“The Origins of Inequality: Insiders, Outsiders, Elites, and Commoners”

Greg Dow and Clyde Reed

July 2009
Abstract. Permanent economic inequality is unknown among mobile hunter-gatherers, but hereditary class distinctions between elites and commoners exist in some sedentary foraging societies. With the spread of agriculture, such stratification tends to become more pronounced. We develop a model to explain the associations among productivity, population density, and inequality. We show that regional productivity growth leads to enclosure of the best sites first, creating inequality between insiders and outsiders. This is followed by the emergence of elites and commoners at the best sites. As this process unfolds, elites and commoners have increasingly unequal food consumption. In some cases, the elite specializes in guarding land while relying entirely on the food produced by commoners. Our analysis is consistent with archaeological evidence from southern California, the northwest coast of North America, southwest Asia, and Polynesia.

Acknowledgement. The authors are grateful for comments at the Workshop on the Emergence of Hierarchy and Inequality organized by Sam Bowles of the Santa Fe Institute in February 2009. The Social Sciences and Humanities Research Council of Canada provided financial support but is not responsible for any opinions expressed.
The Origins of Inequality:

Insiders, Outsiders, Elites, and Commoners

1. Introduction

Mobile foraging groups exhibit a minimal amount of economic inequality. There are typically differences in work tasks and food consumption based on age and sex, and bands may claim rights over specific territories that have unequal value, but there are no hereditary class distinctions (Kelly, 2007). Sedentary foraging societies are more diverse. Some are relatively egalitarian, but others display considerable inequality both across and within communities, and they may be divided into hereditary elite and commoner classes. With the spread of agriculture, such inequality tends to become more pronounced and is usually based on elite control over land (Johnson and Earle, 2000).

We develop a theory to explain the associations among (a) rising productivity; (b) rising population density; and (c) rising inequality. The starting point is a region where agents have free mobility across individual food acquisition locations, which we call 'sites'. These sites vary in quality due to soil characteristics, topography, proximity to lakes, rivers, or a coastline, and so on. There are two technologies: one for acquiring food and the other for excluding outsiders from a site. Exclusion requires a sufficiently large group of insiders who want to enforce a claim on the site, perhaps by building a fence or patrolling a perimeter.

The key exogenous variable is the productivity of labor in food acquisition, which increases over time. Our key endogenous variables are population, property rights, and inequality. Population adjusts in a Malthusian way so that when productivity improves,
these gains are absorbed through population growth rather than increased food per capita for the region as a whole. We show that when this Malthusian assumption is combined with endogenous property rights, productivity growth leads to rising inequality.

Starting from a world of open access, productivity growth implies that at some point it becomes profitable for a group of insiders to close the best sites. The insiders at these sites begin to have more food consumption than outsiders in the commons. Further productivity growth leads to closure of progressively lower-quality sites and a contraction of the commons. Inequality in food consumption between agents at open and closed sites continues to increase. Because food per capita remains constant at the regional level, the agents in the commons suffer increasing poverty, at least initially.

When productivity becomes sufficiently high, stratification emerges at the best sites and then spreads across the region. In the stratified sites, commoners get a 'wage' equal to the food consumption per person available at the remaining open sites. The elite at each closed site enjoys rising luxury, while inequality between elites and commoners continues to increase. Eventually, elites at the best sites may stop contributing to food production and specialize in guarding land.

After developing the model, we apply it to several empirical cases. There is a considerable amount of useful archaeological and ethnographic data on several regions, including the southern California coast, the Pacific Northwest, southwestern Asia, and Polynesia. Each of these cases speaks to a subset of the issues we address, and together they are consistent with the full range of predictions derived from our theory.
Our analysis shows that the following conditions are sufficient to generate rising inequality: (a) sites of heterogeneous quality; (b) an exclusion technology; (c) increasing productivity; and (d) Malthusian population. This does not preclude the possibility that other conditions may also have been sufficient. Historical events can be over-determined in the sense that more than one set of sufficient conditions applies simultaneously. But our approach does have a large amount of predictive content, and as we will argue later, these predictions hold up well in confrontations with evidence. Moreover, our analysis shows that several other alleged sources of early inequality, although perhaps relevant in certain cases, are not logically necessary. These include risk, trade, food storage, public works, land improvements, resource depletion, and craft specialization.

We do not address warfare, although warfare often accompanies the emergence of inequality and clearly has a bearing on control over land by insider groups or elites. Our model focuses on prevention of land appropriation by unorganized outsiders, rather than on warfare between competing coalitions. However, it can be applied to cases in which defensive military technology trumps offensive technology. We will return to the topic of warfare in section 7.

The model is most relevant for societies in which insiders or elites control local resources of high quality that are surrounded by open-access resources of lower quality. We develop two distinct versions of the model: one where the region in question is not geographically circumscribed (agents have access to unlimited land of constant quality), and a second where circumscription is important (agents can move costlessly within a finite region, subject to possible exclusion by insiders at some sites, but face prohibitive costs of leaving the region). The latter situation could involve mobility barriers such as
mountains, desert, or oceans (Carneiro, 1970). In either case, we assume that people can run away from closed sites and take refuge in the commons. This rules out most state societies, except for people living in peripheral locations. It also rules out feudalism and slavery (however, see the discussion of slavery in section 7).

A key advantage of our approach is that the ultimate driving force, productivity, is likely to be exogenous for the societies of interest. This is clearly true when productivity growth results from improved climate, but it is also plausible for technical progress. For foraging societies, such progress is generally slow and involves learning-by-doing over many generations (Dow and Reed, 2008). Although productivity growth tends to be more rapid in agricultural societies, full domestication of wild food plants is currently believed to have taken many centuries and may have been unintentional (Dow, Reed, and Olewiler, 2009). The diffusion of agricultural technology caused exogenous productivity shocks for many societies that had previously pursued a foraging lifestyle.

Several other economists have theorized about the emergence of property rights to land in early societies. North and Thomas (1977) argue that foragers had open access to resources that were subject to depletion from over-harvesting. Population growth made this problem more severe and motivated the development of communal property rights over land, which in turn made agriculture attractive. One difficulty with this argument is that it does not account for the diversity of land tenure systems among foragers (Kelly, 2007). Another is the lack of any correlation between common property problems like over-hunting and the timing and location of early agricultural transitions. Finally, the basis for the population pressure is unclear. North and Thomas argue that if one forager band restrained population to conserve natural resources, it would lose out in competition.
with other bands that did not. But depletion of natural resources should reduce the standard of living for all bands simultaneously, thereby lowering fertility and raising mortality until population becomes stationary.

Bowles and Choi (2002) are also concerned with the co-evolution of agriculture and property rights. In their view, foraging bands typically had communal control over land. This system was replaced by individual property rights over land, livestock, and other resources under agriculture. The trigger was a positive climate shock (the onset of the Holocene) that made agriculture more attractive. The result was a positive feedback loop in which decreased ambiguity of resource possession under agriculture reinforced individual property rights, and individual rights encouraged agricultural development.

Baker (2003) seeks to explain why foraging societies have varying land tenure institutions. The explanatory variables in his model include ecology (resource density and predictability) and technology (both for food production and conflict, where the latter includes the costs of defense). The model involves strategic interactions between groups of insiders and outsiders, where groups must decide how much of their endowed territory they will defend, and whether they will intrude on territories defended by others. Baker shows that if resource density is low, some land is left undefended (in the commons). If resource density is high, groups either defend geographically stable territories, or group claims over territories vary with the current state of the world.

Our theory differs from North and Thomas (1977), Bowles and Choi (2002), and Baker (2003) in several ways. First, these authors only explore the creation of communal or individual property rights over land, without addressing economic inequality across or within social groups. Second, North and Thomas (1977) and Bowles and Choi (2002)
focus on the co-evolution of property institutions with agriculture, while Baker (2003) restricts attention to foraging societies. Our theory applies equally to both foraging and agricultural societies, while explaining why inequality often becomes more pronounced in the latter. Third, the driving force in our story is exogenous productivity growth rather than resource depletion, ambiguity of resource possession, or uncertainty about resource availability. Finally, unlike previous authors we endogenize population and take account of heterogeneity among the sites within a region.

There are several distinct archaeological perspectives on the origins of inequality (Hayden and Villeneuve, 2008). The framework most related to ours is variously called 'behavioral ecology', 'political ecology', or 'political economy'. In this view, political and economic institutions reflect the material interests of elites, and the success or failure of a prospective elite is influenced by ecology, geography, technology, and demography (in addition to Hayden and Villeneuve, see also Kirch, 1984; Earle, 1997; Johnson and Earle, 2000; and Kennett et al., in press). Some key questions in this approach include how elites come to power; how a class of landless commoners emerges; why social ranking becomes hereditary; and why commoners tolerate or support elites.

Our theoretical framework is in the same spirit. One difference is that we focus more narrowly on economic inequality, rather than broader questions involving political hierarchy or social ranking. Another is that we operate at a higher level of abstraction by ignoring the detailed strategies pursued by prospective elites. Instead, we simply assume that if elite formation is both feasible and profitable, it will occur. We also abstract from differences in strength, intelligence, or personality traits that may give some individuals a comparative advantage with respect to leadership or the use of violence.
Another influential view from outside economics is that of Jared Diamond (1997), who proposes a theory of the origin of inequality that combines increased productivity with a political economy constraint involving social cohesion. According to Diamond, the earliest organizational units were egalitarian bands, collections of kin-related individuals working together in groups of no more than a few dozen people. When improvements in climate and technology allowed a sedentary lifestyle, social groups developed into tribes of perhaps a few hundred people. Diamond views tribes as egalitarian units within which social cohesion was maintained through kinship ties.

Diamond, along with most anthropologists, associates the emergence of inequality with the formation of chiefdoms whose population numbered in the thousands. Diamond argues that at high population densities, conflicts between individuals and groups could no longer be resolved through kinship ties alone. The continued existence of such societies therefore required the transfer of coercive power to a ‘chief’, which allowed for the emergence of inequality. Our story parallels Diamond by highlighting productivity growth as a key exogenous variable. However, unlike Diamond we endogenize population density. We also stress exclusion of outsiders, rather than conflict resolution among insiders, as the main causal channel responsible for the origin of inequality.

The rest of the paper is organized as follows. Section 2 briefly discusses a few terminological and conceptual issues. Section 3 develops a formal model of a production site with exogenous parameters representing site quality, productivity, and the outside option of the agents. We identify subsets of the parameter space in which the site is (i) open; (ii) closed without stratification; or (iii) closed with stratification. This provides a self-contained theory of inequality for a region without geographic circumscription.
Section 4 models a region with circumscription by combining the analysis of section 3 with Malthusian population adjustment. The outside option at each individual site now becomes an endogenous variable for the region as a whole. The main result is that under a reasonable equilibrium selection rule, increased productivity always leads to greater inequality. We also give a sufficient condition for population to move in the same direction as productivity and inequality.

Section 5 summarizes empirical predictions derived from the theory. Section 6 reviews relevant data from archaeology and anthropology, including four case histories. We argue that each of these cases is consistent with theoretical expectations. Section 7 concludes by discussing the potential relationships between our model of inequality and kinship, downward mobility, slavery, and warfare. The proofs of all formal propositions are in an appendix available from the authors.

2. **Terminology and Concepts**

Before turning to the formal model, it may be useful to clarify some terminology, because our use of certain words differs from that of the archaeological literature.

For archaeologists, the word 'site' means a place where excavations occur. We use it as a shorthand term for any location or territory with natural resources from which food can be obtained. In the context of agricultural societies, our term 'site' might refer to a small parcel of fertile, well-watered land. In a discussion of foraging societies, it might instead refer to a large territory through which a band customarily roams. The choice of a specific territorial unit thus varies pragmatically with the empirical application.

Archaeologists use a variety of terms such as 'complexity', 'ranking', 'hierarchy', and 'stratification' to describe societies with systematic social, political, and/or economic
inequality (Arnold, 1996). Our focus is on unequal access to food. In the formal model, we treat food as a homogeneous consumption good measured in calories. The model could be extended to include the dimensions of quality, reliability, and variety, but we do not pursue these extensions here. It could also be extended to include housing, clothing, craft goods, or other items that economists would normally count as part of the standard of living. When we refer to 'inequality', we are referring to material goods of this sort.

Inequalities with respect to social status and political power are clearly correlated with economic inequality. However, all three are endogenous variables, and we do not want to explain one endogenous variable by linking it with others that are simultaneously determined. Our goal is to explain economic inequality by linking it to the productivity of food labor, which is arguably exogenous for the reasons discussed in section 1. This enables us to specify necessary and sufficient conditions for the emergence of economic inequality that have an unambiguous causal interpretation.

Another potential source of confusion involves the term 'stratification'. We only use this term when referring to economic inequality within a group that occupies a given site or territory. Specifically, it will be used when there is a land-owning elite and a class of landless commoners who share the same site. Another important form of inequality in our analysis involves unequal food consumption between groups using different sites or territories. Such inequalities between insiders and outsiders can exist even if there is no inequality within any group, and therefore no stratification in our sense.

A variety of economic institutions can be used by elites to extract surplus from commoners. One possibility is that the commoners could 'own' land but be taxed by the elite and required to hand over a fraction of their food output. Such payments could also
be construed as rent for the use of land owned by the elite. Commoners could be required to spend part of their time working on elite land or contributing to projects that the elite finds useful (irrigation systems, roads, monuments, and so on). We are not concerned with these institutional details. It is convenient to frame the model as one in which the elite owns land and pays commoners a wage, but this is inessential: the same economic outcomes could be obtained in other ways.

Archaeologists and anthropologists sometimes explain the rise of an elite class of non-food-producers by saying that technological innovation created a 'food surplus'. This is misleading. Almost any society is technically capable of producing more food than the suppliers of food labor personally consume, simply by having the labor suppliers work longer hours. The important question is whether some prospective elite can profitably coerce others into generating such a surplus and handing it over. In our theory, 'surplus' refers simply to the net gain to an elite agent from not being a commoner. This is not a technological datum. The surplus going to elites (if any) is an endogenous variable that will be determined by a regional equilibrium condition.

The next two sections develop the formal model. Readers who are not interested in the details of this analysis can skip directly to section 5, which summarizes the model's empirical content.

3. Inequality Without Circumscription

The model in this section has two possible interpretations. First, it can be treated as a self-contained theory of inequality for a region with a few high-quality sites that are surrounded by unlimited land of constant quality. Alternatively, it can be regarded as an
initial step in modeling a circumscribed region, a task to be completed in section 4. In either case, the model applies equally to foraging and agricultural economies.

At an individual production site, food output (in calories) is \( \theta sf(L) \) where \( \theta > 0 \) reflects regional climate, technology, and resources; \( s > 0 \) is the quality of the site; and \( L \geq 0 \) is labor time used for food production. The input of land is normalized at unity.

\[ A1 \] The function \( f \) is twice continuously differentiable with \( f(0) = 0; f'(L) > 0 \) for all \( L > 0; f''(L) < 0 \) for all \( L > 0; f'(0) = \infty; \) and \( f'(\infty) = 0. \)

There is a continuum of agents so each individual agent is negligible relative to the number of agents at the site as a whole (L refers to a mass of labor, not the labor of an individual agent). Each agent is endowed with one unit of time. These endowments are used for food production, exclusion of outsiders, or some combination of the two (we ignore leisure). The outside option of an individual agent is \( w > 0 \), which is the food available by migration to another site. There is an infinitely elastic supply of outsiders who will enter the site if they are not excluded and can obtain more than \( w \) by doing so.

If we assume agents have access to unlimited land of given quality, the outside option \( w \) is the food per person available by moving to this land. In this interpretation, there must be an upper bound on the average product of food labor in the external land (by contrast with the assumption made in A1) because otherwise arbitrarily large food per person could be obtained from a hinterland of extremely low population density. This restriction will not be needed for the model with circumscription in section 4.
We refer to insiders who cooperate to prevent land appropriation by outsiders as an *elite*. An elite of size $e$ allocates its time between food labor ($e_f$) and guard labor ($e_g$) subject to $e_f \geq 0$, $e_g \geq 0$, and $e_f + e_g = e$.

A2 Given elite food labor $e_f \geq 0$, the minimum guard labor needed to exclude outsiders is $e_g = g(e_f)$, where $g(0) = e_g^0 \in (0, \infty)$. There is some $e_f^0 \in (0, \infty)$ such that $g(e_f) = 0$ for all $e_f \geq e_f^0$. The function $g$ is twice continuously differentiable on $e_f \in [0, e_f^0)$ with $g'(e_f) < 0$ and $g''(e_f) > 0$. The notation $g'(e_f^0)$ refers to the left derivative at $e_f^0$.

We write $(e_f, e_g) \in E$ whenever $e_g \geq g(e_f)$, as shown in Figure 1. The vertical intercept $e_g^0$ is the amount of guard labor (for example, building a fence or patrolling a perimeter) that would be needed to prevent entry if the elite produced no food. The horizontal intercept $e_f^0$ is the amount of elite food labor that would be sufficient to prevent entry even without specialized guard labor. For example, having closely packed neighbors who could detect and kill intruders might be enough to deter attempts at intrusion. We allow any convex combination of these strategies. A tightly packed group of elite food producers reduces the need for perimeter defense, so $g(e_f)$ is decreasing for $e_f \in [0, e_f^0]$. The strict convexity of $g$ captures complementarities between these two strategies for land appropriation and guarantees that all maximization problems have unique solutions.

Now consider any time allocation $(e_f, e_g) \in E$. Available food is shared equally among the members of the elite. Each member receives

(1) 

$$y(e_f, e_g) = \frac{\max_{c \geq 0} \theta s(f(e_f + c) - w c)}{(e_f + e_g)}$$
where \( c \) is the number of commoners admitted to the site. Commoners do not appropriate land, but the elite may allow them to supply food labor in exchange for a wage \( w \) equal to the outside option. Each member of the elite enjoys the surplus

\[
(2) \quad v(e_f, e_g) = y(e_f, e_g) - w
\]

An elite forms at a production site if and only if \( v(e_f, e_g) \geq 0 \) for some \((e_f, e_g) \in \mathcal{E}\). In this case we say the site is closed. Otherwise it is open. If the elite at a closed site chooses \( c > 0 \), we say that the site is stratified. We denote the total population at a site (elite and commoners together) by \( n = e + c \).

A3 When a site is closed, the time allocation \((e_f, e_g)\) maximizes \( v(e_f, e_g) \) subject to \((e_f, e_g) \in \mathcal{E}\). The commoner population is determined as in (1). When a site is open, there is no elite and the commoner population satisfies \( \theta_{sf}(c)/c = w \).

Maximization of surplus per elite agent is a reasonable behavioral assumption. If elite size is less than the surplus-maximizing level, elite agents can raise their per capita food consumption by recruiting commoners into the elite (commoners are willing to join because this is at least as good as their outside option). If elite size exceeds the surplus-maximizing level, any members who exit through death or migration will not be replaced, because this would not be profitable for the remaining insiders (see section 4 for more on fertility and mortality).

At open sites, no set of agents can gain by excluding others. There is entry into the site if food per capita is above \( w \) and exit if food per capita is below \( w \). Equilibrium thus requires \( \theta_{sf}(c)/c = w \) as stated in A3, with \( e = 0 \) and \( n = c > 0 \).
Proposition 1 (existence and uniqueness). Define \( z = \theta s/w > 0 \). For a closed site there is a unique elite and commoner time allocation \( e_i(z) \in [0, e^0_i] \), \( e_g(z) = g[e_i(z)] \), and \( c(z) \geq 0 \).

Proposition 2 (open sites). There is an \( A > 0 \) such that the site is open for \( z \in (0, A) \) and closed for \( z \in [A, \infty) \). For an open site there is a unique \( c(z) \equiv n(z) > 0 \) that is continuous and increasing on \( (0, A) \). We have \( c(z) \to 0 \) as \( z \to 0 \) and define \( c(0) \equiv 0 \). We also have \( c(z) \to c'(A) \) as \( z \to A \) from below, where \( c'(A) \) uniquely satisfies \( Af[c'(A)]/c'(A) = 1 \).

Proposition 3 (closed and unstratified sites). The functions \( e_i(z), e_g(z), e(z), c(z), \) and \( n(z) \) are continuous on \( (A, \infty) \). These functions are right continuous at \( z = A \). There is a \( B > A \) such that the site is closed and unstratified for \( z \in [A, B] \), and stratified for \( z \in (B, \infty) \). Thus \( c(z) = 0 \) on \( [A, B] \) and \( c(z) > 0 \) on \( (B, \infty) \). There is a constant \( e^A_i > 0 \) such that \( e_i(z) = e^A_i \) on \( [A, B] \). Let \( \eta = e^0_i f'(e^0_i)/f(e^0_i) \) be the output elasticity at \( e^0_i \), where \( \eta \in (0, 1) \).

(a) If \( 1 + g'(e^0_i) \leq \eta \) then \( e^A_i = e^0_i \) and \( c'(A) = e(A) \) so that \( n(z) \) is continuous at the boundary \( z = A \).

(b) If \( \eta < 1 + g'(e^0_i) \) then \( e^A_i < e^0_i \) and \( c'(A) > e(A) \) so that \( n(z) \) drops discontinuously at the boundary \( z = A \).

Proposition 4 (closed and stratified sites).

(a) If \( 1 + g'(e^0_i) \leq 0 \) then \( e_i(z) = e^0_i \) for all \( z \in [B, \infty) \).

(b) If \( 0 < 1 + g'(e^0_i) < \eta \) there are \( C, D \) with \( B < C < D \leq \infty \) such that \( e_i(z) = e^0_i \) for \( z \in [B, C] \) and \( e_i(z) \) is decreasing on \( [C, D] \). If \( g'(0) \leq -1 \) then \( D = \infty \). If \( -1 < g'(0) \) then \( D < \infty \) and \( e_i(z) = 0 \) for all \( z \in [D, \infty) \).

(c) If \( 1 + g'(e^0_i) = \eta \) then \( C = B \) with \( e_i(C) = e^0_i \). All other results are as in (b).
If $\eta < 1 + g'(e_{t}^0)$ then $C = B$ with $e_{t}(C) = e_{t}^A < e_{t}^0$. All other results are as in (b).

In (b), (c), and (d), $e_{g}(z)$ is increasing on $[C, D]$ and $e(z)$ is decreasing on $[C, D]$. In all cases, $c(z)$ and $n(z)$ are increasing on $[B, \infty)$.

It is convenient to interpret these results by supposing that local improvements in resources or technology increase productivity ($\theta$) at one site or perhaps a group of nearby sites. In either case, the outside option ($w$) remains fixed.

Proposition 2 shows that as $z \equiv \theta s/w$ rises, the site first becomes closed at $z = A$. At this point, an elite asserts ownership of the site and blocks further entry. Proposition 3(a) shows that if the constraint slope $g'(e_{t}^0)$ at the corner $(e_{t}^0, 0)$ is steep enough relative to the output elasticity $\eta$, this transition could involve a smooth change in the number of agents at the site, as in Figure 2(a). If the constraint slope at the corner is flatter, then by Proposition 3(b) the transition to a closed site involves an abrupt drop in food labor, the immediate use of guard labor, and an abrupt drop in total population, as in Figure 2(b).

The difference between these cases involves the most profitable way to achieve exclusion. In case (a), the surplus-maximizing way to exclude outsiders at $z = A$ is to avoid specialized guard labor and rely entirely on the density of food producers, so the closure of the site is postponed until the number of agents reaches $e_{t}^0$ as in Figure 2(a). This results in a smooth transition from an open site to a closed one. But in case (b), the surplus-maximizing way to exclude outsiders at $z = A$ involves the use of guard labor as shown by point $e^A$ in Figure 2(b). Once this becomes profitable, there is an abrupt drop in population because the existing agents are not all needed in order to close the site (the total population at $e^A$ is smaller than $c(A)$). Hence a subset of these agents either forces
out the others or permits the insider group to contract through attrition, until surplus has been maximized for those who remain.

Once the site has been closed, there is an interval \([A, B]\) on which the insiders do not find it profitable to recruit commoners as food producers and the site is unstratified. In this interval a higher \(z\) makes the insiders better off but does not affect the size of this group or its time allocation. A commoner class arises only after \(B < z\).

For stratified sites, Proposition 4 shows that several possibilities can arise. If the slope of the constraint is steep enough, as in part (a), the site never moves away from the corner solution \((e_f^0, 0)\) for any value of \(z\) because the cost of guard labor in foregone food output is always too large. In other cases, as \(z\) increases the elite begins to contract after \(C < z\). This process is accompanied by a gradual decline in elite food labor and a rise in guard labor, as the elite's time allocation moves up and to the left along the boundary of the feasible set \(E\). The growth of the commoner class more than compensates for the contraction of the elite, so the total population \(n(z)\) rises.

Two possibilities can occur at high \(z\) values, as indicated in Proposition 4(b). If the constraint slope \(g'(0)\) is steeper than \(-1\), a corner solution with \(e_f = 0\) is impossible, so the elite always contributes some food labor. But if \(g'(0)\) is flatter than \(-1\), for \(D \leq z\) the elite produces no food and specializes in guarding land. Further increases in \(z\) expand the commoner class and enrich the elite but have no effect on elite size or time allocation.

4. **Inequality With Circumscription**

The results from section 3 suffice when the outside option of the agents at a site is exogenous and does not vary with the determinants of productivity at that site. But often the outside option at one site depends on opportunities available elsewhere in the same
region, and climate or technological innovation affects all sites in the region in a parallel way. In such situations, the outside option is endogenous for the region as a whole.

Here we assume that mobility is costless within a region, but agents find it too costly or impossible to leave the region. There is a continuum of production sites. Site quality is indexed by \( s \in [0,1] \) where the number of sites of quality \( s \) is described by a bounded and continuous density function \( q(s) > 0 \).

We use the following terminology to discuss the distribution of food consumption at a particular site. Food per capita for the elite (\( y \) in equation (1)) is called the degree of luxury. Food per capita for commoners (\( w \)) is called the wage. The difference between \( y \) and \( w \) (the surplus \( v \) in equation (2)) is called inequality. Our main analytic result is that under quite general conditions, higher region-wide productivity (whether caused by better climate or technology) leads to more inequality at every closed site.

Time is discrete. In period \( t \), the total adult population for the region is \( N_t \). Each individual adult in period \( t \) has children who survive to become adults in period \( t+1 \). The number of such children is equal to the parent's food income multiplied by a constant \( \gamma > 0 \). This relationship arises because children of richer parents are more likely to survive to adulthood, because children are a normal good in the parental utility function, or both.

We assume that \( \gamma \) is identical for all adults, so the aggregate number of new adults in any period \( t+1 \) is \( \gamma Y_t \) where \( Y_t \) is regional food output in period \( t \). We also assume that a constant fraction \( \delta > 0 \) of the adult population dies in each period after their children are born. The population dynamics for the region are then described by \( N_{t+1} = (1-\delta)N_t + \gamma Y_t \). In a long run equilibrium where \( N \) and \( Y \) are stationary, we have \( Y/N = \delta/\gamma \equiv \beta \). This is the Malthusian feature of the model: as long as the fertility and mortality parameters are
constant, every long run equilibrium yields the same food per capita at the regional level. Productivity gains are absorbed through higher population rather than food per person.

The equilibrium condition \( Y = \beta N \) can be written as

\[
\int_0^1 \theta f[sf[e(x_s) + c(x_s)]q(s)] ds = \beta \int_0^1 n(x_s)q(s) ds
\]

where \( x = \theta/w \in (0, \infty) \)

The integral on the left side is regional food output and the integral on the right side is regional population. The population at an individual site with \( z = \theta s/w \) was denoted by \( n(z) \) in section 3. Here we define \( x = \theta/w \) to be the region-wide ratio of productivity to the wage, so \( z = xs \) and thus the population density at a typical site of quality \( s \) is \( n(xs) \).

Similarly, \( e_f(x_s), e_g(x_s), e(x_s), \) and \( c(x_s) \) denote elite food labor, elite guard labor, total elite size, and commoner labor at a typical site of quality \( s \). Using results from section 3, the integrands in (3) are bounded on the interval \( s \in [0, 1] \) for any fixed \( x > 0 \). They are also continuous, except possibly when \( z = xs = A \) as in Proposition 3(b).

We define two break points in the distribution of site qualities by

\[
a(x) \equiv A/x \quad \text{and} \quad b(x) \equiv B/x
\]

From Propositions 2 and 3, a site is open if \( s \in [0, a(x)) \); closed but unstratified if \( s \in [a(x), b(x)) \); and both closed and stratified if \( s \in (b(x), 1] \).

**Proposition 5** (existence). For each \( \theta > 0 \) there is at least one \( w > 0 \) satisfying (3). There are no solutions to (3) such that \( w > \beta \). For \( \theta \leq \beta A \), \( w = \beta \) satisfies (3). For \( \theta > \beta A \), any solution to (3) has \( w < \beta \).
It is easy to see why $w \leq \beta$ must hold. Substituting $w = \theta \sf[c(x_s)]/c(x_s)$ for open sites, using the definition of surplus for closed sites in (2), and reorganizing (3) gives

$$w(x) + R(x)/N(x) = \beta$$

where

$$N(x) = \int_0^1 n(x_s)q(s)ds$$

and

$$R(x) = \int_{a(x)}^1 v[e_f(x_s), e_g(x_s)]e(x_s)q(s)ds$$

$N(x)$ is regional population and $R(x)$ is regional surplus. Equation (5) shows that labor income per capita $w(x)$ and land income per capita $R(x)/N(x)$, measured in food, must add up to $\beta$, the income level consistent with demographic equilibrium. When $a(x) \geq 1$, the integral $R(x)$ is zero because all sites are open, so $w(x) = \beta$. But when $a(x) < 1$ some food goes to insiders or elites in the form of land rent, so $w(x) < \beta$ must hold.

In the latter case, the commoner wage is below the level $\beta$ needed for population replacement. For site qualities slightly above $a(x)$, insider food consumption is close to $w$ and thus below $\beta$ as well. However, elites at other sites enjoy greater food consumption than is needed for demographic replacement, and this permits a stationary population for the region as a whole. Because the commoner population must be replenished in every period, there is some downward mobility among the children of the elite (see section 7).

We would like to use equation (3) to obtain a relationship between the exogenous productivity $\theta$ and the endogenous wage level $w$. This would generate predictions about how inequality evolves at each site as productivity rises throughout the region over time. It would also yield predictions about how property rights evolve, because $x = \theta/w$ gives the location of the boundary $a(x)$ between open and closed sites as well as the boundary $b(x)$ between unstratified and stratified sites.
In practice, it is easier to treat $\theta$ and $w$ as functions of $x$. Condition (3) gives

$$(6) \quad \theta = \phi_\theta(x) = \frac{\beta N(x)}{Y(x)} \quad \text{and}$$
$$w = \phi_w(x) = \frac{\beta N(x)}{xY(x)}$$

where

$$N(x) \equiv \int_0^1 n(xs)q(s)ds \quad \text{and} \quad Y(x) \equiv \int_0^1 s[f[e(xs) + c(xs)]q(s)ds.$$ 

These functions are continuous on $x > 0$ (see the proof of Proposition 5). As shown in Figure 3, each ray from the origin in $(\theta, w)$ space has slope $1/x$, so each ray is associated with a unique point $[\phi_\theta(x), \phi_w(x)]$ from (6). This point must satisfy condition (3) and no other point on the same ray can do so. When $x$ increases and we move to a flatter ray in Figure 3, one of the following must occur: (i) $\theta$ and $w$ both rise; (ii) $\theta$ rises and $w$ falls; or (iii) $\theta$ and $w$ both fall. Cases along the boundaries between (i) and (ii) or (ii) and (iii) can occur, but it is impossible for $\theta$ to fall and $w$ to rise simultaneously when $x$ increases.

The next result links higher productivity with greater inequality.

**Proposition 6** (inequality). Consider any two equilibria $(\theta_1, w_1)$ and $(\theta_2, w_2)$ and any site quality $s$ such that the site is closed in both equilibria. Denote the corresponding levels of luxury and inequality at the site by $(y_1, v_1)$ and $(y_2, v_2)$. If $\theta_1 < \theta_2$ and $x_1 < x_2$ then $y_1 < y_2$ and $v_1 < v_2$.

The most important feature of this result is that it does not directly involve $w$. This may be surprising because $w$ is the wage that the elite pays commoners at stratified sites, and therefore affects the level of luxury $y$ for the elite at these sites. The wage $w$ also enters our definition of inequality $v$, which is the difference between $y$ and $w$. Even so, higher
productivity generates greater inequality whenever $\theta$ and $x$ move in the same direction, regardless of whether $w$ rises or falls.

Suppose that for each productivity level $\theta > 0$, the equilibrium with the highest wage $w$ is selected. We can show that under this rule the requirements of Proposition 6 are satisfied, and therefore higher productivity always leads to more inequality.

**Proposition 7 (equilibrium selection).** Let $W(\theta) = \{\max \varphi_w(x) \text{ such that } \varphi_\theta(x) = \theta\}$.

(a) $W(\theta)$ is well-defined for all $\theta > 0$.

(b) If $w$ is determined according to $W(\theta)$ then $\theta_1 < \theta_2$ implies $x_1 < x_2$.

**Corollary:** If $w$ is determined according to $W(\theta)$ then increasing productivity implies both increasing luxury and increasing inequality at every closed site.

As shown in Figure 4, the selection rule in Proposition 7 implies that backward-bending segments of the equilibrium locus $[\varphi_\theta(x), \varphi_w(x)]$ can be ignored, because at any point $(\theta, w)$ on such a segment there is another equilibrium with the same productivity and a higher wage. Therefore if the maximum wage is always selected, points along a backward-bending segment will not be observed.

Now consider the path of the wage as productivity grows in chronological time. Start at a productivity level $\theta \leq \beta A$ with $W(\theta) = \beta$. Rising productivity may eventually lead to multiple wage equilibria at the current $\theta$ as in Figure 4. In such situations, the rule $W(\theta)$ ignores equilibria with lower wages and moves along the upper part of the equilibrium locus until a local maximum of $\theta$ with respect to $x$ is reached. Any further growth in $\theta$ causes an abrupt drop to a lower wage level, followed by a resumption of
continuous wage adjustment along the next part of \( W(\theta) \). This may be repeated many times if there are many productivity intervals where multiple equilibria exist.

We believe this is a reasonable prediction about the trajectory of early societies experiencing productivity growth. There is little reason to expect a discontinuity in the relationship between productivity and the wage until continuous adjustment is no longer possible. When a jump becomes unavoidable, we conjecture that the discontinuity will involve the smallest wage reduction needed to restore equilibrium.

Aside from simplicity and empirical plausibility, the equilibrium selection rule in Proposition 7 biases the analysis against our argument that higher productivity leads to greater inequality by always selecting the highest wage permitted by current productivity. Therefore the corollary to Proposition 7, which is immediate from Proposition 6, carries additional weight.

The corollary has another implication. Because \( \theta \) and \( x = \theta/w \) move in the same direction in chronological time, rising productivity causes the site qualities \( a(x) \) and \( b(x) \) to fall in (4). Thus the best sites are closed first. Over time the set of open sites \([0, a(x)]\) contracts and the set of closed sites \([a(x), 1]\) expands. Eventually, the best sites become stratified and the set of stratified sites \([b(x), 1] \) likewise expands. As \( \theta \to \infty \), we have \( x \to \infty \), which gives \( b(x) \to 0 \). Therefore at high productivity levels, almost all of the sites in the region become closed and stratified.

These institutional transitions may unfold gradually, or they may include abrupt jumps in \( x \) due to discontinuities in the wage function in Proposition 7, as illustrated in Figure 4. Thus, a large group of sites may simultaneously become closed or stratified in response to modest productivity gains. Any such discontinuities are 'sudden' only on an
archaeological time scale. Because we are comparing equilibria involving steady state populations, several human generations may need to elapse during the approach to the new equilibrium even when a discontinuity occurs.

A number of empirical implications can be derived from the analysis in section 3. There we characterized a site by the parameter \( z = \theta s/w \). For an individual site of fixed quality \( s \), this becomes \( z = x(\theta)s \), which is increasing in \( \theta \) under the assumptions of Proposition 7. As a result, all implications that follow from rising \( z \) in Propositions 1-4 carry over to rising productivity \( \theta \) here. For example, if stratified sites have falling elite farm labor, rising elite guard labor, rising commoner labor, and rising total population as \( z \) increases in Proposition 4, then the same will be true over time at each stratified site as regional productivity growth occurs. The only difference is that multiple equilibria at the regional level may cause occasional forward jumps in this process.

Under the condition in Proposition 3(a), it is easy to show that regional population \( N(x) \) is an increasing function of \( x \) and therefore of productivity. This is consistent with robust empirical evidence that productivity, population, and inequality rise in tandem (see sections 1 and 6). However, there is an exception to this general rule. If the condition in Proposition 3(b) applies, the closure of additional sites can cause regional population to fall as productivity rises. This requires a substantial discontinuity in population at \( z = A \) and a substantial density \( q(s) \) at the boundary between open and closed sites. Moreover, it can be shown that for sufficiently high productivity levels, this property rights effect is dominated by the rising demand for commoner labor at stratified sites, so that population increases with productivity. We regard this result mainly as a theoretical curiosity, but it could be important in some empirical applications.
It is difficult to sign the relationship between $\theta$ and $w$. Suppose $\theta$ increases and this leads to an increase in $x$ using the assumption of Proposition 7. Several issues are apparent from equation (5). First, $N(x)$ could rise or fall, as explained in the preceding paragraph. But in addition, the effect on aggregate surplus $R(x)$ is unclear. At stratified sites, total surplus may either rise or fall as $x$ rises, depending on the relative importance of rising surplus per elite agent ($v$) or declining elite size ($e$). The outcome depends on the curvatures of the production functions for food and exclusion. Moreover, the effect on $w$ depends on a comparison of the rates of change in $N$ and $R$. We do know from Proposition 5 that $w$ must be falling as the first sites are enclosed. But thereafter, we cannot rule out the possibility that rising productivity could make commoners better off, although it will never make them as well off as they were when all sites were open.

We close with a brief comment on disequilibrium population adjustments. All of the results in this section involve comparisons among long run equilibria with stationary populations. However, in section 6 it will be necessary to consider cases where regional population $N$ is initially below its steady state level and rises over time. We then use the short run equilibrium condition $N = \int_0^1 n(xs)q(s)ds$ where $N$ is treated as exogenous and $x = \theta/w$. Holding productivity $\theta$ constant and using a condition like that in Proposition 7 where the highest equilibrium wage is selected for each $N > 0$, one can show that $w$ is a decreasing function of $N$, perhaps with some discontinuous downward jumps. Because $\theta$ is constant, this implies that $x$ rises as $N$ rises. All of the previous results based on rising $x$ therefore go through unchanged as regional population approaches a steady state.

5. **A Summary of Theoretical Predictions**
Many empirical predictions follow from our model with regional circumscription in section 4. For easy reference, we summarize the most important ones here.

**Initial enclosure.** Starting from a world with no or minimal exclusion of outsiders from production sites, increases in the productivity of food technology (whether due to climate or innovation) should eventually cause some sites to become closed to outsiders. The enclosure process should start with the highest-quality sites in a region and gradually extend to lower-quality sites as productivity continues to rise.

**Closed and unstratified sites.** Initially, closed sites should not have any internal stratification. Agents in higher-quality closed sites should have higher food consumption per capita, and agents at any closed site should be better off than the agents at open sites. The size of the insider group and its time allocation between food and guard labor should not vary with site quality. The population density at such sites should be no higher, and perhaps lower, than the density at the best open sites.

**Stratified sites.** If productivity continues to increase, some closed sites should become internally stratified. This process should begin with the highest-quality sites and gradually extend to lower-quality sites. Commoners at the stratified sites should receive the same food per capita as people at open sites. Inequality should be greatest at the highest-quality sites. Food consumption for elite agents should rise with site quality and productivity. When stratification is first observed, elite agents should be contributing a positive amount of food labor. The size of the elite at stratified sites should stay constant or decline as productivity rises. If it falls, over time more elite labor should be devoted to guarding land while less is devoted to food acquisition. The elite may eventually spend
all of its time guarding land and rely entirely on food produced by commoners. From that point on, further increases in productivity should not change the size of the elite.

6. **Empirical Evidence**

Some archaeological and anthropological background is useful in assessing our theory. Evidence for significant and institutionalized intra-group differences in status and wealth is confined to the last 13,000 years, even though modern humans have existed for about 150,000 years. Before this time, groups remained small, occupied relatively large territories at low densities, and moved to follow resources. Group fission and emigration were generally favored over localized increases in group size and density (Kennett et al., in press). Our theory predicts that inequality should be minimal under such conditions.

Climate change around 13,000 years ago triggered a shift toward agriculture in southwest Asia and probably elsewhere (Dow, Reed, and Olewiler, 2009). This led to a substantial increase in the rate of technological innovation (Dow and Reed, 2008). By 11,600 years ago, the Holocene brought an improvement in mean climate and a dramatic reduction in variance throughout temperate parts of the northern hemisphere (Richerson, Boyd, and Bettinger, 2001). This combination of a benign environment and improving technology led to substantial productivity growth. As our theory predicts, the result was increasing population density, appropriation of the best sites by insiders, and eventually inequality between elites and commoners at the best sites.

This broad portrait is consistent with inferences about nutrition and health from the analysis of skeletal remains. Boix and Rosenbluth (2009) survey archaeological data on human skeletons with specific reference to height, an indicator of nutrition during pre- and adolescent growth spurts, and other attributes of bones that have implications for
human health -- e.g., bone lesions due to anemia and from infections, incomplete enamel formation in teeth, and loss of bone mass. The authors use dispersion measures to proxy for economic inequality. Most relevant for our purposes are findings of a low incidence of inequality in hunting and gathering societies, increasing inequality with the rise of agriculture, and a strong correlation between population density and inequality.

Cohen (1989, 1998) and Cohen and Crane-Kramer (2007) use skeletal evidence to infer that by comparison with hunter-gatherers, early farmers had worse nutrition, shorter stature, and shorter lives. Intensification of agriculture led to increased diversity in health outcomes, with elite health improving and commoners remaining badly off.

Anthropological evidence is consistent with this story. Kelly (2007: ch. 8) notes that most known hunter-gatherer societies are relatively egalitarian, with only a few that have substantial inequality (two of these will be discussed below). The latter group has a number of common characteristics, including highly predictable natural resources, high population densities, large settlement sizes, perimeter defense, and tightly controlled resource ownership. There is typically a hierarchical class structure based on descent or wealth (or both). According to Kelly, population density is the key proximate cause of inequality among hunter-gatherers, with sedentism and storage as supporting factors.

Johnson and Earle (2000) take a similar view of the evolution of human societies, including those with agriculture. Using a large number of anthropological case studies, they argue that technological sophistication and population density are the main causal factors behind the development of larger-scale societies with substantial inequality. As was true for the archaeological data, this perspective fits comfortably with our theory.
We turn next to a series of regional case histories. The first two involve foraging societies and help make the point that the path to inequality does not depend crucially on agriculture. The other two cases address the origin of inequality in agricultural societies.

**Channel Islands.** The Channel Islands are located off the coast of southern California and have good archaeological data going back nearly 13,000 years (unless otherwise specified, information is from Kennett, 2005, and Kennett et al, in press). The residents were hunter-gatherers who depended heavily on the rich marine environment for food. Site quality varied with terrain, the presence of kelp zones, and watershed size. The quality of potential coastal village locations has been ranked using a Geographic Information System, with the quality of location based solely on environmental variables and independent of archaeological findings.

Communities expanded slowly before 1500 BP (before present), and first and second ranked habitats were filled in during this period. In part, the slow growth in population can be attributed to periods of warm and variable sea-surface temperatures that resulted in low marine productivity. Not surprisingly, sites were settled in order of quality. After 1500 BP there was a rapid population expansion, and local population densities rose as people nucleated into villages. Over the period 1500-1300 BP, third and fourth ranked habitats came into use, resulting in the occupation of all viable locations. This period marked the end of open access at the best sites, which was replaced by competition among communities and unequal control of local resources.

Skeletal evidence of insider-outsider inequality in the period after 1500 BP can be inferred from the findings of Lambert and Walker (1991). *Cribra orbitalia* is a condition that develops as part of a child's response to anemia. Comparing contemporaneous
skeletal remains across the Channel Islands reveals, for example, less than a 30% incidence of *cribra orbitalia* on Santa Cruz (largest of the northern Channel Islands and rich in diverse plants and animals) compared to an incidence over 70% on San Miguel (a small isolated island with a shortage of fresh water and terrestrial resources). This difference is explained by differences in water contamination, exposure to parasites, and the nutritional adequacy of diet.

By 1300 BP there was a large increase in lethal violence involving projectile points. It is likely that defense was a public good fostering corporate group formation. Clear social inequality (inferred from burial practices) emerged by around 650 BP. By the time of first contact with Europeans in 1542 AD, total population on the Channel Islands was around 3,000 people living in at least 22 villages, which varied in size and sociopolitical importance. The village was the primary political unit, although sometimes multiple villages belonged to the same polity. The society was socially ranked and included hereditary chiefs whose kinship system was patrilineal and patrilocal, and who often married into chiefly families on the mainland. The majority of the people had a matrilineal and matrilocal kinship system.

Productivity markers include the following. The development of the single-piece fishhook (made from shell) occurred between 2500-2100 BP. The bow and arrow was introduced between 1500-1300 BP. After 1300 BP there was further development of new fishing tackle and advances in fishing technology related to plank canoes and toggling harpoons. These developments resulted in exponential increases in fishbone densities, and fish from mid-water, deep-ocean, and open-ocean habitats became increasingly common in faunal assemblages. Fish contributed 50-65% of the faunal remains. It is
probable that the bow and arrow raised productivity for the other 35-50% of the faunal diet. Another productivity-raising consequence of plank canoes was the exploitation of comparative advantage through specialization and trade between the islands and with the mainland after 1300 BP. There is evidence of large-scale craft specialization on the islands as well as significant trading of groundstone (from the islands) for goods from the mainland, acorns being a prime example.

In the period between 1300-650 BP, population continued to increase, probably due to learning-by-doing with respect to these new technologies. At the same time, there is evidence that some locations experienced relative productivity declines. Climate conditions became dry and unstable. A series of severe droughts made potable drinking water unavailable at fourth-ranked locations and some third-ranked locations. Increased surface water temperatures reduced the availability of some marine food supplies, especially at sites in which kelp forests were relatively small.

Our model from section 4 correctly predicts several aspects of this history. For the period prior to 1500 BP, total population was low relative to the availability of good sites, it was infeasible or unprofitable to restrict entry into the high-quality sites, and thus there is no evidence of economic inequality. Subsequent productivity increases involving fishing, hunting, and trade led to rapid population growth. At this point access to the best sites was restricted, more marginal sites were occupied, and insider-outsider inequality becomes visible in the archaeological record.

The explanation for elite-commoner inequality by 650 BP is more complicated, because productivity was rising due to better technology but at least sometimes falling simultaneously due to climate change. Because the population for the Channel Islands as
a whole continued to rise during 1300-650 BP, we infer that the former effect dominated for the region as a whole. This is consistent with the eventual (although slow) emergence of stratification. However, productivity at inferior sites did drop due to climate change, both absolutely and relative to the best sites. The result was to reduce the outside option for people who did not have a claim on land at the best sites, making them more willing to accept unequal consumption in exchange for living and working at these sites. Drier average conditions, severe droughts, and high surface water temperatures would all have contributed to stratification by tilting relative productivities in favor of the best sites.

The Northwest Coast (information is from Ames, 1995 except where noted). The northwest coast of North America, which ranges from northern California to southeastern Alaska, is the classic anthropological example of a sedentary hunter-gatherer society with locally high population densities, inequality, and warfare. The NWC has a wet and mild climate with rich, predictable resources. The most important food resource was salmon, which was highly abundant at certain times and places, and could be stored. Secondary resources included other fish, shellfish, marine mammals, berries, and terrestrial hunting.

At the time of European contact in the 1770s, the basic social unit was the 'house', a large cedar-plank dwelling with 80-150 occupants. Houses were organized into small autonomous villages of about 180-500 people. The most fundamental social distinction was between free people and slaves, who were captured in war and had no rights of any kind. Free people were divided into elite and commoner classes, with systematic ranking within the elite based on genealogical distance from a chiefly line. Marriages usually occurred within these classes and one's status was inherited.
Chiefs did not have much coercive control over members of their houses, except for slaves. Commoners had no permanent allegiance to a house and were generally quite mobile, with chiefs often competing for new house members. In effect, commoners were labor suppliers and chiefs were employers. If dissatisfied, commoners could either move to another house or create a new house on inferior land.

In theory, land and other resources were owned by the house as a whole, but in practice chiefs managed the use of most natural resources as well as the house's capital stock. The prestige of a house and its chief depended upon the ecological productivity of the house's territory and the population it could support. Competition for good sites could lead to warfare where villages were attacked, the residents were driven off, and local resources were appropriated by the victors.

Kelly (2007: 321-328) notes that there is a gradient on the NWC from relatively uniform resources in the south to increasingly localized resources in the north. In the northern subregion, resource ownership was more tightly defined, residential groups were more sedentary, settlement sizes were larger, social ranking was more systematic, and the inheritance of wealth and status was more important.

The earliest known occupants of the NWC were mobile foragers dating to about 10,000 BC (Ames, 2004). There is no evidence for status differences or conflict at this time. Residential patterns became more stable during the middle Holocene and the first houses date to about 3000-4000 BC. Cemeteries may have existed as early as 4000 BC but were clearly present by 2000 BC, and indicate "long-lived, stable territorial social groups" (Ames, 2004: 63). Grave goods suggest that inequality, although probably not permanent ranking, may have existed by 2000 BC.
There is little evidence that the people of the NWC specialized more in particular foods (such as salmon) over the course of the Holocene, or that human harvesting led to significant resource depletion (Butler and Campbell, 2004). However, there is evidence of rising technological proficiency, particularly between 2500-1000 BC, based on the use of net weights, toggling harpoons, fish weirs, storage boxes, and improved canoes (Ames, 2004: 63). By 1800 BC - 500 AD, population levels were higher than in earlier periods, residential sites were used for longer periods, settlement sizes increased, and community-level social or political organization developed. Ames (1995: 181) believes that the later elite evolved from the genealogical core of large permanent extended households starting around 1500-1000 BC. These large households controlled good resource locations and provided the basis for chiefly descent groups.

Technology continued to improve and populations peaked sometime during 200 - 1770 AD. During the latter period, inequality was expressed in markedly different house sizes as well as major differences in the richness of burial mounds. Fortifications became increasingly common after 800 AD and warfare intensified.

The Northwest Coast supports our theoretical model in several ways. First, it is consistent with a causal channel running from productivity growth to population growth and inequality. Second, the south-to-north gradient in the degree of site heterogeneity corresponds to a gradient in the degree of inequality. Third, the institutional structure of these societies, with mobile commoners and open access to low-quality land, parallels our modeling framework. Finally, the NWC highlights the point that our theory applies to foraging societies with appropriate ecological, geographic, and technological features. Agriculture may promote inequality, but is not a necessary condition for it.
Southwest Asia. This region is of special interest because agricultural economies first evolved there. Before agriculture, during a period that lasted about 1500 years and ended around 13,000 BP, the climate was mild and rainfall was plentiful. At this time the region was occupied by early Natufians, who were largely sedentary foragers. Variability in mortuary practices at specific sites provides some evidence of elite-commoner inequality (Kuijt and Prentiss, 2009). This is not surprising from the standpoint of our theory, because excellent climatic conditions led to abundant natural resources, high productivity, increased regional population, and increased sedentism at the best sites. Although differences in nutrition for this period have not yet been documented from skeletal remains, our theory suggests that insider-outsider inequality should have preceded the emergence of elite-commoner inequality within sites.

The village of Abu Hureyra provides evidence of insider-outsider inequality in the period after 13,000 BP. Abu Hureyra was located on the ecotone between the Euphrates Valley and the woodland-steppe in what is today northwest Syria (unless otherwise indicated, data are from Moore et al., 2000). Abu Hureyra was founded around 13,500 years ago and inhabited by sedentary foragers. The remains of food plants at this time reflect a diverse diet typical of hunter-gatherer societies. Abu Hureyra enjoyed a number of advantages over most other sites: it was near the Euphrates River and therefore not dependent on rainfall for fresh water; it was on the annual migration route for gazelle; and food sources included wild cereals as well as nuts and tubers from the nearby woodland. These advantages resulted in Abu Hureyra becoming one of the largest and longest-lasting villages in the region.
Around 13,000 years ago, climate change throughout the northern hemisphere brought a temporary return to Ice Age conditions known as the Younger Dryas, which lasted over 1000 years. During this period many settlements were abandoned, sedentary lifestyles were replaced by mobile foraging in large parts of southwest Asia, and skeletal data indicates a region-wide decline in nutritional levels, health status, and population (Smith, 1991; Mithen, 2003).

Nevertheless, population levels increased substantially at Abu Hureyra during the Younger Dryas and analysis of skeletal remains indicates that the nutrition level remained constant there. Local population growth under adverse climate conditions indicates that people were probably migrating from elsewhere in the region. The lack of evidence for violence or fortifications at Abu Hureyra suggests that any such migration was peaceful, but kinship ties and threats of potential violence would have imposed constraints on the immigration process (Bar-Yosef and Meadow, 1995). Those allowed into Abu Hureyra prospered, while those who remained outside faced starvation. Cultivation of wild plants, leading eventually to full domestication, began at Abu Hureyra and similar refuge sites in the Younger Dryas (Dow, Reed, and Olewiler, 2009), which ended around 11,600 BP.

Our theory suggests the following interpretation. Unlike the case of increasing productivity discussed in section 4, the negative climate shock of the Younger Dryas caused a region-wide decline in productivity associated with reduced rainfall. This was relatively more severe at sites with less reliable sources of ground water (see Dow, Reed, and Olewiler, 2009). Within the framework of section 3, the result was a smaller decline in productivity at Abu Hureyra and other refuge sites ($\theta$) as compared with productivity elsewhere ($w$). This caused an increase in the ratio $z = \theta s/w$ at Abu Hureyra. As shown
in section 3, the response should be growing population density at high-quality sites until these sites become closed and insider-outsider inequality emerges. We believe this is an accurate description of events in southwest Asia during the Younger Dryas.

Interestingly, the evidence of grave goods disappears with the advent of the Younger Dryas and only reappears, and then infrequently, more than 2000 years later during the Late Pre-Pottery Neolithic B period, when agriculture was better established. Differential burial practices in this period have been interpreted as evidence of limited forms of social differentiation among individuals and households. However, skeletal evidence on nutrition is not sufficient to support inferences about inequality across or within settlements (Smith and Horwitz, 2007). There is general agreement on the lack of convincing evidence for the existence of hereditary elites in the Pre-Pottery Neolithic (Kuijt and Goring-Morris, 2002).

Again, our theory suggests an interpretation. With the return of more plentiful rainfall and warmer temperatures after the Younger Dryas, previously abandoned sites became viable once again and began to attract increased population. This is consistent with the observation that at Abu Hureyra, population declined immediately after the end of the Younger Dryas (Moore et al., 2000), as people presumably migrated elsewhere in the region. Although the arrival of agriculture created an impetus toward productivity growth, fully domesticated crops took a long time to emerge (Balter, 2007), and this productivity effect was partially counteracted by climate effects that tended to discourage concentration of population at the best sites within the region. This accounts for the long lag before the reappearance of elite-commoner inequality after the Younger Dryas, and the relatively weak nature of the social differentiation that did emerge.
The end of the Pre-Pottery Neolithic was associated with population dispersal, perhaps due to soil depletion or a climate shock, and this ended any movement toward greater inequality for about 2500-3000 years. Only in the Bronze Age do archaeologists see re-aggregation of population, the emergence of city-states, and widespread economic stratification.

**Polynesia.** This is an enormous region of the Pacific Ocean bounded roughly by the Hawaiian Islands, Easter Island, and New Zealand. The islands of this region differ widely in their sizes, climate, topography, soil fertility, and ecosystems. All Polynesian societies are descended from a common ancestral culture based on domesticated plants and animals, hunting, foraging, and fishing. Migration and colonization over thousands of years had generated a diverse assortment of social structures by the time of European contact. The resulting trajectories can be reconstructed from archaeology, linguistics, and ethnographic accounts. Polynesia therefore provides an ideal laboratory for investigating how the natural environment influences social evolution. All page references below are to Kirch (1984) except where indicated.

Polynesian societies are organized as 'conical clans', where social rank is based on genealogical distance from a founding ancestor. When migrants colonized a new island chain, population growth led to fission into subclans, where lower-ranking clans occupied economically self-contained but less valuable areas (29-33). In most cases, population densities were roughly at steady-state levels by the time of European contact.

Conditions at contact provide us with a cross-sectional sample where differences in productivity across islands should be associated with differences in population density and inequality. The degree of inequality varied widely. In the more egalitarian societies,
economic decisions were made by all adult males in village meetings, and food resources transferred to chiefs were largely redistributed back to the producing households. At the other extreme, there were hereditary divisions between land-owning elites and landless commoners. In these cases, the food resources transferred to chiefs were used to finance elite consumption, public works, political activities, and warfare.

An example involving moderate inequality is the island of Futuna (Hayden and Villeneuve, 2008). This is a small volcanic island that had a population density of 15-50 per km$^2$ before contact. Land was owned by elders in corporate descent groups and the most valuable agricultural land was owned by the highest-ranking elite families, with strong competition among local leaders to attract followers. It is unclear whether village elites benefited significantly from their positions due to obligations for the redistribution of resources to followers, although Hayden and Villeneuve believe that on Futuna and elsewhere in Polynesia, "considerably less was redistributed than was received" (41). In terms of our theory, this is a case where (a) there are insiders and outsiders at each village and (b) there is moderate stratification within villages.

Across Polynesia, there is a strong correlation between population density and degree of stratification. The most highly stratified island chains, defined as those with the sharpest divisions between landowning elites and landless commoners, are generally agreed to include Hawaii, Tonga, and the Society Islands, with Hawaii as the most extreme case. Sahlins (1958: 11-12) adds Samoa to this list, while Goldman (1970: 20-21) instead adds Mangareva. Of 10 island chains for which data are available, the four with the highest population densities per unit of arable land are Hawaii, Samoa, the Society Islands, and Tonga (Kirch, 1984: 37, 98).
Our theory leads us to expect that these will also be the islands with the best endowments of natural resources and thus the highest productivity of food labor. This expectation is confirmed. Sahlins (1958: 126-130) identifies Hawai‘i, the Society Islands, and Samoa as the cases with the best productivities (no data were available for Tonga), based on their wide range of crops, diversity of environmental zones, opportunities for land improvement, abundance of edible wild plants, abundance of fish, and reliable water supplies. Of these, Hawaii stands out as having the best resource endowment, and the next best below this top group is Mangareva. All nine of the other islands or chains examined had at least one serious environmental limitation, such as seasonal water shortages, vulnerability to drought or storms, poor quality soils, lack of environmental diversity, or limited fishing opportunities due to the absence of a barrier reef and lagoon.

Kirch (1984) endorses this assessment of relative productivities. He comments: "[T]he Tongan islands . . . are richly endowed with unusually fertile soils that permitted an intensive level of dryland agriculture" (221). The eight major islands of the Hawaiian chain are described as well-watered, with a high degree of environmental diversity and "vast tracts of fertile land" (245). "Opportunities for [agricultural] intensification were virtually unsurpassed within the Polynesian region" (282).

Some information is available on the time path that led to stratification for Tonga and Hawaii. There is little evidence of status differences in the early phase of settlement at Tonga. But by the time of European contact, "the emphasis on fencing of gardens and indeed the degree to which the island is carved up into walled parcels, bespeaks a permanency of land division" (223).
For Hawaii, colonization began at the best sites around 400 AD. The population slowly expanded into more marginal locations, and after about a thousand years virtually all lowland areas were occupied, including the most marginal. Densities in the most fertile areas became quite high, about 250 people per km$^2$. Earle (1997: 42) estimates that the Hawaiian population peaked at roughly 160,000 people around 1500 AD.

Linguistic evidence indicates that in the early stages, land was held by corporate descent groups. There is little indication of inequality before 1400 AD, but after that a few houses became much more elaborate and required considerably more labor to build. The construction of religious monuments increased dramatically (Earle, 1997: 44). By European contact, the population was sharply divided into a hereditary landowning elite and a class of landless commoners (Kirch, 1984: 257-258; for details on the economic organization of the Hawaiian islands in this stratified phase, see Earle, 1997: 75-89).

Taken as a whole, the Polynesian evidence fits our theory nicely. In the early stages of colonization when good land was abundant, there was little inequality. As population grew and good land became scarce, inequality arose between insiders and outsiders at favorable locations. On the best-endowed islands, continued population growth led to sharp stratification. These dynamics are consistent with our expectations (see the remarks on population at the end of section 4). The cross-sectional correlations among productivity, population, and inequality at European contact are also consistent with our expectations from section 4.

7. **Extensions of the Theory**

Having reviewed the empirical support for our theory, we conclude with some potential extensions. These include kinship, downward mobility, slavery, and warfare.
**Kinship.** A key observation is that people who share access to territory and keep outsiders from using it often have close kinship ties, either genetically or by marriage. Stratified societies generally have hereditary elites who claim rights to territory based on descent from a common ancestor and maintain careful records of genealogical relations within the elite. Proximity to a chiefly or kingly line is generally a source of enhanced status, while the ancestry of commoners is of little interest to anyone. Examples are provided by the Channel Islands, the Northwest Coast, and Polynesia (see section 6).

We suspect that such patterns arise because kinship groups have advantages in producing local public goods. This follows from familiar evolutionary arguments about altruism among genetic relatives. One such public good is exclusion of outsiders, and another is the control of commoners. In our model, groups linked by close kinship ties should have exclusion technologies with a larger feasible set $E$ (see the model in section 3). As productivity rises over time, such groups will pre-empt other groups by being the first to enclose sites of any given quality level. Once sites become stratified, the insiders will become a permanent hereditary elite.

**Downward mobility.** Our Malthusian framework implies that regional population typically increases as productivity increases, and that per capita food consumption for the region (averaging over elites and commoners) remains constant. However, the inequality between elites and commoners implies that these two groups will have unequal numbers of surviving adult children per person. Moreover, at sites of sufficiently high quality the elite must have more children than what is needed for demographic replacement. In this situation, a stable elite size can only be maintained if there is a mechanism that moves some children of the elite into the commoner class. There must be greater downward
mobility at higher-quality sites, because the elites at these sites are richer and therefore have more children per person.

Three mechanisms for implementing downward mobility that may be important in practice are primogeniture (the first born inherits elite status, others do not); proximity to a chiefly line of descent as a basis for preferential access to land (where the least related people drop out of the elite in each generation); and factional conflict within the elite for control over land. It is also possible that the elite may have a higher mortality rate than commoners due to greater participation in external warfare.

Although the need for downward mobility is most evident in stratified societies, the same problem arises when some sites are closed but none are stratified. Consider a hunter-gatherer society where bands control territories, and some of these territories are better than others. The resulting insider-outsider inequality means that the bands with the best territories have more children than are needed for replacement, while bands with poor territories or those reliant on the commons have fewer children than are needed for replacement. Unless there is some social mechanism for transferring people from rich bands to poor bands in each generation, insider-outsider inequality will tend to erode as population accumulates through natural growth at higher-quality sites. One mechanism that may play this role is fission of groups into subgroups when local population density becomes high, with one subgroup moving away to land of lower quality.

**Slavery.** In section 1, we emphasized that our model applies most naturally when commoners are free to leave stratified sites. But some sedentary hunter-gatherers such as those of the Northwest Coast did have slavery (Ames, 2008), and slavery is widespread among agriculturalists. We thus sketch an extension of the model to address this point.
Suppose a chief or an elite wants to constrain some commoners from exiting a site in order to give them a level of food consumption below the outside option \( w \). The slaves can still try to escape to the commons where \( w \) is available, but the probability that they will succeed is less than one. There is a positive probability of being punished or killed, which is worse than \( w \). Thus the expected payoff from fleeing is below \( w \), and the elite only needs to pay slaves this expected value in order to forestall escape attempts.

Whether it is profitable to use such a strategy depends on the cost of tracking down escaped slaves, the penalties that are imposed, and so on. The key tradeoff for the elite is between lower wages and higher enforcement costs. Although killing slaves who are caught might maximize deterrence, recaptured slaves are valuable and thus may be subject to lesser penalties. In practice there could be a spectrum of 'unfree' people who face varying probabilities of capture if they run away and varying punishments if caught. This implies a corresponding spectrum of living standards that such people would have to receive in order to deter escape attempts.

A crucial question in this story is whether slave owners (for example, neighboring chiefs or elites) collude in the capture and return of slaves, or compete with each other by keeping captured runaway slaves. This may be a matter of indifference to slaves (unless the neighboring chiefs will set them free), but it affects the profitability of slavery for the local elite. Collusion is always difficult to enforce when there is a benefit from cheating on the collusive arrangement, but repeated game equilibria may support cooperation in some cases. Of course, this problem tends to disappear if smaller chiefdoms are replaced by larger ones for reasons involving military technology and public finance. Other things
being equal, societies with centralized coercive power and control over large territories
should be most likely to rely on slavery.

**Warfare.** There is a strong empirical correlation between inequality and warfare
(illustrated by the Channel Islands, the Northwest Coast, and Polynesia in section 6). The
formal model in this paper only addressed exclusion of individual outsiders, not conflict
between organized groups. However, one can readily interpret the model as describing a
situation where defensive military technology trumps offensive technology, so groups of
insiders can always exclude groups of outsiders. Here we briefly sketch a richer analysis.

Every society probably has theft and retaliation against thieves. In a trivial sense
this behavior tends to create or redress inequalities because the motives are redistributive.
At a more organized level, small-scale societies often engage in raiding (for food, people,
or other valuables). These raids do not yield permanent control over land or other natural
resources, and any resulting inequalities are likely to be transient. We therefore consider
warfare motivated by a desire to expel an existing population from desirable territory, or
to rule over such a population.

One might think that such warfare requires land of heterogeneous quality, but at
least in theory this may not be necessary. Assume resources are uniformly distributed. If
every group initially has the same land to labor ratio, one group could try to push another
group off its land. If this ploy succeeded, the victors would have gained a higher ratio of
land to labor and thus a higher standard of living. But there are two practical difficulties.
First, if the groups are initially of equal size, it is unclear how one group can prevail over
the other. Second, a higher land/labor ratio implies lower population density, which may
make it impossible to defend the group's newly enlarged territory. Warfare thus appears
more likely in a world of heterogeneous site qualities, where groups with inferior land but large populations may be tempted to move up the quality ladder. Of course, this requires a story about why larger populations occupied lower-quality territories in the first place.

Although empirically inequality and warfare are strongly correlated, at least in some cases it seems that inequality arose first. In southwest Asia, evidence of inequality, which is inferred from mortuary practices and the existence of public buildings, is dated well before the first evidence of inter-village conflict, which is inferred from defensive walls (Bar-Yosef, 2002).

Inequality also seems logically prior to warfare in two ways. First, there is little reason to bear the costs and risks of an attack if one does not expect to retain possession of conquered land for a reasonable amount of time. If conquerors must immediately face attacks from other groups, warfare will be unprofitable unless defensive technology is at least somewhat reliable. From an analytic point of view, it thus makes sense to start with the simplest case where defensive technology is always reliable. In effect, this is what our inequality model assumes.

Second, there is little reason to bear the costs of an attack unless one expects to get a sizable benefit in return. As noted above, holding both one's current land and the new land one wants to acquire may not be feasible. Thus attacks are most likely to be motivated by a desire to abandon one's current land and take over someone else's. This motive can easily arise in a world with insider-outsider inequality, and the theory in this paper can be used to determine the potential payoff from such a strategy. Similarly, with stratified sites one elite may seek to displace another elite, leave existing commoners in place, and appropriate the associated land rents. This motive is likely to arise for elites
that have inferior sites but more military power than the target elite. Again, it would be necessary to explain how such a situation arose in the first place.

For real military technologies, defense does not always trump offense. Moreover, temporary shocks to population or resources may increase the probability of a successful attack or decrease its costs (including opportunity costs) for a coalition of outsiders or a rival elite. As inequality grows, the temptations for such attacks become greater. In the case of stratification, we can also expect factional struggles within the elite over the rents from land ownership. Again, the model in the present paper provides a framework within which one can define the payoffs from such strategies.
References


Ames, Kenneth M., 2008, Slavery, household production, and demography on the southern Northwest Coast: Cables, tacking, and ropewalks, Ch. 6 in Catherine M. Cameron, ed., *Invisible Citizens: Captives and Their Consequences*, University of Utah Press, Salt Lake City, 138-158.


Bar-Yosef, Ofer and Richard H. Meadow, 1995, The origins of agriculture in the near
east, Ch. 3 in T. Douglas Price and Anne Birgitte Gebauer, eds., Last Hunters, First
Farmers: New Perspectives on the Prehistoric Transition to Agriculture, School of
American Research Press, Santa Fe, New Mexico, 39-94.

Boix, Carles and Frances Rosenbluth, 2009, Bones of contention, presented at Santa Fe
Institute Workshop on the Emergence of Hierarchy and Inequality, February

Bowles, Samuel and Jung-Kyoo Choi, 2002, The first property rights revolution, Santa Fe
Institute, October 28, presented at Workshop on the Co-evolution of Behaviors
and Institutions, SFI, January 10-12, 2003, available at

Butler, Virginia L. and Sarah K. Campbell, 2004, Resource intensification and resource
depression in the Pacific Northwest of North America: A zooarchaeological
review, Journal of World Prehistory 18(4), December, 327-405.

Carneiro, Robert L., 1970, A theory of the origin of the state, Science, New Series,
169(3947), August 21, 733-738.

Cohen, Mark Nathan, 1989, The evidence of prehistoric skeletons, Ch. 7 in Health and
the Rise of Civilization, Yale University Press, New Haven, 105-129.

Cohen, Mark Nathan, 1998, The emergence of health and social inequalities in the
archaeological record, Ch. 14 in S. S. Strickland and P. S. Shetty, eds., Human

Cohen, Mark Nathan and Gillian M. M. Crane-Kramer, 2007, Editors' summation, Ch. 22
in Mark Nathan Cohen and Gillian M. M. Crane-Kramer, eds., Ancient Health:


Kennett, Douglas J., Bruce Winterhalder, Jacob Bartruff, and Jon M. Erlandson, in press, An ecological model for the emergence of institutionalized social hierarchies on


Sahlins, Marshall, 1958, *Social Stratification in Polynesia*, University of Washington Press, Seattle, WA.


The Exclusion Technology

For A Production Site

**Figure 1**
Continuous Population

When Site Changes From Open

To Closed At $z = A$

Figure 2(a)
Discontinuous Population From \( c^{-}(A) \) to \( e^{A} \) When Site Changes From Open To Closed At \( z = A \)

Figure 2(b)
Productivity And The Wage As Functions Of $x = \frac{\theta}{w}$

$[x_1 < x_2; \text{see text for cases (i), (ii), and (iii)}]$

**Figure 3**
Maximum Wage $W(\theta)$
For Each Productivity Level

Figure 4