggregate income in the past century and a half, following several millennia in which that share did not exceed about ten percent. The scholarly explanations for the recent increase in the scale of government focus mostly on increased demand for public services, and particularly on the increased demand for redistribution, reflecting the impact of interest groups in democracies. Underlying these explanations is the premise that the increase in income and the spread of democracy have led modern governments to give more attention to the public's welfare. Olson's (1993) non-teleological perspective leads us to question the applicability of these explanations. We find it unlikely that the appetite of the Bourbons or the Romanovs for tax revenue was any less than that of the revenue maximizing ancient Pharaohs. That is, they did not collect more because they did not want to, but because they could not. If anything, one would thus expect the greater attentiveness of modern democracies to public welfare to have lowered the burden of taxation below the maximum that the Leviathans of previous centuries would have charged had they ruled today.[39] This consideration leads us to conclude that tax revenue, as a share of income, has increased in the recent past due to an increase in the state's capacity to tax rather than due

[38] The Nuzi records from the mid-first millennium reveal that existing laws sought to preserve patrimonial land by forbidding the sale of private land to non-relatives. The evidence also reveals how these laws were circumvented by disguising land sales as "adoptions," where the seller adopted the purchaser as a son and became a tenant on his former land (Zaccagnini 1984). Jas (2000) quotes Warriner (1948, pp. 21, 104), who noted that the ancient land tenure regimes in Northern and Southern 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."

[39] Our argument against the democratization explanation accords with the findings by Mulligan et al. (2004), who find no significant difference between modern democracies and non-democracies in the pattern of economic and social policies.

to democratization.

In analogy to our contention that the Neolithic Revolution gave rise to the state, not by raising productivity but by increasing transparency and appropriability, we argue that it is the transformation of the tax technology that was induced by the Industrial Revolution which accounts for the recent increase in the scale of the state, rather than any rise in productivity per se. The shift away from agriculture in the past century or two may have eliminated the significance of the environmental factors that we emphasized above. However, we contend that an analogous channel of increased transparency explains how the shift to industry has led to the unprecedented recent increase in the state's share of total product.

Our model thus offers a general theoretical framework and terminology in which to imbed the ideas of Kau and Rubin (1981) and of Kleven et al. (2009), who attribute the modern growth of the state to a decline in the cost of collecting taxes. These scholars argue in particular that the cost of collecting taxes decreased as a result of a shift away from self-employment (in agriculture) to market production by hired labor: the attendant massive paper trail created by the increase in record keeping and accounting exposed the private sector to income taxation, and in effect enabled the state to turn private companies into tax collection agencies. We contend that these arguments illustrate exactly what is captured here by the idea that increased transparency of production induces a transformation in the state's ability to tax.[40]

5 Conclusion

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

Based on this theory, we conduct a comparative analysis of the salient institutions of the earliest states. In particular, we argue that the rapid rise of a powerful central state in Egypt, its subsequent

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

resiliency, the weakness of its cities and the lack of land-owning peasantry can all be explained by the high transparency of Egyptian farming, both at the local level and at the state level. The same paradigm explains also key institutional differences between Northern Mesopotamia, where transparency was low, and Southern Mesopotamia, where it was high at the local level, but much lower at the central level.

Our environmental theory of early institutions contributes to the understanding of antiquity by providing and applying a new paradigm. Reflecting the ideas of Polanyi (1957) and Finley (1999), it is common among anthropologists, archaeologists and historians of antiquity to maintain that ancient economies were fundamentally different from modern ones, and that economic theory, with its focus on the central role of markets, cannot be applied to the study of antiquity. Even though this perception has eroded somewhat in recent years, the underlying grounds for rejecting the applicability of economic theory to the study of the ancient world have not changed (see Van de Mieroop 1999, 2004). North (1977, p. 706) similarly conceded that economic historians have not begun to account for non-market allocative systems, and thus cannot say much about societies in which markets had little allocative effects. Accordingly, we avoid here altogether the role of markets, and yet apply economic theory to explain how hierarchical, non-market relations shaped the institutions of antiquity and determined the distribution of resources.

Our theory of institutions also sheds light on key modern economic and political concerns, unrelated to antiquity or to the impact of the environment. In particular, the theory underscores the role of differential degrees of informational asymmetry in understanding the architecture of hierarchical institutions. Although the prevailing perception is that asymmetry of information hinders efficiency and encumbers taxation, our proposed framework reveals that the lack of transparency of agents' activities ('privacy') may be protective of agents' freedom, and may possibly promote their material well-being.

References

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

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

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

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

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

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

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

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

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

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

[11] Banerjee, Abhijit V., Paul J. Gertler, and Maitreesh Ghatak, (2002), "Empowerment and efficiency: tenancy reform in West Bengal," *Journal of Political Economy*, 110, 239-280.

[12] Banerjee, Abhijit V. and Ghatak, Maitreesh. (2004), "Eviction threats and investment incentives," *Journal of Development Economics* 74, 469-488.

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

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

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

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

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

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

[19] Buonanno, Paolo, Ruben Durante, Giovanni Prarolo, and Paolo Vanin, (2012), "Poor Institutions, Rich Mines: Resource Curse and the Origins of the Sicilian Mafia," Carlo Alberto Notebooks No. 261.

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

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

[22] Chattopadhyay, Suhas (1979), "Operation Barga" Social Scientist, 8, 41-48.

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

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

[25] Chwe, Michael Suk-Young (1990), "Why Were Workers Whipped? Pain in a Principal-Agent Model," *The Economic Journal*, 100, 1109-1121.

[26] Clark, Gregory and Hamilton Gillian (2006), "Survival of the Richest : The Malthusian Mechanism in Pre-Industrial England", *Journal of Economic History*, 66/3, 707-736.

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

[28] Dandamaev, Muhammad A. (1984), *Slavery in Babylonia: From Nabopolassar to Alexander the Great (626–331 BC)*, translated by Victoria A. Powell, Northern Illinois University Press.

[29] Dari-Mattiacci, Giuseppe (2013), "Slavery and Information," *Journal of Economic History,* 73, 79–116.

[30] Deininger, Klaus and Jin, Songqing (2006), "Tenure security and land-related investment: Evidence from Ethiopia" *European Economic Review*, 50, 1245–1277.

[31] De la Sierra, Raul Sanchez (2013), "On the Origin of States: Stationary Bandits and Taxation in Eastern Congo," working paper, Colombia University.

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

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

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

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

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

[37] 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.

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

[39] Fenske, James (2014), "Ecology, Trade, and States in Pre-Colonial Africa," Journal of the European Economic Association, 12, 612-640.

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

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

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

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

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

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

[46] Hossain, M., 1982, "Agrarian reform in South Asia—a review of recent experiences in selected countries," in: Steve Jones, Puran Chandra Joshi, and Miguel Murmis (eds.), *Rural Poverty and Agrarian Reform*, Allied Publishers, New Delhi, 142– 161.

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

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

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

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

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

[52] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[75] Roth, Martha T. (1997) *Law Collections from Mesopotamia and Asia Minor*, 2nd edition, Scholars Press.

[76] Segal, Ilya and Michael D. Whinston (2012), "Property Rights" in Robert Gibbons, and John Roberts (eds.), *The Handbook of Organizational Economics*, Princeton University Press, 100-158.

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

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

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

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

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

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

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

[84] Tirole, Jean (1986), "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics, & Organization*, 2, 181-214.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ONLINE APPENDIX

Appendix A: Hiding Output

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

An incentive scheme, $b > 0, d \in \{0, 1\}$, induces truthful reporting of the agent if:[1]

$$b + \delta V \geq (H - L) + ((q(1 - d) + (1 - q))\delta V. \tag{A1}$$

where $H - L$ is the output stolen by the agent when he reports low instead of high output, and V denotes the present value of the agent's utility from being employed in agriculture in a stationary equilibrium with truthful reporting. The agent's incentive constraint is binding in the optimal solution (otherwise the principal can lower the bonus payment b) and so:

$$b = (H - L) - q\delta dV. \tag{A2}$$

The value function $V(b, d)$ associated with truthful reporting (analog of (2) in the basic model) is:

$$V(b, d) = \frac{pb}{1 - \delta(1 - \mu d)}. \tag{A3}$$

Plugging (A3) into (A2) and simplifying yields an incentive constraint:

$$b = (H - L)\left(1 - \frac{\delta pqd}{1 - \delta + \delta d\,(\mu + pq)}\right). \tag{A4}$$

The principal's objective is:

$$\pi = \underset{b,d\in\{0,1\}}{max}\, p(H - L) + L - pb - \mu dx - m, \tag{A5}$$

subject to (A4).

Thus, two types of contracts may be optimal: one with $d = 0$ ('pure carrot') and another with $d = 1$ ('carrot and stick'). The threshold transparency level $\hat{q}$ that determines the level above which the'carrot and stick' is optimal is given by the solution of the following equation (analogous to (4) in the basic model) that equates the expected profit to the principal under the two contracts:

$$\frac{\hat{q}}{1 - \hat{q}} = \frac{(1 - p)\,x}{p\delta(H - L)}[1 - \delta(p + \hat{q}\ - 2p\hat{q})]. \tag{A6}$$

[1] Notice that the incentive constraint is relevant only in case the state of nature is good and output is high.

A pure carrot contract is optimal if $q < \hat{q}$. It is given by:

$$d_c = 0, b_c = H - L, \text{ and } V_c = p(H-L)/(1-\delta). \tag{A7}$$

A stick and carrot contract is optimal if $q > \hat{q}$. It is given by:

$$d_s = 1, b_s = (H-L)\left(1 - \frac{\delta pq}{1-\delta(p+q-2qp)}\right), \qquad V_s = \frac{p(H-L)}{1-\delta(p+q-2qp)}. \tag{A8}$$

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

Appendix B: Costly Monitoring

Suppose that the model is identical to the basic model except that the principal can observe a signal $\sigma \in \left\{\tilde{l}, \tilde{h}\right\}$ about the agent's effort at cost $c \geq 0$ (in units of output) instead of on the state of nature as in the basic model. The accuracy of the signal is $q \in [1/2, 1]$, such that:

$$Pr(\tilde{h}|h) = Pr(\tilde{l}|l) = q \ ; \ Pr(\tilde{h}|l) = Pr(\tilde{l}|h) = 1 - q.$$

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

As in the basic model, $\gamma > 0$ is the periodic cost of exerting high effort, the agent's alternative employment outside of agriculture tenancy provides utility of zero and the agent's periodic utility, U, when engaged in agriculture equals his expected income, to be denoted by I, less the cost of effort. In particular, when exerting high effort, this periodic utility is: $U = I - \gamma$.

We denote the present value of the agent's utility from being employed in agriculture by V, and denote by $\delta \in (0, 1)$ the agent's discount factor.

The principal is assumed to rely on the following incentive scheme. If output is high, then the principal retains the agent with certainty and pays the agent $\omega + b$, where $b \geq 0$ is a bonus payment. If output is low, then the agent is still paid the basic subsistence wage $\omega = \gamma$.

When output is low, if the signal indicates that the agent was exerting high effort ($\sigma = \tilde{h}$), then the principal retains the agent. But if output is low and the signal indicates that the agent was shirking ($\sigma = \tilde{l}$), then the principal may dismiss the agent.

We denote by $d = 1$ the strategy of dismissal upon low output and a signal indicating low effort: $\sigma = \tilde{l}$ and $Y = L$, and retention of the agent otherwise, and by $d = 0$ the strategy of always retaining the agent. If the agent is dismissed, the principal incurs a fixed cost $x > 0$ (in units of output). We assume that this cost is large enough to ensure that it will not be desirable to dismiss the agent when output is low ($Y = L$) and the signal indicates high effort.

Thus, the principal can either imply a contract with $d = 1$ in which he incurs the monitoring cost c, or she can employ a contract with $d = 0$ and no monitoring.

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

$$V = pb + [1 - Pr(dismiss|e = h)]\delta V. \tag{B1}$$

For convenience, we denote the probability of a bad harvest and a good signal by $\mu = (1-p)(1-q)$. The probability of dismissal upon high effort is then $d\mu$. V is thus determined by the contract parameters b and d and the parameters: μ, p and δ as follows:

$$V(b,d) = \frac{pb}{1-\delta(1-\mu d)}. \tag{B2}$$

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

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

subject to providing the agent with incentives to exert high effort (identical to the basic model):

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

where $V = V(b,d)$ as in (B2).

Since $\omega = \gamma$, we can rewrite the principal's objective function and the agent's incentive constraint as follows:

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

s.t.

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

Thus, we obtain that modeling monitoring as a (costly) signal on effort, yields a maximization problem that for $c = 0$ is identical to the maximization problem in the main model. More generally, the larger is c the higher would be the threshold $\hat{q}$ above which the optimal contract is 'stick & carrot', without any change in the qualitative results. This indicates that the larger is c - the more costly it is to obtain a signal on effort as in this model or on the state of nature, as in the main model - the larger is the range of parameters for which the solution is 'pure carrot'. This means that if $c > 0$ then the threshold $\hat{q}$ is strictly larger than $1/2$ for lower values of the cost of replacement x.

Appendix C: Conditioning Effort on State of Nature

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 constraint is relevant only when the state of nature is good, $\theta = G$. Under a 'pure carrot' contract, the agent's incentive constraint reduces to:

$$b \geq \gamma.$$

In a stationary equilibrium, the value of the agent's discounted utility when he exerts high effort when the state of nature is good, and when he obtains a per-period utility of γ when the state of nature is bad (he doesn't exert effort in those periods) satisfies:

$$V = [pb + (1-p)\gamma] + [1 - Pr(dismiss)]\delta V, \tag{C1}$$

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

$$V_s = \frac{pb + (1-p)\,\gamma}{1 - \delta\,(1 - (1-p)\,(1-q))}. \tag{C2}$$

The principal's objective function remains unchanged. The incentive constraint under 'stick and carrot' is:

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

This implies that the principal sets the bonus so that:

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

By replacing $b_s = \gamma - q\delta V$ in (C2) we obtain that

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

as in the basic model, and

$$b_s = \gamma\left(1 - \frac{q\delta}{1 - \delta\,(p + q - 2pq)}\right). \tag{C3}$$

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

$$p\gamma - [pb_s + (1-p)\,(1-q)\,x]\,, \tag{C4}$$

and hence, since $b_s < \gamma$, for a sufficiently large x there exists a threshold $\hat{q}$ above and below which the principal chooses $d = 1$ and $d = 0$, respectively. In particular, plugging (C3) into (C4) and setting it equal to zero, yields $\hat{q}$ as the unique solution in the interval $[0, 1]$ of the following quadratic equation ($\hat{q}$): by:

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

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

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

Appendix D: Probabilistic Dismissal

In this appendix we consider again the basic model, but we allow the principal to dismiss the agent upon observation of low output and a good signal with any probability $d \in [0,1]$ as opposed to just $d \in \{0,1\}$ as in the main text. We recast the principal's problem as the minimization of discretionary expenditure:

$$\min_{d\in[0,1],b} pb + \mu x d, \tag{D1}$$

subject to the agent's incentive constraint:

$$pb = \left(1 + \frac{pqd\delta}{1-\delta(1-d\mu)}\right) \geq \gamma. \tag{D2}$$

The agent's incentive constraint must be binding in the optimal solution. Plugging the value of b from (D2) into (D1) yields the principal's objective function

$$\gamma\left(1 - \frac{\delta pqd}{1-\delta+\delta d\,(\mu+pq)}\right) + \mu x d. \tag{D3}$$

as a function of d alone.

Differentiation of the principal's objective function with respect to d yields:

$$-\frac{\gamma\delta qp(1-\delta)}{A^2} + \mu x \tag{D4}$$

where $A = 1-\delta+\delta d(\mu+pq)$.

Inspection of (D3) reveals that the expression on the left of (D3) is convex in d while the expression on the right is linear and increasing in d. Comparison of the values of these two expressions at $d=0$ reveals that if

$$q \leq \frac{(1-p)x(1-\delta)}{\delta\gamma + (1-p)x(1-\delta)} \tag{D5}$$

then the value of d that maximizes the principal's objective function (sets the derivative (D4) equal to zero) is negative. Because d is a probability, this means that the optimal probability of dismissal in this case is $d=0$. Comparison of the values of these two expressions at $d=1$ yields another condition on q such that the value of d that maximizes the principal's objective function is larger

than one. Because d is a probability, this means that the optimal probability of dismissal in this case is $d = 1$.

Thus, there exist two threshold values $\underline{q}$ and $\overline{q}$ such that for $q < \underline{q}$ the optimal $d = 0$; for $q > \overline{q}$ the optimal $d = 1$; and for $\underline{q} \leq q \leq q < \overline{q}$ the optimal value of d (obtained from solving the first-order-condition equation $D4 = 0$) is given by:

$$d = \frac{1-\delta}{\delta(\mu + pq)} \left(\sqrt{\frac{\gamma \delta pq}{(1-\delta)\mu x}} - 1 \right) \tag{D6}$$

If the right-hand-side of (D5) is larger than .5 or, equivalently,

$$\frac{(1-p)x}{\gamma} > \frac{\delta}{1-\delta} \tag{D7}$$

then $\underline{q} > .5$, which means that the pure carrot contract is optimal for some values of the accuracy parameter q. Inspection of (D7) reveals that this is the case if the cost of dismissal x is sufficiently large and/or the agent is impatient (δ is small) so that the threat of dismissal is less effective.

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

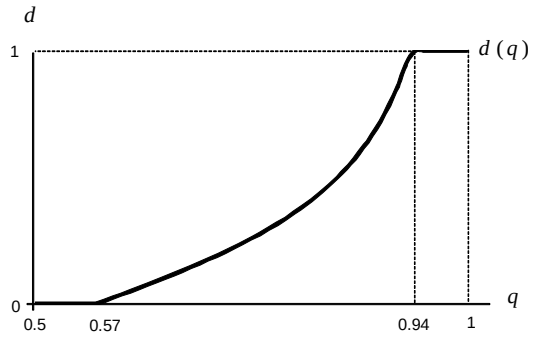


Figure 5: The optimal dismissal probability, $d \in [0, 1]$, as a function of transparency q

As in the basic case, the agent's bonus is maximal when $q < \underline{q}$. In the range above $\underline{q}$, as the probability of dismissal increases, the bonus decreases – since the increased threat of dismissal is used as a substitute incentive device. The bonus continues to decrease further in the range where $q > \overline{q}$, where the dismissal probability reaches its upper limit ($d = 1$). The principal's net expected revenue (taking into account the costs of dismissal) is constant below the threshold $\underline{q}$ and increases monotonically in q above $\underline{q}$.

Appendix E: Warning before Dismissal

In this appendix we allow the principal to warn the agent an optimally chosen number of times when output is low and the signal about the state of

nature is good before actually dismissing the agent. That is, we assume that the principal optimally selects an integer number n of "bad signals," or times at which will observe $Y = L$ and $\sigma = \widetilde{G}$ before it dismisses the agent. The number of "warnings" prior to dismissal is thus given by $n-1$. The basic model is therefore one where n is restricted to the set $\{1, \infty\}$.

Let $V(n)$ denote the value of being employed in agriculture for an agent with n bad signals left. If $n = 1$ then the agent is dismissed the next time $Y = L$ and $\sigma = \tilde{G}$. The agent is dismissed immediately upon $n = 0$ and so $V(0) = 0$. Let $b(n)$ denote the bonus payment to the agent when $Y = H$ as a function of the number of bad signals that remain n.

The value function $V(n)$ satisfies the following recursive equation:

$$V(n) = pb(n) + \mu\delta V(n-1) + (1-\mu)\delta V(n). \tag{E1}$$

The agent's incentive constraint, which as before is binding in the optimal solution, can be simplified to:

$$pb(n) = \gamma - pq\delta(V(n) - V(n-1)). \tag{E2}$$

By combining (E1) and (E2) we obtain the following recursive formulation for $V(n)$:

$$V(n) = A + BV(n-1), \tag{E3}$$

where the constants A and B are given by:

$$A = \frac{\gamma}{1-\delta+\delta(\mu+pq)}; B = \frac{\delta(\mu+pq)}{1-\delta+\delta(\mu+pq)}. \tag{E4}$$

Observe that $0 < A$ and $0 < B < 1$.

Given that $V(0) = 0$, the solution for $V(n)$ in terms of the parameters of the model is:

$$V(n) = \frac{A(1-B^n)}{1-B}. \tag{E5}$$

It therefore follows that:

$$b(n) = \gamma/p - q\delta AB^{n-1}. \tag{E6}$$

Observe that the bonus payments to the agent increase with n. It can be immediately verified that $b(1)$ and $V(1)$ are identical to b_s and V_s of the basic model, while b_c and V_c coincide to the limits of $b(n)$ and $V(n)$ from (E6) and (E5), respectively, as n tends to infinity.

We now solve for the optimal number n. Denote the principal's discount factor by δ_P, and denote the discounted expected discretionary costs for the principal (that include bonus payments and dismissal costs) starting from the point where it employs an agent has k bad signals left until dismissal under a policy where agents are dismissed after n bad signals and are induced to exert high effort in every period by $c(k, n)$.

For $k = 1$:

$$\varphi(1, n) = pb(1) + \mu(x + \delta_P c(n, n)) + (1 - \mu)\delta_P c(1, n).$$

And for $1 < k \leq n$:

$$c(k, n) = pb(k) + \mu\delta_P c(k - 1, n)(1 - \mu)\delta_P c(k, n).$$

These two equations simplify to:

$$c(1, n) = \alpha b(1) + \beta x/\delta_P + \beta c(n, n), \tag{E7}$$

and

$$c(k, n) = \alpha b(k) + \beta c(k - 1, n), \tag{E8}$$

where the two constants α and β are given by:

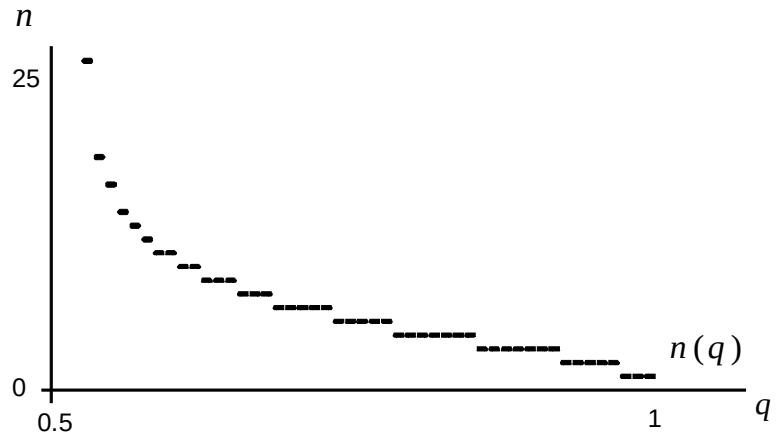
$$\alpha = \frac{p}{1 - \delta_P + \mu\delta_P}; \quad \beta = \frac{\mu\delta_P}{1 - \delta 2 + \mu\delta_P}. \tag{E9}$$

Equations (E7) and (E8) can be explicitly solved for $c(n, n)$ as a function of the underlying parameters of the model as follows:

$$c(n, n) = \frac{\gamma}{1 - \delta_P} + \frac{\beta^n x}{\delta_P(1 - \beta^n)} + \frac{\alpha q \delta A(B^n - \beta^n)}{(1 - \beta^n)(B - \beta)}. \tag{E10}$$

It is reassuring to confirm that the solution of the equation $c(1, 1) = c(\infty, \infty)$ for q yields the threshold $\widehat{q}$ from the basic model, and is independent of the principal's discount factor δ_P.

The following figure describes the optimal n (the n that minimizes (E10)) as a function of the level of transparency q, for the same parameters used to illustrate the basic model. The additional parameter δ_P is set to $\delta_P = 0.98$.[2]



[2] A lower discount rate for the principal reduces the discounted cost of dismissal and shifts the curve of optimal n's downwards.

Figure 6: The optimal number of “bad signals" before dismissal, n, as a function of transparency q

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

Finally, it should be noted that in our calibration the probability of a bad signal (upon exerting effort) is $\mu = 0.2(1-q)$. Hence, a bad signal or warning is not issued more frequently than about every five years. In this case, the expected time needed for five warnings is much larger than the expected life span of an adult farmer, and so is effectively equal to infinity.

Appendix F: Endogenous Population Size

In this appendix we allow the principal to control the size of individual plots. This generalization yields new predictions with respect to the effect of transparency on the size of the population.

Suppose that output from a plot of size λ is:

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

The agent's cost of high effort is denoted by $\gamma(\lambda)$. The cost function $\gamma(\lambda)$ is assumed to be increasing and convex and to be such that $\gamma(0) = 0$. A larger plot size is associated with a larger cost of training a new agent. We therefore assume that the replacement loss is given by $x(\lambda) = \lambda x$.

If the size of the land is controlled by the principal is T, then the number of plots (and agents) is given by T/λ. The principal is assumed to maximize her expected payoff from the entire land under her control. Thus, her problem is:

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

s.t.

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

The analysis of the basic model where $\lambda = 1$ applies to any $\lambda > 0$. Both the subsistence and incentive constraints are binding in the optimal solution, which implies that $\omega = m + \gamma(\lambda)$. If the signal about the state of nature is uninformative (q is sufficiently low), a ‘pure carrot’ contract where:

$$d_c = 0, \qquad b_c = \gamma(\lambda)/p \tag{F1}$$

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

$$\lambda_c \gamma'(\lambda_c) - \gamma'(\lambda_c) = \frac{m}{2}. \tag{F2}$$

Similarly, if the signal about the state of nature is sufficiently informative (q is sufficiently high), then a 'stick and carrot' contract where:

$$d_s = 1, \qquad b_s(q, \lambda) = \frac{\gamma(\lambda)}{p} - \frac{q\delta\gamma(\lambda)}{1 - \delta(p + q - 2pq)}. \tag{F3}$$

is optimal. The principal's problem in this range is equivalent to the selection of λ to minimize $T(m + \gamma(\lambda) + pb_s(q, \lambda)]/\lambda$. As before, the optimal solution λ_s is given by the unique solution to the first order condition:

$$\lambda_s \gamma'(\lambda_s) - \gamma'(\lambda_s) = \frac{m}{2 - \frac{pq\delta}{1-\delta(p+q-2pq)}}. \tag{F4}$$

The convexity of $\gamma(\lambda)$ implies that the left-hand-side of (F2) and (F4) is increasing in λ. The fact that the right-hand-side of (F2) is smaller than that of (F4) and the right-hand-side of (F4) is increasing in q implies that the optimal plot size under the 'stick and carrot' regime λ_s increases with transparency q, and is larger than the optimal plot size under the 'carrot' regime λ_c.

The fact that $\lambda_s > \lambda_c$ is due to the fact that when the stick is in use, it costs less to incentivize the agent, and so the principal may as well assign a larger plot size to the agent, which would allow it to economize on the fixed cost of agents' maintenance. The larger plot size implies, of course, a smaller population.

The extra decision variable λ leads to a 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, consider the case where the cost function $\gamma(\lambda)$ has a constant elasticity $\lambda\gamma'(lambda)/\gamma(\lambda) = K$, calibrated so that $\gamma(1) = \gamma$ so that the optimal plot size under the 'pure carrot' regime is still equal to one ($\lambda_c = 1$). This guarantees that under the 'pure carrot' contract every aspect of the economy is identical to that of an economy with a fixed plot size. However, the higher revenue under the 'stick and carrot' regime implies that the new threshold transparency $\hat{q}_\lambda$ for switching into the 'stick and carrot' contract is lower than before. At the transparency threshold $\hat{q}_\lambda$ the agents are made discretely worse off when they are switched from a 'pure carrot' contract to a 'stick and carrot' contract. But beyond this point, since each agent's net per-period 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 $\hat{q}$ agents are better off than under the fixed plot case. This is compatible with increased revenue to the principal, since the number of agents is smaller.

These results are similar to those depicted in Figure 1. If we set $T = 1$ so that the principal's expected income is identical to her income under a fixed

plot size, then the threshold $\hat{q}_\lambda$ 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.

To conclude, this appendix shows that if plot size is endogenous then as economic activity becomes more transparent, the lower is population density.

Appendix G: The Urban Sector

In the model, we implicitly assume that all those individuals who do not belong to the elite and are not employed in agriculture belong to the urban sector. To simplify, we assume further that the urban sector does not trade with the farming sector. That is, the provision of protection and the collection of tribute ('protection' revenue) is the only interaction between the two sectors. We also simplify by consideration of a model with a single tier of government, where the governor is identical to the king. The food collected by the governor is evidently not consumed entirely by her. This food revenue provides the means for supporting an army that provides protection to the farming sector and secures the governor's monopoly on the extraction of revenue from farming activity. This food supply also sustains the artisans who supply various amenities (including luxury items) for the governor and his dependents, and may also possibly be exchanged for prestige goods from abroad. Since some of the food that reaches the urban sector is in some sense wasted on sumptuary meals or on imports, the ratio of the average food collection to the food required for long-term maintenance of farmers (m) provides an estimate of an upper bound on the size of the urban sector that is supported by the farming sector.[3]

More significant than the relative sizes of the two sectors is the very different uncertainty in food supply that they face. The essence of this issue can be clarified by considering what happens in bad years. At the level of the individual farmer bad years occur with probability $1 - p_1 p_2$. At the governor's level, however, they occur less frequently, with a lower probability of $1 - p_2$. This reflects the fact that the governor's revenue bundles together the revenue from many independent plots, and thus provides an insurance against idiosyncratic plot bad states. However, our model also identifies a difference in the severity of bad harvests due to village bad states. In this case, our assumptions imply that the output of each farmer is L_1, and the revenue collected by the governor is $L_2 = N_1 [L_1 - (m_1 + \gamma)]$. In the numerical calibration presented in the main text we set $L_1 = m_1 + \gamma$. This implies that the income retained after a bad harvest enables farmers to survive until the next harvest, but the governor and the urban sector obtain no revenue at all. This extreme result is clearly due to our simple model and to this particular calibration; but it reflects a general phenomenon: a larger share of the farming output remains in the periphery after bad harvests.

[3] If farmers are employed in the construction of monuments over the Summer, and are paid for their extra effort by the state, as was customary in Egypt, this too would have to be taken into account.

This captures another important and ill-understood aspect of ancient economies in which the urban sector was likely to be more vulnerable to downward shocks to output. This implies that hunger and starvation are likely to be concentrated particularly among the lower strata of the urban sector: servants, small artisans and the like. This implication is in line with our presumption that this segment of society is demographically vulnerable, and may not have reproduced on its own, other than through an inflow from the farming sector. In addition, under the circumstances assumed here, the vulnerability of the urban sector implies that whereas farmers need only store food within the year, inter-annual storage is an absolute necessity for the urban sector, as a buffer for years where the harvest is small. This inter annual storage, however, should not be considered as providing insurance for the farming sector, but rather as serving the urban sector.[4]

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