Geography, Transparency and Institutions*

Joram Mayshar† Omer Moav‡ Zvika Neeman§

April 7, 2017

Abstract

We propose a theory in which geographic attributes explain cross-regional institutional differences in: (1) the scale of the state, (2) the distribution of power within state hierarchy, and (3) property rights to land. In this theory, geography and technology affect the transparency of farming; and transparency, in turn, affects the elite's ability to appropriate revenue from the farming sector, thus affecting institutions. We apply the theory to explain differences between the institutions of Ancient Egypt, Southern Mesopotamia and Northern Mesopotamia, and also discuss its relevance to modern phenomena.

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

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

*We have benefited from comments from Daron Acemoglu, Bob Allen, Josh Angrist, Ernesto Dal Bo, Eddie Dekel, Diana Egerton-Warburton, Christopher Eyre, James Fenske, Oded Galor, Maitreesh Ghatack, Jeremy Greenwood, Avner Greif, 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. Mayshar's research was supported in part by The Falk Institute for Economic Research in Israel. Address: Mt. Scopus. Jerusalem 9190501, Israel.

‡Department of Economics, University of Warwick; School of Economics, Interdisciplinary Center, Herzliya; CAGE and CEPR. Moav's research is supported by the Israel Science Foundation (Grant No. 73/11). Address: Department of Economics, University of Warwick, Coventry CV4 7AL, United Kingdom.

§Eitan Berglas School of Economics, Tel-Aviv University. Address: Ramat Aviv 6997801, Israel.

1 INTRODUCTION

Following North (1981), recent theories about the success of nations ascribe a paramount role to the protection of property rights. Acemoglu and Robinson (2012) argue that the greatest detriment to economic prosperity is the presence of extractive institutions that compromise property rights. Ancient Egypt, however, had a prosperous civilization, built the great pyramids, and was stable over several millennia, in spite of having an extractive government and no land property rights for its peasant farmers.

We thus propose that North's thesis about the importance of property rights pertains to post-agricultural societies, where private capital accumulation assumes a dominant role, but is less relevant for understanding agricultural societies where land is the main capital asset. This calls for an alternative theory to explain the success of some nations in the preindustrial world and the failure of others. In this paper, we seek to explain variations among pre-modern farming societies in: the scale of the state; the relative power of the center versus the periphery; and the land tenure regime. Unlike Acemoglu and Robinson (2012), who argue that institutions are by and large determined by the vagaries of human history, we propose a mechanism that explains how differences in institutions are the result of differences in geography and technology.

Our basic argument is that the government's ability to appropriate revenue from the farming sector was a key factor that accounts for differences between the institutions of earlier states; and that this ability to appropriate was significantly affected by the transparency of production, and, in turn, by geographical and technological conditions. In a nutshell, we attribute the resilience and control of Ancient Egypt's central government, the relative weakness of its regional cities, and the peasantry's non-ownership of land, to the fact that its farming was highly transparent, and thus readily appropriable. From this perspective, Egypt is a polar case. Low transparency, on the other hand, explains the existence of owner-occupied farming and the relative weakness of the states in other regions, such as ancient Northern Mesopotamia.

We choose to illustrate the applicability of our model by focusing on the ancient civilizations of the Near East during the late fourth to the second millennia BCE, since these were pristine cases of societies under relatively stable economic and military conditions, prior to the emergence of monetized taxation and the military and administrative innovations that facilitated the creation of empires. Although we focus on the role played by tax technology in ancient states, we believe that our theory is pertinent to other established states as well, providing an important insight in understanding pre-modern, agriculture-based states in general.[1] Moreover, our appropriability theory can also help explain some modern phenomena. First, since social institutions exhibit substantial inertia, our explanation of the institutions in farming-based societies can improve our understanding of current ones. And, to the extent that institutions impact the prosperity of nations, our model can expose deep-rooted factors that account for the current variation in the wealth of nations.[2] Secondly, several scholars attribute the unprecedented increase in the relative scale of government in the past century to a decline in the cost of collecting taxes.[3] Our transparency theory provides a formal structure and a broader perspective to this argument. Thus, according to our theory, there is an analogy between the long-term effects of the Agricultural Revolution in antiquity and the modern Industrial Revolution: in both cases the increased transparency of production transformed the state's tax capacity.[4]

To better understand how our approach relates to the extensive literature on state intuitions, we present the literature survey after presenting the model (in the next section) and its application to the civilizations of antiquity (in the section following the model). However, we want to clarify at the outset that there are two key issues that we do not deal with in this paper. First, we assume that the state is already established and that its government has a monopoly on the power to coerce. Thus, we do not discuss the emergence of the state, nor do we address the possible existence of external rivals and of warfare. Implicitly, in the spirit of Olson (1993), we posit that a governing hierarchy is an outcome of the advantage possessed by a dictator who monopolizes theft in the form of taxation, over uncoordinated theft that destroys incentives. Our focus is on how the sovereign's access to information impacts its ability to appropriate. Second, we abstract from differences in land productivity, which is the focus of a large body of literature that seeks to explain differences in early state development. In this tradition, Diamond (1997), for example, attributes the relative backwardness of New Guinea to its low land productivity relative to that in Eurasia. And Dal Bó et al. (2015) contend that states developed to provide defense in order to resolve what they identify as "the paradox of civilization," namely the vicious circle by which increased

[1] The notion of a 'tax technology' was proposed by Mayshar (1991).

[2] See Bockstette et al. (2002) and Spolaore and Wacziarg (2013).

[3] Kau and Rubin (1981) and Kleven et al. (2016) contend that the shift away from self-employment in agriculture into production by hired labor transformed the capacity to tax, since it was accompanied by a paper trail that rendered private production much more transparent to the modern state and thus facilitated income taxation.

[4] Consistently with our claim, de la Sierra (2016) 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 conflict between rival armed groups and to the monopolization of violence in the coltan rich regions; whereas an increase in the price of gold, which is easier to conceal and is hence less transparent, did not.

land productivity encourages predation, but farmers' enhanced insecurity discourages the investment that generates that increase in productivity.

The theory that we propose is based on a variant of the conventional principal-agent paradigm.[5] We address here three features that we incorporate in this framework. First, we focus on variation in the extent of information asymmetry between agents, representing tenant farmers/tax payers, and the principal, representing an absentee land-owner or the government. In particular, the principal observes a signal about the state of nature that determines the productivity of farmers' effort. On the basis of this signal, the principal infers with some error whether the agent worked diligently or not. The accuracy of this signal is our main exogenous variable, representing the degree of transparency of farming. The level of inundation of the Nile in Egypt is an example of such a signal. Second, we limit the incentive scheme that is available to the principal by assuming that in addition to remuneration (carrot), the only feasible sanction (stick) is the threat of dismissal upon suspected shirking. We assume that such dismissal is costly also for the principal.[6] In the spirit of Shapiro and Stiglitz's (1984) "efficiency wages" theory of employment contracts, this implies that unlike the standard applications of the principal-agent framework, the agent's outcome is not pinned down to his outside option. Third, to make the threat of dismissal meaningful, we embed the model in a multi-period setting.

The model's results are fairly intuitive: when the signal is more accurate, the role of the carrot is smaller, the role of the stick is larger, and the principal collects more revenue. Our interpretation of these results is that greater transparency induces a form of servitude, since the tenant is denied tenure and may be evicted upon suspected shirking. On the other hand, when there is low transparency, the agent retains more of the output without facing the threat of dismissal.[7]

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

[6] Our assumption that the sanction is in the form of a threat of eviction is consistent with the literature on tenancy contracts (e.g. Banerjee and Ghatak 2004). One might question why we do not allow for corporal punishment as an incentive device, as was common with slaves; since this is painful for the agent but plausibly imposes only minor costs on the principal. We do not attempt to resolve this puzzle here, but note that Chwe (1990) points out that corporal punishment is rare in labor relations, even though it is common for criminal offences. Moreover, we also note that the peasants in ancient Egypt and Mesopotamia were almost invariably free tenants, rather than slaves, while slaves were not usually employed in agriculture (Dandamaev 1984, 277). We surmise that this may be due to the fact that in the absence of the threat of dismissal, slaves (unlike tenants) require close ongoing supervision.

[7] 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

In North's (1981) depiction of the evolution of property rights in western societies since the Middle Ages, property rights are granted by an authoritarian government that seeks to maximize its revenue. This is the case also in our framework. However, in North's formulation the elite grant property rights to encourage private investment by the non-elite; that is, property rights serve as a commitment device to overcome the hold-up problem of ex-post expropriation. In contrast, in our framework, private investment does not have a role. By focusing on the information constraints that hinder the appropriation of output, our theory offers an alternative explanation for the emergence of property rights to land. When transparency is high enough, the threat of dismissal — an indicator for the lack of property rights — is the prime motive for the agent to exert effort. But with sufficient opacity — when the cost of erroneous dismissal outweighs the benefits — the absolute, non-benevolent state, willingly gives up the option to dismiss, thus granting farmers de facto title to the land they cultivate. That is, according to our theory, property rights to land are explained by the extent of information asymmetry.

In a two-layered extension of the model, designed to explain variations in the extent of state centralization, we examine the role of different degrees of transparency at different tiers of the governmental hierarchy. We show that when local farming activity is sufficiently transparent, not only to the intermediary (governor), but also to the upper level of the hierarchy (king), the intermediary retains a smaller share of the revenue and is subject to dismissal. On the other hand, if farming activity is sufficiently opaque to the king, the governor retains autonomy and a larger share of revenue.

We contend that the success of early central states, such as ancient Egypt, was due to high global transparency that enabled the central authority to keep the subordinated intermediary lords at bay, and to extract a larger share of revenue from the periphery to the center. In contrast, the fragile and fragmented structure of the early states in Northern Mesopotamia reflects the region's low local and global transparency. In an intermediate case, high local but low global transparency of farming in Southern Mesopotamia resulted in strong local urban elites that managed to retain power in the face of repeated attempts to subjugate them to a unified central state.

We note that our theory complements recent attempts to attribute key features of Imperial China to its fiscal capacity. According to Ma (2011), Imperial China achieved long-term success by replacing a hereditary feudal system with a rotating meritocratic bureaucracy.

that when the principal can choose costly monitoring to obtain a signal about 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. Dari-Mattiacci (2013) provides a similar theory, based on information asymmetry, to explain slavery.

Denying tenure to local provincial officials prevented their ability to acquire the local informational advantage that would otherwise have given them independent power. We find this administrative innovation to be analogous to the (natural) lack of informational advantage by provincial officials in Ancient Egypt, which gave greater power to the Pharaohs.[8]

2 THEORY

We develop a version of the conventional principal-agent model to facilitate analysis of the effect of the extent of informational asymmetry. We 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, and produces either high or low output. High output is obtained if and only if the agent exerts high effort and the state of nature is “good.” Each agent decides whether to exert high or low effort. His payoff is the payment received from the principal, less his cost of exerting effort. The principal incentivizes agents using a ‘carrot’ in the form of a bonus payment upon delivering high output, and a ‘stick’ in the form of dismissal as punishment for suspected shirking. We assume that dismissal is painful for the agent, who is forced out of farming and into the urban sector, where he becomes a servant and enjoys no rents. Dismissal is also costly for the principal (the state). The principal designs a contract that maximizes her expected periodic income. This income equals the total output produced by all agents, net of the payments to the agents, and net of the cost incurred by dismissing agents suspected of shirking. The principal does not observe the state of nature, but observes a signal on this state. The accuracy of this signal is our main exogenous variable, representing the degree of transparency of farming. The model’s main result is that dismissal is conditioned on failure to deliver high output, coupled with a signal that is sufficiently accurate and that indicates that the state of nature was likely to have been “good.”

2.1 The Basic Model

The annual output (Y) produced by the agent and the agent’s choice of effort (e) can be either low or high: $Y \in \{L, H\}$, and $e \in \{l, h\}$, respectively. The state of nature is also binary: either good or bad: $\theta \in \{G, B\}$. Output is a function of the effort exerted by the agent and the state of nature, whereby output is high if and only if the state of nature is

[8] On the other hand, Sng (2014) and Sng and Moriguchi (2014) seek to explain the weakness of early modern Imperial China by focusing on the agency problems that resulted from the size of the Empire. They posit that the vast size of the Empire created inherent difficulties in supervising local intermediaries, who used their power to extort taxes, whereas the central state sought to keep tax rates low to prevent revolts.

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 ex ante probability that the state of nature is good is denoted by: $p \in (0, 1)$. The agent chooses the level of effort before he learns the state of nature. After choosing the level of effort, both the agent and the principal observe a public signal about the state of nature: $\sigma \in \{g, b\}$. The accuracy of this signal, $q \in [0.5, 1]$ is such that:

$$Pr(g|G) = Pr(b|B) = q;\ Pr(g|B) = Pr(b|G) = 1 - q.$$

The accuracy level q represents the degree of transparency of production. If $q = 1$ then the signal perfectly reveals the state of the world (in this case, if $\sigma = g$ and $Y = L$, the principal can be certain that the agent shirked); if $q = 0.5$ then the signal is uninformative.

Thus, we model transparency and land productivity as exogenous, and abstract from the fact that both could be affected by the principal and the agent. The principal might invest, for example, in monitoring, or in increasing productivity through irrigation systems. Farmers could impact transparency and productivity by the choice of crop type, or by investing in land improvements. We contend, however, that our abstraction should not have a qualitative effect on the theory's prediction as long as exogenous geographical factors have a major impact on transparency and productivity.

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 low effort, and $\gamma > 0$ is the annual cost of exerting high effort. We assume that even low output is sufficient to cover the cost of upkeep of an agent who exerts high effort: $L \geq m + \gamma$. We assume also that $H > L + \gamma/p$. This implies that it is desirable for the principal to incentivize the agent to exert effort.

Both the agent and the principal are assumed to be risk neutral. 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. The utility of a dismissed agent is normalized to zero. We assume that the agent has no other sources of income or wealth, and that he cannot save or borrow. The agent's intertemporal discount factor is denoted by $\delta \in (0, 1)$.

The principal's incentive scheme is such that if output is low, she pays the agent a basic wage ω. If output is high, she pays the agent $\omega + a$, where $a \geq 0$ is an added bonus. The basic wage ω must sustain an agent who exerts effort until the next harvest: $\omega \geq m + \gamma$. When output is high the principal retains the agent. The agent is also retained when output is low and the signal indicates that the state of nature is bad ($\sigma = b$). But if output is low and the signal indicates that the state of nature is good, the principal may dismiss the agent and replace him with another. For simplicity, we assume that the principal employs a non-probabilistic dismissal strategy, that is, the dismissal probability d satisfies: $d \in \{0, 1\}$.[9] Thus, there are only two types of contracts: $d = 0$ and $d = 1$. If the agent is dismissed, then the principal incurs a fixed cost $x > 0$ that represents the cost of dismissing the agent and the present value of lost output while recruiting and training a new agent.

The contract strikes the optimal balance between the use of the carrot (a) and the stick (d). 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, and 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 ($Y = L, \sigma = g$), which occurs with probability $\mu = (1 - p)(1 - q)$ if the agent exerts high effort.

The expected cost of including the stick in the contract, μx, is thus decreasing with transparency q. We assume that the dismissal cost x is sufficiently high to preclude the possibility that the agent will be dismissed whenever output is low, irrespective of the signal, and thus to guarantee that sufficiently low transparency renders dismissal too costly and results in a 'pure carrot' regime. In particular, we assume that $x > \hat{x} = p\delta\gamma / (1 - \delta/2)(1 - p)$. Thus, a low x in a high transparency region would reinforce the choice of a 'stick and carrot' contract.

The principal chooses $a \geq 0$, $\omega \geq m + \gamma$ and $d \in \{0, 1\}$ to maximize $\pi = p(H - a) + (1 - p)L - \mu d x - \omega$, subject to providing the agent with the incentive to exert effort. The following proposition describes how the optimal contract depends on the precision of the public signal q, that is, on the transparency of production.

Proposition. If $x > \hat{x}$, then the optimal contract selected by the principal has the following properties:

1. the agent's basic wage is set at its lowest possible value: $\omega = m + \gamma$.
2. There exists a threshold $\hat{q} \in (0.5, 1)$ such that:

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

if $q > \hat{q}$ the optimal contract is a stick and carrot' contract: $d_s = 1$, and $a_s = \frac{\gamma}{p}\left(1 - \frac{pq\delta}{1 - \delta(p + q - 2pq)}\right)$;

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

The proof of this proposition is provided in Appendix 1.[10]

[9] In online Appendix C we consider the case where the dismissal probability is unrestricted: $d \in [0, 1]$.

[10] By Malthusian considerations, if farmers' expected income exceeds the subsistence level, we should

2.1.1 Discussion

We illustrate the results of this proposition in a graph (Figure 1) for a simple calibration. We set: $H = 1.1$, $L = 0.6$ and $p = 0.8$, so that a bad harvest with a significantly lower crop occurs about once every five years, and the expected crop size of each plot is set to one: $E(Y) = pH + (1 - p)L = 1$.[11] To be consistent with tenants' output share of about two thirds and with the relatively high cost of maintaining a family throughout the year, we set the subsistence cost to $m = 0.5$ and the effort cost to $\gamma = 0.1$, thus making the basic wage $\omega = 0.6$. Given an interest rate (in grain) of one third, as was customary 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.[12]

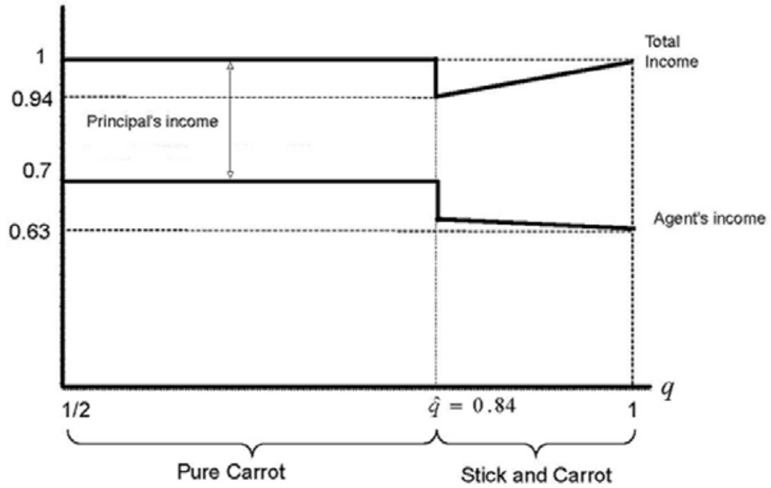

Figure 1. Periodic expected income as a function of signal accuracy

In this figure, the agent's expected income I 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 principal's expected income. The

expect the farming population to grow. In Online Appendix F we close the model as far as population size is concerned by assuming that any excess workers from the rural sector, including dismissed agents, are employed outside of farming, where the wage is low (particularly during famines) and does not guarantee reproduction.

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

[12] With these parameters $\hat{q} > 0.5$ is achieved already with $x = 0.48$. However, in the version of the model in which the dismissal probability is continuous, (online Appendix D), a higher x is required for obtaining a range of $\hat{q} > 0.5$ in which $d = 0$ is optimal. Thus, for consistency, we set $x = 2$.

figure clearly identifies the two regimes: 'pure-carrot' and 'stick-and-carrot,' and the switch between them at the critical transparency level $\hat{q}$.

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 (since expected output is 1) and the expected income of both the principal and the agent is independent of q. In 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: $p(H - L) + L$.

In contrast, in the stick-and-carrot regime, when $q > \hat{q}$:

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

and the expected total income is:

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

The expected total income reveals that the stick-and-carrot contract is socially inefficient. This is because it entails an expected loss of μx, since the agent may be dismissed even though he works diligently. The efficiency loss μx declines as accuracy improves, and at the limit, when the signal is accurate ($q = 1$), 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 in the latter range 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 discretely by the expected cost of dismissal $\mu(\hat{q})x$. Beyond that threshold, his expected per-period income declines with q.

Comparing the outcome when the signal fully discloses the state of nature ($q = 1$) with the outcome when the signal is highly inaccurate ($q < \hat{q}$) is revealing. In both cases the diligent agent is never dismissed and the economy is efficient. However, the distribution of income is quite different. The agent's income above subsistence falls from $I_c - (m = \gamma) = \gamma$ ($= 0.1$ in the example) in the range of the opaque signal to $I_s - (m = \gamma) = \gamma - p\delta\gamma/[1 - \delta(1 - p)]$ ($= 0.03$) when $q = 1$, as the bonus that is required to dissuade the agent from shirking is reduced to a minimum.[13]

[13]When the agent is very patient (δ =1), his utility from employment in agriculture dissipates entirely.

These results confirm that when transparency is sufficiently low, the agent-tenant is never dismissed and could be considered a de facto owner of the land that he cultivates. In contrast, when transparency is sufficiently high, the farmer may be evicted and thus cannot be considered to have ownership rights to the land. In this range, an increase in transparency implies that the probability of (wrongful) dismissal (μ) declines, so that expected cost of including a stick in the contract decreases. This enables the principal to rely more on the stick of dismissal and less on the carrot of bonus payments. This entails a correspondingly smaller share of output for the tenant and an increase in the revenue appropriated by the state. The effect of increased transparency on the optimal combination of the stick and carrot is robust and does not depend on our specific modeling assumptions. It reflects the logic that the credible threat of using a stick reduces the cost of incentivizing the agent with a carrot.

One may argue that the principal has an incentive to renege, by avoiding paying the bonus to the agent, or, alternatively, by failing to dismiss the agent when this is called for, and negotiate ex-post in order to avoid the cost of dismissal. We have in mind, though, a patient principal who faces many agents, repeatedly. Informally, we envision that the agents are likely to believe that if the principal reneges once, even if on one agent only, she is likely to continue doing so in the future and on other agents. Under these circumstances, the principal's commitment to the contract becomes credible, since reneging would render her unable to incentivize agents by using the carrot or the stick.[14]

2.2 A Two-Level Hierarchy Model

This subsection provides a key extension of the basic model by incorporating both landlord-farmer relationships and those between the ruler and intermediary officials. This extension has important implications for the study of state concentration and the power structure between the center and the periphery. Our main conclusion is that if farming activity is sufficiently transparent, not only to the local elite but also to the center, the intermediary retains a smaller share of the revenue and is subject to dismissal. If, however, farming activity is sufficiently opaque to the king, the local official (governor) retains autonomy and

[14] We address the concern over inefficient costly dismissal further in online Appendix D, where we consider a more complex contract-form in which the principal warns the agent a number of times (determined endogenously) upon suspected shirking, before final dismissal. The qualitative results of the model regarding the effect of transparency q on the optimal contract are unchanged in both extensions. A related concern is that the contractual relationship could terminate for reasons other than dismissal upon suspected shirking. But as long as contracts last for more than one season this should not change the qualitative results of our model. One could extend it by adding a parameter for an exogenous probability of separation: the higher that probability, the less effective the stick of dismissal.

a larger share of revenue. We also introduce here an alternative depiction of the moral hazard problem that is more likely to apply to the relations between local elite and the center: misreporting tax collection.[15]

We assume now a two-tier case, where each plot is located within a district, and where officials at the district level mediate between the tenant-farmer and the king. This two-tier case can easily be extended to add more tiers. We attach subscripts 1 or 2 to the variables at each level of the hierarchy, from the bottom up.

Two independent state variables are assumed to 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 district specific. The plot specific state can be thought of as injury to the tenant during the critical harvest time, or damage due to localized flood or fire. The district specific state would be something affecting the entire district, such as widespread drought or blight. We denote the probability that each plot of land is in a plot-specific good state by $p_1 \in (0,1)$, and the corresponding probability for each district by $p_2 \in (0,1)$. We assume that the plot-specific states are independent across plots within a district, and independent also of the district state. 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 district's states of nature are good ($\theta_1 = \theta_2 = G$), which pertains with probability $p_1 p_2$.

The district 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 district specific state is good ($\theta_2 = G$), then the governor receives plot-specific signals $\sigma_1 \in \{g, b\}$ for each plot in the district. These signals are accurate with probability $q_1 \in [0.5, 1]$ and are (conditionally) independent across plots. The relations between the district governor and the farmers under her control are just as in the basic model. The contract selected by the governor will thus have the same structure as before: it specifies a basic wage $\omega_1 = m + \gamma$, a bonus a_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 district's state and the plot's signal are good ($\theta_2 = G, \sigma_1 = g$). Thus, subject to the farmer exerting high effort, he is dismissed with probability: $\mu_1 d_1 = (1 - p_1) p_2 (1 - q_1) d_1$. That is, 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 upon high effort, and the probability of dismissal is $\mu_1 d_1$ instead of μd. Thus, the governor chooses a pure-carrot contract ($d_1 = 0$) if transparency is below

[15] Our theory of state hierarchy is consistent with Olson's (1993) non-functional approach: hierarchy serves as part of a uni-directional extraction mechanism, and does not serve in the management of downstream activities, as is customarily depicted.

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.[16]

We also assume that the number of plots in each district, N_1, is sufficiently large so that the total revenue obtained by the district governor can be substituted by its expected value. The governor's revenue is then limited to two possible outcomes, depending on the district-specific state of nature θ_2: L_2 in a bad year ($\theta_2 = B$) and H_2 in a good year ($\theta_2 = G$), where:

$$\begin{aligned} L_2 &= N_1 \left[L_1 - (m + \gamma) \right], \\ H_2 &= H_2(q_1) = N_1 \left[p_1(H_1 - L_1 - a_1) + L_1 - (1 - p_1)(1 - q_1)d_1 x_1 - (m + \gamma) \right]. \end{aligned}$$

The parameters a_1 and d_1 are those selected by the governor (as a function of q_1). As in the basic model, beyond a threshold $\hat{q}_1$, the good-year revenue H_2 is increasing in q_1.

For the relations between the king and the district governor, we employ a variant of our basic model in which, instead of possibly exerting low effort, the governor may hide some output in good years and report to the king L_2 instead of H_2. At this level of the hierarchy we assume an information structure analogous to the one in the basic model. The king does not know the specific states θ_2 for any of the districts, but he receives an independent signal $\sigma_2 \in \{g, b\}$ about each of the district states, whose accuracy is denoted by the probability $q_2 \in [0.5, 1]$.

The king is assumed to employ a two-edged incentive scheme analogous to the one above: a bonus a_2 if the governor reports collecting H_2, and 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 the district harvest is expected to be high. The king thus chooses $a_2 \geq 0$ and $d_2 \in \{0, 1\}$ to maximize:

$$\pi_2 = \max_{a_2 \geq 0, d_2 \in \{0,1\}} p_2(H_2 - a_2) + (1 - p_2)[L_2 - (1 - q_2)d_2 x_2].$$

subject to providing the governor with the incentive to report truthfully.

The details of the solution to this problem are very similar to the solution to the basic model, and are reported in Appendix 2. Once again, the balance of power between the king and the district governor depends on the transparency of the district economy to the king, q_2. When local conditions are sufficiently opaque to the king, the intermediary governor enjoys substantial autonomy in that she retains a (relatively large) fixed part of her collected revenue and always keeps her position. But if the transparency of the local provincial economy to

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

the king is sufficiently high, then the governor is subject to dismissal and retains a relatively lower share of the revenue collected.

3 APPLICATION: THE MAJOR CIVILIZATIONS OF THE ANCIENT NEAR EAST

Our theory provides the following predictions that link transparency to institutions. According to the basic model:

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

(2) When farming conditions are more transparent, the state's capacity to tax is higher and the inequality between the elite and the farming population is greater.

And, according to the hierarchical extension of the model:

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

In this section we demonstrate that these three predictions are consistent with the institutions that prevailed in the three major civilizations of the ancient Near East during the late fourth to the second millennia BCE: Northern (upper) Mesopotamia, Southern Mesopotamia (Babylonia, Sumer), and Egypt. Naturally, we do not pretend that our simple theory can explain all the institutional differences between these civilizations, nor that our theory is the only one that addresses these differences.

These three civilizations were listed above in accordance with their age, but are reviewed below in reverse order. Intensive agriculture was first adopted in the highlands of southern Anatolia and northern Mesopotamia in the seventh millennium BCE. Agriculture was adopted in the alluvial planes of Southern Mesopotamia and in the Nile Valley only two and three millennia later, respectively. It was in Sumer, however, that the first major city-states were formed in the fourth millennium (Liverani 2006). But the first central territorial state was formed in Egypt, in about 3000 BCE, starting from a core in Upper (southern) Egypt (Kemp 2006). The rapidity of the formation of a central state, and its subsequent stability, are among the key features that distinguish between ancient Egypt and Southern Mesopotamia, leading Baines and Yoffee (1998, 268) to conclude "the two civilizations are profoundly different."

Trigger (2003) and other scholars note multiple distinguishing features between these ancient civilizations. One of them is land tenure arrangements. In Egypt the land nominally belonged to the King, and in southern Mesopotamia land was typically owned by the temples and the urban elite. In both regions land was thus cultivated by tenants; but in Northern Mesopotamia land was mostly owner-cultivated. Another major distinguishing feature concerns the role of cities. Fortified city-states existed in pre-dynastic Egypt, but Egyptian cities ceased to be fortified after the formation of the central state and played a limited role as administrative centers. This led Wilson (1960) to characterize Ancient Egypt as "a civilization without cities." In contrast, for most of the time up to the first millennium BCE, the alluvial plains of southern Mesopotamia were ruled by rival, fortified city-states that retained their independence and resisted repeated attempts to unify Mesopotamia under a central state. This led Adams (1981) to characterize southern Mesopotamia as "the Heartland of Cities." The highlands of Northern Mesopotamia on the other hand gave rise to more limited city-states.

We now review each of these three civilizations separately, and demonstrate how the geographical features and resulting transparency of agriculture in each region can account for their distinctive institutional characteristics. To summarize, we argue that ancient Egypt occupies a polar extreme, with farming that was highly transparent both at the local and the global levels. Northern Mesopotamia is closer to the other extreme, with low transparency at both the local and the global levels.[17] Southern Mesopotamia, we suggest, presents an intermediate case, being comparatively transparent at the local level, but quite opaque to any potential central state.

3.1 Egypt

The Nile flows northwards, receiving its water mainly from the early-summer monsoon rains in eastern Africa. As a result, it surges in summer, at which time it floods the narrow river valley in Egypt. The Egyptian basin irrigation system was based on lateral dikes across the river valley, constructed to retain the flood water. The trapped water soaked the land and deposited nutrients for about two months, before it was drained back to the Nile, in time for the sowing of the staple cereals (mostly barley). The moisture trapped in the soil was the sole source of water during the growing season. Harvest was in late March, before the hot winds could parch the grain stalks and cause the kernels to disperse. This form of farming within the Nile Valley originated at the southern tip of Upper Egypt in the fifth millennium BCE, from where the Egyptian central state subsequently emerged.[18] The homogeneity of

[17] Agriculture in Northern Mesopotamia was, however, significantly less opaque than in the more arid regions of the Ancient Near East. Noy-Meir (1973) demonstrates the extreme effects of spatial variations in micro-climate and terrain quality on the heterogeneity of desert plant populations.

[18] For brevity, we focus on the Nile Valley, thus avoiding the Nile delta and the Fayum depression. The basin irrigation system prevailed with surprisingly minor variations for about five millennia, until the construction of the first Aswan Damn in the early twentieth century. Willcocks (1899) and Butzer (1976) provide detailed descriptions of this system.

the land within each basin implied very high local transparency.

Since few details about tenancy arrangements in ancient Egypt have survived, historians often employ evidence from the more recent past. In describing district life in Egypt from the medieval period up to the nineteenth century, Baer (1969, 17) contends that it was characterized by three phenomena: (a) the village head periodically redistributed land among the peasants; (b) the village inhabitants were collectively responsible for tax payments; (c) the village as a whole was responsible for maintaining irrigation infrastructure and for providing labor for public works. Eyre (1997, 378) and Eyre (1999, 51-52) similarly maintains that in ancient Egypt, farmers did not have secure tenure and the village community as a whole was responsible for paying taxes. The village head exercised tight control over village land and 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.[19]

This description supports our prediction that the threat of dismissal (or relocation) of individual farmers was a widely used incentive device in Egypt and that land was not owned by the cultivating farmers. Indeed, the prevailing notion in ancient Egypt was that the entire land belonged to the Pharaoh (Baines and Yoffee 1998, 206), even if this coexisted in various periods with a practice by which much of the land was de facto owned by the temples, by various lay organizations, and by powerful individuals (Manning 2003, 65-98). That is, even when land in the Nile Valley was privately held, it was owned by absentee landlords who did not work the fields.[20]

This state of affairs is consistent with prediction (1). The high local transparency of farming eliminated the main disadvantage of absentee land ownership, and left peasants vulnerable by denying them an information advantage. Significantly, in the few known cases where private land lease documents survived from antiquity, the contracts were for one year only (Hughes 1952), providing further support for our proposed mechanism that tenants were constantly under the threat of eviction.[21]

[19] Eyre (1997, 1999) contends that the divorce between land-ownership and actual farming was endemic to Egypt and persisted until the mid-twentieth century. According to Baer (1969, 62-78), even the major agrarian reforms during the nineteenth century, which gave land title to the cultivating peasants, ended with much of the land reverting back to large absentee landlords after the small cultivators failed to pay the required taxes.

[20] Hughes (1952, 1-2) summarizes that in the first two millennia of the historic period (the third and second millennia BCE) there was never "a large body of small landholders who managed and worked their plots themselves . . . the lowest classes were largely serfs on the domains of Pharaoh, the wealthy and the temples."

[21] Another feature that reduced the advantages of long-term leases in the Nile Valley was that land fertility was sustained by the Nile's annual deposits, so that land could not in effect be over-exploited. In addition, agrarian capital investment was by way of dikes and local canals that were constructed and maintained

Transparency should not be confused with predictability. The fluctuations in the Nile's annual inundation level were substantial and caused significant unpredictable annual variations in crop output. Particularly high inundation would break the lateral dikes and flood villages, causing as much of a threat as very low inundation levels. The timing, length and severity of the hot spring winds at harvest time contributed to the uncertainty. However, in any given year, the conditions that farmers faced were fairly homogeneous within each irrigation basin system, and also across basin systems. As a result, farming activity was highly transparent not only locally, but also to the central government. The Nile's annual peak inundation was recorded as early as the third millennium BCE (Kemp 2006, 64). Nilometers that measured the inundation level were set up along the Nile, and it appears that the Pharaohs used this information as a control device. Cooper (1976, 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 the case also in antiquity.[22]

The Nile's global transparency enabled the Pharaohs to employ a stick-intensive 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), the high transparency of farming helps explains why the Pharaohs were able to run a lean state bureaucracy and to siphon off a substantial share of the tax revenue, without engaging in direct control. In turn, it explains why the provincial centers retained so little independent power. This is consistent with Eyre's (1994, 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." Indeed, at least in the early Old Kingdom period, the positions of governors and state bureaucrats were by a revocable appointment, and nonhereditary.[23] The non-secure status of these state bureaucrats is closely related to the relative weakness of the cities in the different districts. These cities essentially remained administrative centers, without amassing

communally. The homogeneity of farm land within each basin also reduced the importance of plot specific skills, implying that the cost of replacing an agent was not high, reinforcing the adoption of dismissal as part of the incentive scheme.

[22] The transparency of Egyptian farming was due also to the relative ease of monitoring farming activity in real time by inspectors traveling along the Nile. Kemp (2006, 254-6) provides evidence for such a monitoring expedition from 1140 BCE.

[23] Baines and Yoffee (1998, 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."

substantial independent wealth to threaten the predominance of the center.

The high transparency at all levels of the state hierarchy can also explain the rapidity of the formation of a strong central state in Egypt, and its remarkable subsequent stability.

3.2 Southern Mesopotamia

As in Egypt, farming in arid Southern Mesopotamia relied entirely on riverine irrigation. The water regime in the Tigris and the Euphrates, however, is very different from that in Egypt. Both these rivers flow southward, and are fed by the winter rains and by spring melting snow in the mountains of modern Turkey and Iran. The long distance between these mountain ranges and Southern Mesopotamia meant that water levels were low in October-December when irrigation was most needed, but high in the harvest season in late spring (Adams 1981, 3-6; Postgate 1994, 178; Wilkinson 2013). This mismatch prevented irrigation by flooding, as in Egypt. Cereals were cultivated on the outer slopes of levees, including the levees of abandoned courses of the rivers. An extended canals system was required to deal with the water shortage in the cultivation season: capturing water upstream and directing it towards the fields. Furthermore, since water quantity was insufficient to irrigate all the arable land, control mechanisms were required to distribute the water.[24] The swelling of the rivers in the spring posed another major threat of flooding the ripe fields at harvest time, and had to be overcome by diverting excess water into the marshy flood plain at the lower end of the cultivation zone (Adams 1981, 245; Wilkinson 2003, 89; Wilkinson 2013).[25]

These major problems apparently delayed the adoption of extensive agriculture in Southern Mesopotamia well after agriculture flourished in Northern Mesopotamia, and after irrigation systems were established in southwest Iran (Wilkinson 2003, 72-76). In addition to the intricate canal system that overcame these problems, agriculture in Southern Mesopotamia benefitted from another innovation: the cultivation in deep furrows, plowed by oxen, in narrow and long fields that sloped down from the feeding canal towards the marshy plain (Liverani 2006). This method enabled conservation of seed and water, and also helped divert the saline topsoil away from the plants.

Farming conditions in Southern Mesopotamia were complex (Wilkinson 2013). Even

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

[25] Unlike in Egypt, the soil nutrients were not replenished automatically and salt was not washed away. The need to replenish land fertility and the shortage of water combined to establish a system of relatively frequent land fallow.

fields within the same zone could vary in quality, depending on how high they were above the saline water table in the adjacent marsh. The overriding factor though was the dependency of cultivation on rationed water, which was controlled upstream, and which could have been directed elsewhere. Farmers were thus completely dependent on the local elite who controlled the flow of water at various canal junctures. In turn, the elaborate canal system provided the local elite with significant means of control and with information on the state of agriculture.[26]

Accordingly, we 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 sharecroppers, who were overseen by a hierarchy of intermediaries, under the ultimate control of dominant elite families who resided in the urban centers and controlled each city's temple (Renger 1995; Liverani 2006).[27] In accord with prediction (2), this high local transparency explains why powerful early city-states were able to form and to persist 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 late-fourth and third millennia BCE. Writing originated in about 3200- 3100 BCE in the largest of these cities, Uruk, when its population reached about twenty thousand (Yoffee 2005, 43).

At the same time, the complexity of the irrigation system required skilled local managers with a "thorough knowledge of local conditions on a day-to-day basis" (Hunt 1987, 172). Unlike the case of Egypt, the local managing elite in Southern Mesopotamia were thus indispensable and irreplaceable. In other words, we interpret farming activity in Southern Mesopotamia as rather opaque to any distant central government. Consistent with prediction (3), 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 Mesopotamian civilization. Thus, even when an early city-state in Southern Mesopotamia managed to conquer a competing city-state, it still needed the cooperation of the elite of the subjugated city to obtain on-going tax revenue from the conquered territory. It was the specific knowledge possessed by the local elites, we contend, that assured the autonomy of Southern Mesopotamian cities.

This helps explain why several aggressive attempts to unify Southern Mesopotamia under one of the rival city-states in the third and second millennia BCE ended in failure after a

[26] One may argue that such direct control provides power of coercion that goes beyond mere information. In our framework, however, the state is assumed to possess the power to coerce. Thus, implicitly, we view control as a form of transparency.

[27] As in Egypt, in addition to remitting farm output to the elite, the peasants were required to provide compulsory labor services (corvée) to repair and extend irrigation infrastructure, and also for temple lands.

relatively short period — in marked contrast 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 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 by 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.[28] Our explanation of this historic pattern is consistent with Yoffee's (2005) description of the fate of Sargon's earliest central state. According to Yoffee, Sargon was well aware of the intermediation problem. When he ascended to power he 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, 292-293, 2005, 143).

Furthermore, not only that the city-states in Southern Mesopotamia resisted subjugation to outside power, they also resisted local despots. These city-states were typically governed by hereditary kings by the rule of law, and with councils of elders and assemblies at their side, consisting of members of landed elite, merchants and artisans (Van de Mieroop 2013). We suggest that this pattern of governance prevailed since would-be despots were unable to raise sufficient revenue to sustain coercive power without the cooperation of the local elite who possessed specific information on the intricate countryside and the economy.[29]

3.3 Northern Mesopotamia

Farming started in Northern Mesopotamia long before it was adopted in Southern Mesopotamia. And urbanization was identified there already in the late fifth and early fourth

[28] The Neo-Assyrian Empire in the first millennium BCE devised various administrative methods to subject conquered states. In particular, they adopted bi-directional deportations, in which the elite of a conquered state were deported and replaced by people from elsewhere. But, significantly, even under the Neo-Assyrian, Neo-Babylonian and Persian empires, the elites in the cities in southern Mesopotamia retained much of their former autonomy (Van de Mieroop 1997, 128-139, 2013).

[29] Sinopoli et al. (2016, 390) identify an even more pronounced pattern of distributed power in the governance of city-states in antiquity in diverse regions of the world (including the Indus Valley, Greece and Western Africa). But rather than focus on what handicapped would-be despots, they attribute the pervasiveness of this pattern to "both ideological commitments and material benefits to the actors involved."

millennia BCE; but declined in the later part of the fourth millennium.[30] The geographic conditions in the highlands of Northern Mesopotamia are quite different from those in riverine Southern Mesopotamia and Egypt. Agriculture was mostly rain fed. Due to the uncertain and idiosyncratic nature of rainfall, and to the relative unevenness of the terrain, farming was comparatively opaque even at the local level.[31] Wilkinson (1994, 2003, 210) concludes that the settlement pattern in Northern Mesopotamia was characterized by a large scattering of roughly equivalent, nucleated units. Each unit was 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, 503).

Without disputing this observation, we take issue with Wilkinson's explanation that this pattern was due to the fact that no center was able to dominate another, since none had an "overwhelming situational or demographic advantage" (2003, 210). 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 territorial states under city leaders who happen to 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 lands. In a more pronounced version of the situation in Southern Mesopotamia, and consistently with our third prediction, we propose that the localized nature of the early city-states in this region was due to the opacity of farming activity that limited the span of control of its urban centers.[32]

The relatively low transparency of farming in Northern Mesopotamia, even at the local level, can also explain the drastically different land tenure regime in that region. In contrast to the tenancy pattern in Egypt and Southern Mesopotamia, owner-operated farming was prevalent in Northern Mesopotamia from early on. Cuneiform documents from the mid-second millennium BCE from Nuzi (near modern Mosul) reveal that while the local kings and the elite owned large estates, the temples did not possess economic power or land, and

[30] 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 consistent with the presumption of the elite's inability to raise the needed resources to secure the countryside from banditry, which forced the peasants to seek refuge within the walls of the central city, and with the relatively small territorial span of these early city-states.

[31] See Wilkinson (1994) and Jas (2000).

[32] The costly transport of the crop tribute to the center over land was another major contributing factor for the limited span of early potential states in Northern Mesopotamia, in comparison with riverine Southern Mesopotamia and Egypt, where in-kind tax revenue was transported by rafts and boats.

much land was owned by nuclear families who worked their patrimonial property. The Nuzi evidence also reveals that land ownership in Northern Mesopotamia was in a constant state of flux. Small landholders regularly lost title of their land to rich families through debt and sale under duress (Zaccagnini 1999; Jas 2000). However, the persistence of owner-occupied farming indicates that the process of land consolidation must have been matched by an opposing process by which large, presumably less efficient, estates were gradually dissolved. The prevalence of owner-cultivated private farming in Northern Mesopotamia is consistent with prediction (1) that low transparency makes tenancy less profitable to absentee owners.[33]

4 RELATED LITERATURE

Since the body of related literature is large, we shall review only the leading alternative theories on the pattern of state governance in the Ancient Near East and some related general theories on statehood.[34]

We start with Wittfogel's (1957) influential hydraulic theory of "oriental despotism," according to which large-scale irrigation infrastructure was necessary to realize the agricultural potential in riverine environments. Strong, despotic states are presumed to have been a prerequisite for constructing and administering these irrigation projects. Wittfogel's many critics pointed out, however, that the irrigation systems in ancient Egypt and Mesopotamia (and elsewhere) were constructed communally, prior to the emergence of a strong central state. Moreover, even after a central state emerged, these irrigation systems were managed locally, rather than from the center. Due to the cogency of these counter-arguments, Wittfogel's theory is now considered defunct. But this leaves unexplained the correlation that he pointed out between riverine environments and strong ancient states. Our theory explains this correlation by reversing the direction of causality in Wittfogel's theory. It is not that a despotic state was required to construct and to operate irrigation systems, but rather that irrigation-based agriculture provided transparency and facilitated state control.[35]

[33] Jas (2000) quotes Warriner (1948, 21, 104), who noted that the different land tenure regimes between Northern and Southern Mesopotamia in antiquity 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."

[34] With regard to the related literature on property rights, we only note that in our framework rights to land do not arise spontaneously (as in Demsetz 1967), but are granted by an authoritarian government (as in North 1981).

[35] Billman (2002) provides additional evidence from an early irrigation system in the Moche valley in the arid northern coast of Peru in 400 BCE-800 CE. He argues that the use of irrigation created an opportunity for leaders "to control land and the flow of water;" thus enabling them "to finance the creation of centralized,

An alternative functional theory posits that the early state served a redistributive purpose. Thus, Adams (1981, 244) views the Mesopotamian city-states as having been formed to cope with uncertainty in farming output, through precautionary storage against years of shortage: "In the largest sense, Mesopotamian cities can be viewed as an adaptation to the perennial problem of periodic, unpredictable shortages. They provided concentration points for the storage of surpluses." Our framework, however, suggests that the attested extensive inter-annual storage in ancient Egypt and Mesopotamia may have served primarily to protect the urban elite against revenue shortfalls in years of famine, rather than to aid the farming population in the countryside.

Probably the most widespread alternative theory is the idea (referred to in our introduction) that the emergence of the state was due to the increased productivity of agriculture and to the surplus that was thus created. It is argued that this surplus generated via diverse channels (such as population pressure, trade, or required protection) a need for government. In an influential variant of these ideas, Carneiro's (1970) "environmental circumscription" theory poses that states emerged only in circumscribed areas that trapped the agrarian population and restricted its ability to avoid subjugation by fleeing elsewhere. In his comprehensive study of the history of government, Mann (1986, 38-40, 75-102, 108-115) uses the metaphor of a "social cage" to explain the success of ancient Egypt and of other early states. Both Mann (1986) and Allen (1997) argue that Egypt's success was due to the deserts that isolate the Nile Valley and inhibited the peasants from avoiding subjugation via out-migration, thus enabling the state to extract surplus from the farming sector. From our perspective, while this entrapment theory fits Egypt, it does not adequately explain the institutional differences that were examined above, for example, between northern and southern Mesopotamia.

Tilly (1975) offered another influential conflict theory that relates the emergence of centralized states to their capacity to tax. In seeking to explain the evolution of European states since the middle-ages, he contends that new military technologies disrupted the international equilibrium and forced states to consolidate in order to finance ever costlier wars. Tilly famously stated: "War made the state, and the state made war" (Tilly 1975, 42). Finer (1997) applied this warfare theory more broadly to explain the history of government, referring to this positive feedback theory as the "extraction–coercion cycle" (Finer 1997, 15-19).[36]

It is evident however that warfare theory cannot explain the state's success in ancient Egypt, since, as argued for example by Dal Bó, et al. (2015), Egypt's natural circumscription insulated it from the outside and implied that once a central state was formed, it was not

hierarchical political organizations;" leading to the formation of an early territorial state.

[36] Gennaioli and Voth (2015) test Tilly's theory to argue that tax capacity indeed increased since the Middle Ages due to the necessity of financing wars.

seriously threatened by competing states. This was in contrast to Mesopotamia, where local rivalries and nomadic banditry were a perennial problem. Stasavage (2010) proposed an alternative transparency theory – where the transparency is of government activity, rather than of production – that challenges Tilly's theory on other grounds. He contends that the compact geographic span of small pre-modern European city-states generated greater transparency of their governments and enabled these city-states to raise the necessary resources, via taxes or credit, and thus withstand aggression and retain their independence.

As we see it, the critical element missing in the warfare theory is an explanation of what enabled a victor to extract ongoing revenue from a conquered territory to make the conquest viable and long-lasting. In other words, while admitting that fiscal capacity contributes to a state's military capacity, we question the general validity of the reverse causal relation that greater military capacity necessarily increases fiscal capacity. These considerations, though, highlight that we assume here, as mentioned in the introduction, an isolated state with an absolute power to coerce; yet without incorporating rivalry between competing polities or taking into account the resources required to maintain such power and to deter secession.[37]

Moreover, we have avoided altogether assigning the state with the typical function of providing public goods, including the maintenance of law and order. We acknowledge in this respect the contribution of Levi (1988), who considers, as we do, the constraints on the government's capacity to tax, but emphasizes how the provision of public goods and adherence to constitutional constraints may foster cooperative compliance by tax-payers, as a substitute for the sole reliance on coercion.

Finally, we note that in a companion contribution (Mayshar et al., 2016) we consider both the emergence of the state and the role of increased productivity, issues that we do not address here. In that paper we focus on a different aspect of the ability to appropriate, contending that the transforming feature of the Neolithic Revolution which gave rise to social hierarchy was the increased appropriability of crops, rather than increased productivity. In particular, we argue and provide empirical support for our claims that even after the adoption of highly productive agriculture, state institutions did not emerge in regions where farming relied on non-seasonal roots and tubers that are typically perishable and largely non-appropriable. Complex hierarchies and state institutions emerged only in regions of the world, such as the Ancient Near East, where farming relied on seasonal and non-perishable cereal crops, since such crops require storage from one harvest to the next and are thus highly vulnerable to appropriation.[38]

[37] The literature on these issues is extensive. See most recently Boix (2015).

[38] Huning and Wahl (2016) provide additional evidence in support of our current transparency theory, and also for our additional claim about the secondary role of productivity. They extend our present model by

5 CONCLUSION

Stigler (1961) stated that "knowledge is power." We apply this maxim to examine how the extent and the structure of informational asymmetry shaped the institutions of pre-modern, agricultural state societies. Our overarching contention is that through its effect on the tax technology, the transparency of production affects the scale of the state, its hierarchical structure, and land tenure practices. This theory helps explains why ancient Egypt was rapidly united and was subsequently very stable and highly centralized, while Sumer remained a complex of competing city-states for several millennia. It also explains why land in Egypt belonged (at least nominally) to the Pharaoh, while in Southern Mesopotamia land belonged to the temples and to the elite, and in Northern Mesopotamia there was substantial owner-occupied farming.

Our environmental theory of early institutions offers a new paradigm for understanding antiquity, with an emphasis on how differences in the extent of information asymmetry affect hierarchical extractive institutions. While we apply our theory to the institutions of antiquity, we propose that it can be applied to all predominantly agricultural state societies. More generally, and unrelated to environmental considerations, our theory sheds light on how production technologies can impact the state's capacity to tax and shape institutions. In particular, whereas the prevailing perception is that asymmetry of information hinders efficiency, our framework reveals that the lack of transparency of agents' activities ('privacy') may in fact be beneficial to them in protecting their freedom, and possibly also in promoting their material well-being.

---

viewing spatial homogeneity of soil quality as a proxy for transparency, and relating it to state size, under the assumption that states' income determines military spending. Consistent with our theory, they find a robust positive relationship between observability and size of medieval German territories. They also find that low observability is correlated with the existence of city-states, and show that the resulting political fragmentation in the medieval period is recognizable in Germany until today. Moreover, they show that land productivity does not have a significant effect on the territorial size of these city-states.

References

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

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

[3] Allen, Robert C. 1997. "Agriculture and the Origins of the State in Ancient Egypt." *Explorations in Economic History* 34 (2): 135-154.

[4] Baer, Gabriel. 1969. *Studies in the Social History of Modern Egypt.* Chicago: University of Chicago Press.

[5] Baines, John and Norman Yoffee. 1998. "Order, Legitimacy, and Wealth in Ancient Egypt and Mesopotamia." In *Archaic States,* eds. Gary M. Feinman and Joyce Marcus. Advanced Seminar Series. Santa Fe: School of American Research Press, 199-260.

[6] Banerjee, Abhijit V. and Ghatak, Maitreesh. 2004. "Eviction Threats and Investment Incentives." *Journal of Development Economics* 74 (2): 469-488.

[7] Billman, Brian R. 2002. "Irrigation and the Origins of the Southern Moche State on the North Coast of Peru." *Latin American Antiquity* 13 (4): 371-400.

[8] Bockstette, Valerie, Areendam Chanda, and Louis Putterman. 2002. "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth* 7(4): 347-69.

[9] Boix, Carles. 2015. *Political Order and Inequality: Their Foundations and Their Consequences for Human Welfare.* Cambridge: Cambridge University Press.

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

[11] Carneiro, Robert L. 1970. "A Theory of the Origin of the State." *Science,* 169(3947): 733-738.

[12] Chwe, Michael Suk-Young. 1990. "Why Were Workers Whipped? Pain in a Principal-Agent Model." *The Economic Journal* 100(403): 1109-1121.

[13] Cooper, Richard S. 1976. "The Assessment and Collection of Kharāj Tax in Medieval Egypt." *Journal of the American Oriental Society* 96(3): 365-382.

[14] Dal Bó, Ernesto, Pablo Hernández, and Sebastián Mazzuca. 2015. "The Paradox of Civilization: Pre-Institutional Sources of Security and Prosperity." NBER Working paper 21829.

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

[16] Dari-Mattiacci, Giuseppe. 2013. "Slavery and Information." *Journal of Economic History* 73(1): 79-116.

[17] De la Sierra, Raul Sanchez. 2016. "On the Origin of States: Stationary Bandits and Taxation in Eastern Congo." Working paper, University of California Berkeley.

[18] Demsetz, Harold. 1967. "Toward a Theory of Property Rights." *The American Economic Review,* Papers and Proceedings of the Seventy-ninth

[19] Annual Meeting of the American Economic Association 57(2): 347-359.

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

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

[22] Eyre, Christopher J. 1997. "Peasants and 'Modern' Leasing Strategies in Ancient Egypt." *Journal of the Economic and Social History of the Orient* 40(4): 367-390.

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

[24] Finer, Samuel E. 1975. "State and Nation-Building in Europe: The Role of the Military." In *The Formation of National States in Western Europe*, ed. Charles Tilly. Princeton: Princeton University Press.

[25] Gennaioli, Nicola and Hans-Joachim Voth. 2015. "State Capacity and War." *Review of Economic Studies* 82(4): 1409-1448.

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

[27] Hughes, George Robert. 1952. *Saite Demotic Land Leases.* Chicago: University of Chicago Press.

[28] Huning, Thilo R. and Fabian Wahl .2016. "You Reap What You Know: Observability of Soil Quality, and Political Fragmentation." *EHES working papers in economic history.* 101.

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

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

[31] Kau, James B. and Paul H. Rubin. 1981. "The size of Government." *Public Choice* 37(2): 261-274.

[32] Kemp, Barry J. 2006. *Ancient Egypt: Anatomy of a Civilization*, Second edition, Abingdon: Routledge.

[33] Kleven, Henrik Jacobsen, Claus Thustrup Kreiner and Emmanuel Saez. 2016. "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries." *Economica* 83(30): 219-246.

[34] Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley: University of California Press.

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

[36] 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.

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

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

[39] Mayshar, Joram. 1991. "Taxation with Costly Administration." *Scandinavian Journal of Economics* 93(1): 75-88.

[40] Mayshar, Joram, Omer Moav, Zvika Neeman, and Luigi Pascali. 2015. "Cereals, Appropriability and Hierarchy," CEPR Discussion Paper 10742.

[41] North, Douglass. C. 1981. *Structure and Change in Economic History*. New York: W.W. Norton & Co.

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

[43] Olson, Mancur. 1993. "Dictatorship, Democracy, and Development." *American Political Science Review* 87(3): 567-576.

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

[45] 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.

[46] Shapiro, Carl and Joseph E. Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *The American Economic Review* 74(3): 433-444.

[47] Sinopoli, Carla M., Roderick J. McIntosh, Ian Morris and Alex R. Knodell. 2015. "The distribution of power in early cities" In *World History,* vol. 3: *Early cities in comparative perspective*, ed. Norman Yoffee. Cambridge University Press, 381-393.

[48] Sng, Tuan-Hwee. 2014. "Size and Dynastic Decline: The Principal-Agent Problem in Late Imperial China, 1700–1850." *Explorations in Economic History* 54: 107-127.

[49] Sng, Tuan-Hwee and Chiaki Moriguchi. 2014. "Asia's little divergence: state capacity in China and Japan before 1850." *Journal of Economic Growth* 19(4): 439-470.

[50] Spolaore, Enrico and Romain Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature* 51(2): 325-369.

[51] Stasavage, David. 2010. "When Distance Mattered: Geographic Scale and the Development of European Representative Assemblies." *American Political Science Review* 104(4): 625-643.

[52] Stigler, George. 1961. "The Economics of Information." *Journal of Political Economy* 69(3): 213-225.

[53] Tilly, Charles. 1975. "Reflections on the History of European State-Making." In *The Formation of National States in Western Europe*, ed Charles Tilly. Princeton: Princeton University Press.

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

[55] Ur, Jason A. 2010. "Cycles of Civilization in Northern Mesopotamia, 4400–2000 BC." *Journal of Archaeological Research* 18(4): 387-431.

[56] Van De Mieroop, Marc. 2013 "Democracy and the Rule of Law, the Assembly, and the First Law Code." In *The Sumerian World*, ed. Harriet Crawford. Abingdon: Routledge, 277-289.

[57] Warriner, Doreen. 1948. *Land and Poverty in the Middle East.* New York: Royal Institute of International Affairs.

[58] Wilkinson, Tony J. 1994. "The Structure and Dynamics of Dry-Farming States in Upper Mesopotamia." *Current Anthropology* 35(5): 483-520.

[59] Wilkinson, Tony J. 2003. *Archaeological Landscapes of the Near East.* Tucson: The University of Arizona Press.

[60] Wilkinson, Tony J. 2013. "Hydraulic Landscapes and Irrigation Systems of Sumer." In *The Sumerian World*, ed. Harriet Crawford. Abingdon: Routledge, 33-54.

[61] Willcocks, William. 1899. *Egyptian Irrigation.* Second edition. London: E. and F. N. Spon.

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

[63] Wittfogel, Karl A. 1957. *Oriental Despotism: A Comparative Study of Total Power.* New Haven: Yale University Press.

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

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

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

Appendix 1 - Proof of the Proposition

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

$$V = [\omega + pa - m - \gamma] + [1 - d\mu]\delta V, \tag{A1}$$

where $\mu = (1-p)(1-q)$ is the probability of a bad harvest and a good signal, and $d\mu$ is the probability of dismissal. Solving From (A1):

$$V(a,d) = \frac{\omega + pa - m - \gamma}{1 - \delta(1 - d\mu)}. \tag{A2}$$

The principal selects $a \geq 0$, $\omega \geq m + \gamma$ and $d \in \{0, 1\}$ to maximize:

$$\pi = \max p(H - a) + (1-p)L - \mu dx - \omega, \tag{A3}$$

subject to incentivizing the agent to exert high effort:

$$\begin{aligned} &p[a + \delta V] + (1-p)[q + (1-q)(1-d)]\delta V + \omega - m - \gamma \geq \\ &p[q(1-d) + (1-q)]\delta V + (1-p)[q + (1-q)(1-d)]\delta V + \omega - m, \end{aligned} \tag{A4}$$

where $V = V(a,d)$.

Since ω cancels out from (A4), it is optimally set to $\omega = m + \gamma$, thus confirming (1). Plugging (A2) into (A4) and simplifying yields the incentive constraint:

$$pa\left(1 + \frac{pqd\delta}{1 - \delta(1 - d\mu)}\right) \geq \gamma. \tag{A5}$$

Part (2) follows from the maximization of (A1) subject to (A5). Because the Principal sets a 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] to the quadratic equation $\pi(a_c, 0) = \pi(a_s, 1)$, that can be expressed as:

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

To see that $\hat{q} > 0.5$ if $x > \hat{x} = p\delta\gamma/(1-\delta/2)(1-p)$, note that while the left-hand-side of (A6) is convex and increasing from zero to infinity as q increases from zero to one, the right-hand-side is positive and linear in q. The threshold $\hat{x}$ is obtained by requiring that for $\hat{q} = 0.5$, the right-hand-side (A6) be equal to one.

Finally, the third pure strategy of dismissal of the agent upon observing 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 $x > \hat{x}$. □

Appendix 2

The incentive constraint for the governor is:

$$a_2 \geq (H_2 - L_2) - q_2 d_2 \delta_2 V_2,$$

where $\delta_2 V_2$ is the governor's discounted value of 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 (A1):

$$V_2 = p_2 a_2 + [1 - d_2(1-p_2)(1-q_2)]\delta_2 V_2, \tag{B1}$$

From (B1) it is possible to solve for $V_2(a_2, d_2)$ as in (A2), and then to solve explicitly for the king's optimal incentive scheme a_2 and d_2. Thus, subject to parameter restrictions on x_2 and δ_2 that are analogous to those above, there exists a threshold $\hat{q}_2 > 0.5$ such that if district 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 king's per-period revenue is $\pi_{2c} = L_2$, independently of the state of nature, and the governor retains $a_{2c} = H_2 - L_2$ whenever the district state of nature is good, and zero otherwise.

On the other hand, when district 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, on the basis of observing $\sigma_2 = g$, but the governor reports low revenue. This occurs with probability $(1 - p_2)(1 - q_2)$. In this regime, following a similar derivation to the one in the above, $d_{2s} = 1$ and $a_{2s} = (H_2 - L_2) - q_2\delta_2 V_{2s}$, where:

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

The king's expected revenue in this case 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.

For Online Publication

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

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

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 a) and so:

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

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

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

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

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

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

The principal's objective is:

$$\pi = \max_{a,d\in\{0,1\}} p(H-L) + L - pa - \mu dx - m, \tag{A5}$$

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:

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

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

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

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

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

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 $\omega + a$, where $a \geq 0$ 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:

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

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 a and d and the parameters: μ, p and δ as follows:

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

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

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

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

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

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

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

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

s.t.

$$pa + pqd\delta V(a,d) \geq \gamma. \tag{\textit{B-IC}}$$

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: 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:

$$\min_{d \in [0,1], a} pa + \mu x d, \tag{C1}$$

subject to the agent's incentive constraint:

$$pa = \left(1 + \frac{pqd\delta}{1 - \delta(1 - d\mu)}\right) \geq \gamma. \tag{C2}$$

The agent's incentive constraint must be binding in the optimal solution. Plugging the value of a from (C2) into (C1) yields the principal's objective function

$$\gamma \left(1 - \frac{\delta pqd}{1 - \delta + \delta d \left(\mu + pq\right)}\right) + \mu x d. \tag{C3}$$

as a function of d alone.

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

$$-\frac{\gamma \delta q p (1 - \delta)}{A^2} + \mu x \tag{C4}$$

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

Inspection of (C3) reveals that the expression on the left of (C3) 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

$$q \leq \frac{(1-p)x(1-\delta)}{\delta\gamma + (1-p)x(1-\delta)} \tag{C5}$$

then the value of d that maximizes the principal's objective function (sets the derivative (C4) 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 $C4 = 0$) is given by:

$$d = \frac{1-\delta}{\delta(\mu + pq)} \left(\sqrt{\frac{\gamma\delta pq}{(1-\delta)\mu x}} - 1 \right) \tag{C6}$$

If the right-hand-side of (C5) is larger than .5 or, equivalently,

$$\frac{(1-p)x}{\gamma} > \frac{\delta}{1-\delta} \tag{C7}$$

then $\underline{q} > .5$, which means that the pure carrot contract is optimal for some values of the accuracy parameter q. Inspection of (C7) 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:

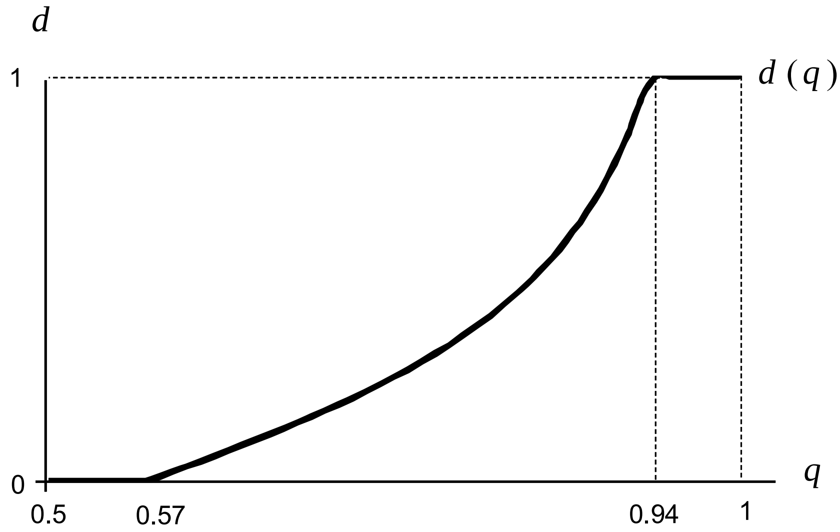

Figure 5: The optimal dismissal probability, $d \in [0, 1]$, as a function of transparency q

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 D: 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 $\sigma = g$ 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 $\sigma=g$. The agent is dismissed immediately upon $n=0$ and so $V(0)=0$. Let $a(n)$ 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:

$$V(n)=pa(n)+\mu\delta V(n-1)+(1-\mu)\delta V(n). \tag{D1}$$

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

$$pa(n)=\gamma-pq\delta(V(n)-V(n-1)). \tag{D2}$$

By combining (D1) and (D2) we obtain the following recursive formulation for $V(n)$:

$$V(n)=A+BV(n-1), \tag{D3}$$

where the constants A and B are given by:

$$A=\frac{\gamma}{1-\delta+\delta(\mu+pq)};B=\frac{\delta(\mu+pq)}{1-\delta+\delta(\mu+pq)}. \tag{D4}$$

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:

$$V(n)=\frac{A(1-B^n)}{1-B}. \tag{D5}$$

It therefore follows that:

$$a(n)=\gamma/p-q\delta AB^{n-1}. \tag{D6}$$

Observe that the bonus payments to the agent increase with n. It can be immediately verified that $a(1)$ and $V(1)$ are identical to a_s and V_s of the basic model, while a_c and V_c coincide to the limits of $a(n)$ and $V(n)$ from (D6) and (D5), 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 $k=1$:

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

And for $1 < k \leq n$:

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

These two equations simplify to:

$$c(1,n) = \alpha a(1) + \beta x/\delta_P + \beta c(n,n), \tag{D7}$$

and

$$c(k,n) = \alpha a(k) + \beta c(k-1,n), \tag{D8}$$

where the two constants α and β are given by:

$$\alpha = \frac{p}{1-\delta_P+\mu\delta_P}; \quad \beta = \frac{\mu\delta_P}{1-\delta 2+\mu\delta_P}. \tag{D9}$$

Equations (D7) and (D8) can be explicitly solved for $c(n,n)$ as a function of the underlying parameters of the model as follows:

$$c(n,n) = \frac{\gamma}{1-\delta_P} + \frac{\beta^n x}{\delta_P(1-\beta^n)} + \frac{\alpha q\delta A(B^n-\beta^n)}{(1-\beta^n)(B-\beta)}. \tag{D10}$$

It is reassuring to confirm that the solution of the equation $c(1,1) = c(\infty, \infty)$ for q yields the threshold $\hat{q}$ 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 (D10)) 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]

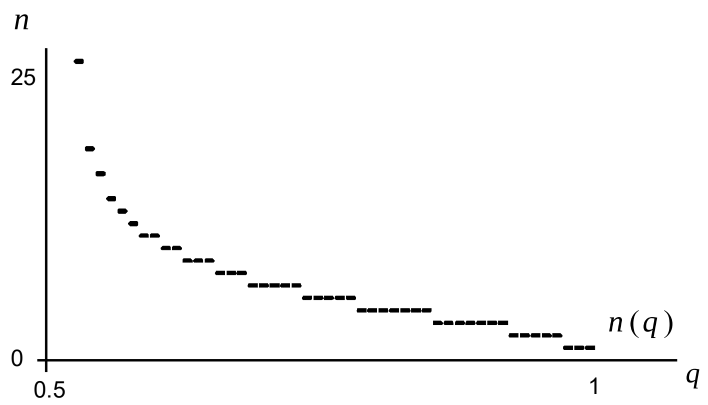

Figure 6: The optimal number of "bad signals" before dismissal, n, as a function of transparency q

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.

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

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 E: 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:

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

s.t.

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

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

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:

$$d_c = 0, \qquad a_c = \gamma(\lambda)/p \tag{E1}$$

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:

$$\lambda_c \gamma'(\lambda_c) - \gamma'(\lambda_c) = \frac{m}{2}. \tag{E2}$$

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

$$d_s = 1, \qquad a_s(q, \lambda) = \frac{\gamma(\lambda)}{p} - \frac{q\delta\gamma(\lambda)}{1 - \delta(p + q - 2pq)}. \tag{E3}$$

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

$$\lambda_s \gamma'(\lambda_s) - \gamma'(\lambda_s) = \frac{m}{2 - \frac{pq\delta}{1-\delta(p+q-2pq)}}. \tag{E4}$$

The convexity of $\gamma(\lambda)$ implies that the left-hand-side of (E2) and (E4) is increasing in λ. The fact that the right-hand-side of (E2) is smaller than that of (E4) and the right-hand-side of (E4) 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 pa 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 F: 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

[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.

the next harvest, but the governor and the urban sector obtain no revenue at all. This extreme result is clearly due to our simple model and to this particular calibration; but it reflects a general phenomenon: a larger share of the farming output remains in the periphery after bad harvests. This 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.