A THEORY OF STATE FORMATION AND THE ORIGINS OF
INEQUALITY †

Carles Boix ‡

February 9, 2012

† This paper includes chapters 1 and 2 of the book manuscript “Political Order, Inequality and Growth”. (*) Previous versions of the material employing to write this paper were presented at the Juan March Institute, University of Virginia, the New York meeting of the “September Group” Princeton University, Duke University, Yale University, Stanford University, the University of Chicago, Harvard University, Cornell University and the University of Maryland. I thank the comments of all their participants, in particular to those of Alicia Adserà, Sam Bowles, Dan Gingerich, Stathis Kalyvas, Carol Mershon, Thomas Romer and Milan Svolik.

‡ Princeton University. E-mail: cboix@princeton.edu.
1

EQUALITY AND THE STATE OF NATURE

To examine the emergence of inequality and the foundations of political order, that is, how political authority formed *ex nihilo*, Chapter 1 engages in a thought experiment, modeled after modern political or state-of-nature theory, that starts by assuming a world characterized by the most sparse initial conditions: one where a set of individuals, endowed with identical resources (time and perhaps some innate abilities), interact among themselves in the absence of any political institutions or state, that is, without any organization having the monopoly of force over them. Under such a condition of anarchy, described in Section 1, individuals rely on two different strategies to survive and prosper. They may follow a productive strategy, which consists in the allocation of their resources to the production (and voluntary exchange) of goods and services. Alternatively, they may adopt an expropriatory or extractive strategy, that is, they may direct their resources to appropriate the assets or returns of other individuals in a forcible manner.

Chapter 1 then shows (in Section 2) that, even when those individuals have incentives to follow an expropriatory strategy in independent, one-shot interactions, they may cooperate ‘spontaneously’, that is, without any permanent political institutions, over time. However, the outcome of spontaneous or self-enforced cooperation is only feasible if the resources of all those agents remain similar, that is, if there are no changes in the initial condition of equality that characterizes the state of nature. If the distribution of resources, and the stream of income each one receives, shifts in favor of part of the population, the cooperation equilibrium breaks down. Moreover, even though equality remains central to that result, time horizons, predation technologies and mobility turn out to have important effects on the maintenance of spontaneous cooperation.
Cooperation may take place, as discussed in Section 3, in the form of two settings: a world in which atomized individuals live in isolation while tolerating each other and, more plausibly, a system in which individuals live together in a stable society. The foundations of the latter may be emotional (such as the ties that bind parents and children together) and/or material (taking advantage of production complementarities and sharing risks in a volatile environment). Yet, the conditions spelled in Section 2 must still be present in all those cases to elicit cooperation in the absence of a third party enforcing it.

The chapter concludes with two additional sections that marshal a wide array of ethnographic and archaeological sources on foraging groups with no permanent political institutions to probe the theory. Section 4 shows that human communities with no political hierarchies exhibit similar interpersonal patterns of consumption, high levels of wealth equality, very limited levels of intergenerational transmission of (non-genetic) assets. Section 5 explores the rather numerous strategies that individuals employ in an active and systematic manner to sustain some fundamental homogeneity among themselves and therefore to uphold the basis of cooperation. Those practices, which range from mocking others to ostracizing and even killing them, are conducted to preempt any individual deviation from the cooperation equilibrium and the eventual formation of an elite. The examination of those strategies brings in two important insights. First, it explains why growth and development cannot happen under conditions of strict equality: to last in a condition of spontaneous cooperation, the group needs to preclude anyone from innovating, hoarding and so on. Second, it offers a theory of social behavior that accounts for the fact that primitive societies tend to exhibit high levels of both random and individualized violence (including very high homicide rates) – something that anthropologists and ethnographers have not been able to explain with a coherent theory – and discusses how those levels of violence are compatible with, that is, do not lead to the demise of, the spontaneous (i.e. non-institutionalized) maintenance of order.
INITIAL CONDITIONS

To understand the dynamics that led from equal, primitive economies to complex, unequal historical societies, consider a world populated by two types of agents, $i$ and $j$. These two agents may be thought of as individuals living together or as individuals (or households) populating some contiguous territory. They may also be thought of as representative individuals of homogeneous groups, that is, groups formed by agents who, because of their close ‘resemblance’ (in a sense to be made more precise shortly) to each other, can cooperate in a stable, permanent manner.

Initially, all individuals are equal. They have the same labor endowment or available time $L$, which they can allocate between two different types of activities. On the one hand, they can devote some fraction of their time $L$ to a ‘productive’ activity, that is, to the extraction or generation of output (or income) from some resource under their control – e.g. hunting or gathering in some circumscribed area. Formally, agent $i$ will receive some income $r_i$ according to a standard Cobb-Douglas production function $r_i = A_i T_i^\alpha L_i^{1-\alpha}$ where $A$ denotes the state of technology (such as, in this initial world, their natural predispositions or abilities hunting, fishing, etc.) and $T$ stands for a fixed stock of land. (The same notation here used, for the sake of brevity, for $i$ applies as well to $j$.)

On the other hand, $i$ (as well as $j$) may direct some portion of his time to predation that is, to steal the output of the other agent. Say that the fraction of time spent on predation is $\lambda$ (with $0 \leq \lambda \leq 1$). As a result, $i$ will obtain (from looting) $\lambda^\omega (1-\omega) r_j$, where $r_j$ is the return from the productive activities of individual $j$; $(1-\omega)$ is the maximum part of $j$’s output that can be appropriated through looting, with $\omega$ denoting the portion that the looted individual $j$ can shelter.

---

1 For previous work modeling the economic activity of individuals as a choice between production and predation, see Pareto (1966 [1902]), Haavelmo (1954) and Hirschleifer (1994, 1995).
from expropriation through a set of strategies (to be discussed later in the chapter in more detail) such as consuming the producing output immediately or escaping from the looter; and $\theta$ is a technology parameter ($\theta \geq 0$) that measures the effectiveness of the time that $i$ allocates to extraction. Since the parameter $\theta$ is shaped by the predation abilities of both individuals, we can think of $\theta$ as the ratio between the effectiveness of $j$’s predation technology and the effectiveness of $i$’s predation technology. Accordingly, as $i$’s technology of looting improves, $\theta$ moves closer to 0: $i$ needs to spend little time predating to maximize the income obtained from looting. Notice that, since the effectiveness of $j$’s predation technology is the reverse of $i$’s, $j$’s technology parameter can be denoted as $1/\theta$.

The overall income of individual $i$ would be, in principle, that is, in a decision-theory setting:

$$y_i = (1-\lambda) \cdot r_i + \lambda^\theta (1-\omega) r_j = (1-\lambda) \cdot A_i T_i^\alpha L_i^{1-\alpha} + \lambda^\theta (1-\omega) A_j T_j^\alpha L_j^{1-\alpha}$$

where, again, $\lambda$ (with $0 \leq \lambda \leq 1$) is the fraction of time $i$ spends looting.

To simplify the analysis, assume that $T$ and $L$ are identical across all agents. Next define

$\hat{A}_n = A_n T_n^\alpha L_n^{1-\alpha} / A_j T_j^\alpha L_j^{1-\alpha}$ so that $(A_i T_i^\alpha L_i^{1-\alpha} / A_j T_j^\alpha L_j^{1-\alpha}) = \hat{A}_i$ and $(A_j T_j^\alpha L_j^{1-\alpha} / A_j T_j^\alpha L_j^{1-\alpha}) = 1$. The value of $\hat{A}_i$ will then indicate the ratio of income (resulting from production) between individual $i$ and agent $j$.

We can rewrite $y_i$ as:

$$y_i = (1-\lambda) \cdot \hat{A}_i + \lambda^\theta (1-\omega) \quad (1.1)$$

To determine the strategies that each agent will follow, i.e. the allocation of their time $L$, consider now the payoffs that accrue to $i$ and $j$ when they choose their course of action simultaneously. First, if both agents simply produce ($\lambda_{ij} = 0$), their payoff will be $\hat{A}_i$ and 1 for $i$ and $j$ respectively. Second, if both sides loot, and assuming that, under a condition of generalized looting, all output not directly consumed by each agent will be lost or destroyed, agent $i$ will be only able to retain the portion she produces directly or $(1-\lambda) \cdot \hat{A}_i$. In turn, $j$ will only keep $(1-\lambda)$. 
Finally, if agent $i$ loots while agent $j$ does not, the looting agent receives an income $y_i$ as defined in function (1.1) above. In turn, and given that the remaining income of the non-looter (the part produced by $j$ and not expropriated by the looter) will be reduced by some proportion due to the destructive effects of pillaging, $j$ will obtain $(1-\beta)(1-\lambda)$, where the parameter $\beta$ ($0 \leq \beta \leq 1$) measures the destructiveness of predation. The same result (now with the payoffs reversed) applies if it is $j$ who loots and $i$ who produces. While $j$ gets $(1-\lambda) + \lambda \omega (1-\omega) \hat{A}_i$, $i$ keeps $(1-\beta)(1-\lambda) \hat{A}_i$. The set of strategies and payoffs is represented in Figure 1.1.

**Spontaneous Cooperation under a Condition of Equality**

We are now in a position to examine the circumstances under which both parties will settle in a cooperation equilibrium in which both choose to produce and no one devotes any time to predation. To make the analysis realistic, let us assume that both $i$ and $j$ interact in an infinite, repeated game, with some discount rate $\delta$ that, for the time being, is identical for all players.

To make the analysis simple (and focused on the problem at hand), assume next that both players have two strategies available to them. They can follow a trigger strategy PP, producing always until looting occurs and then responding with looting ever after. Alternatively, they can adopt a constant looting strategy LL.$^2$

To have sustained peace, the following conditions should take place. For player $i$, the value of choosing a production strategy must be larger than looting a producer in the first period and then facing a looting solution always ever after. Formally:

---

$^2$ Since $(1-\beta)(1-\lambda) < (1-\lambda)$, just producing when looted is not a feasible option. The notation and solution in this subsection partly follows Dixit (2008). On a discussion of the existing literature on cooperation in the context of infinitely repeated games with a focus on experimental evidence, see Dal Bo and Fréchette (2009).
\[
\frac{\hat{A}_i}{1-\delta} > (1-\lambda)\hat{A}_i + \lambda^\theta (1-\omega) + \frac{\delta}{1-\delta}(1-\lambda)\hat{A}_i
\]

Rearranging terms:
\[
\hat{A}_i > \frac{\lambda^\theta}{\lambda}(1-\delta)(1-\omega)
\]

(1.2)

Similarly, for player \(j\) the following inequality (where producing always is better than looting the producer and then moving into a permanent situation of conflict) must hold:
\[
\frac{1}{1-\delta} > (1-\lambda) + \frac{\lambda^\theta}{\lambda}(1-\omega)\hat{A}_i + \frac{\delta}{1-\delta}(1-\lambda)
\]

Hence:
\[
\frac{1}{\hat{A}_i} > \frac{\lambda^{1/\theta}}{\lambda}(1-\delta)(1-\omega)
\]

(1.3)

Putting (1.2) and (1.3) together, we can conclude that continuous production will be an equilibrium if:
\[
\frac{1}{(1-\delta)(1-\omega)} \frac{\lambda}{\lambda^{1/\theta}} > \hat{A}_i > \frac{\lambda^\theta}{\lambda}(1-\delta)(1-\omega)
\]

(1.4)

The equilibrium condition expressed in (1.4) shows that, even in a Hobbesian world, that is, in a world where there is no authority in place, it is possible for individuals to sustain peace, avoid conflict and engage in productive activities – again without having to resort to any centralized mechanism of coercion employed to refrain them from looting. This is a result well known in the formal study of human cooperation (Axelrod 1984).

More importantly, the expression (1.4) defines the conditions under which a non-predatory equilibrium is possible absent any formal institution or third party guaranteeing it. Above all, cooperation is only sustainable if there is some fundamental equality of income (derived from productive activities) among individuals or across human communities. To see why, consider a world in which agent \(i\) becomes much richer than agent \(j\) and \(\hat{A}_i\) rises to the point
that the left-hand side in (1.4) does not hold any longer. Agent $j$ has no incentive then to produce and shifts to a looting strategy. The outcome of spontaneous order simply unravels. The same result takes place if $j$’s production returns are much higher than $i$’s. In this instance, as $A_i$ becomes much lower than 1, the right-hand side in expression (1.4) does not hold and agent $i$ is the one who shifts away from the production strategy, triggering a situation of generalized looting.

Historically, the condition of equality seems to have prevailed in primitive foraging societies, which were predominant across the world from the emergence of mankind to the introduction of agriculture, for two main reasons. First, their technology of production was simple and accessible to everyone. Second, the marginal product of land was relatively similar: partly because population densities were low and foragers tended to exploit territories of similar value (in terms of the quality of soil, weather conditions, availability of water, etc.); partly because their technological know-how limited their ability to exploit the land beyond hunting and gathering. In line with the conclusions of the model, foragers maintained a system of social cooperation without relying on state coercion.

In the following sections I explore empirically the relationship between having limited or bounded inequality and maintain a production equilibrium. Before that, however, I examine in the remainder of this section the ways in which the outcome of spontaneous cooperation depends on three additional factors: $\delta$ or the discount rate; $\omega$ or the proportion that cannot be expropriated by the looter; and $\theta$ or the efficacy of the looting technology itself.

**The Shadow of the Future**

If players put little value on future payoffs, i.e. if the discount rate $\delta$ is close or equal to 0, they will only cooperate with each other if their incomes are very similar. Take, for example, the expression in (1.2), which characterizes $i$’s incentives to cooperate. If the discount rate declines
(while holding the other two parameters, $\omega$ and $\theta$, close to 1), the value of the right-hand side comes close to 1. It only takes a small decline of $\hat{A}_i$ below unity (implying that $j$ is more productive than $i$) for the mathematical inequality not to hold any longer. However, as soon as the discount rate rises, the output of $i$ ($\hat{A}_i$) can increase, that is, there can be more inequality, without the cooperation equilibrium.

The weight that individuals give to future income may be in part the result of personal idiosyncratic factors. However, the discount rate will be affected, in a more systematic way, by the size of the population of the group within which all the interactions take place and by the correlated probability with which players may interact with the same partners over time. Suppose a world in which individuals are paired off with each other through a random mechanism every day. As the number of players increases, the likelihood that the same pairing of individuals will occur over time declines. The growth of the population of players undermines the expectations of future exchanges that constrain the behavior of individuals. The shadow of the future declines and cooperation is less frequent.\(^3\) In fact, the formal literature has repeatedly shown that the possibility of cooperation and the size of the group are negatively related (Axelrod 1984, Olson 1993).\(^4\)

**The Technology of Predation**

The possibility of cooperation is also a function of the kind of technologies of predation available to individuals $i$ and $j$: if their technology of predation is similar, then $\theta$ fluctuates around $1$. The agents’ incentives to cooperate are strongly defined by the extent of inequality. However, as one of the agents becomes a relatively more effective looter, the temptation to

---

\(^3\) For a similar point, see Fearon and Laitin (1996), p. 718.

\(^4\) For neurobiological evidence showing that the size of the neocortex limits the ability of individuals to keep track of social interactions, see Dunbar 1992.
renege from cooperation rises. Take, again, expression (1.2) and examine the behavior of individual $i$. As the looting technology parameter $\theta$ goes up, the value of the right-hand side increases – unless $i$ is very productive ($A$ is large enough), he will have an incentive to plunder $j$.

In a primitive world where sophisticated weapons are absent, all individuals are roughly similar in strength. Since the risk of injury and death in any conflict among equals is very high, men attack each other by surprise, taking full advantage of striking the opponent first. As noted by ethnographers for primitive societies, “most serious attempts at killing and most killings are done when the victims of the attack can be caught helpless (…) and, above all, little capable of effectively harming the attackers” (Gat 1999: 564) and engage in strategies similar to the ones recorded for nonhuman primates such as chimpanzees, who organize in groups to isolate and attack single individuals (LeBlanc 2003: 80-81). The production of violence is a highly time-intensive activity. Given that all production activities (hunting and fishing) were also labor-intensive, any time devoted to extraction implied a proportionate (linear) decline in direct output. In such a world, $A$ (as well as $\delta$ and $\omega$) determined the extent to which “spontaneous cooperation” was possible. However, as soon as $\theta$ moves away from 1, that is, as soon as the marginal return to production and looting started to diverge across individuals, the production equilibrium became increasingly fragile – a topic I fully explore in Chapter 2 in the context of the process of state formation.

**The Exit Option**

Finally, the incentive to loot other is shaped by the strategies producers themselves employ to reduce the value of what a looter can steal. This reduction is embodied by the parameter $\omega$, which may be affected by at least three factors. First, as soon as output can be stored, the size of $\omega$ declines or, in other words, the value of a potential plunder increases for the looter. Hence, for example, any new process to conserve food should increase the likelihood of
conflict, other things being equal. Second, it is affected by the patterns of consumption of producers: if the producer consumes everything immediately after obtaining it (for example, by swallowing all the strawberries as soon as she picks them), predation will be less common.

More interestingly, looting and conflict are a function of the ease with which producers can move away in response to any potential aggression. When exit costs are low (essentially because production does not change in the new territory to which producers go), the incentives to plunder are also low: as soon as the looter threatens to plunder, the producer moves away, effectively making $\omega$ equal to 1 and the final value of output equal to 0. As a matter of fact, we know that foraging groups often split into smaller subgroups in response to internal conflict (Hirschman 1981; Chagnon 1983) and that they bear high levels of rotation (mostly following interpersonal conflicts and disagreements) at the individual level (Leacock and Lee 1982).

However, exit is only feasible, at least in a primitive world, if population densities are low and if the marginal productivity of land is similar across territories. As soon as territories are heavily populated, moving into a new area becomes extremely costly. Similarly, if there are high discontinuities in output per unit of land, producers have a much lower incentive to abandon the area they live in response to a threat of looting.

The dynamics of population growth, conflict and exit were probably a main driver of the gradual expansion of mankind from territorially concentrated areas to the whole globe since the late Pleistocene. As the number of humans grew, partly as a result of some natural drift and partly as a way to secure new producers as well as future care-takers for the elderly, conflict within their communities must have increased: as noted in the previous discussion on the determinants of the discount rate $\delta$, free riding becomes easier and cooperation harder as the size of groups grows. In that sparsely populated world, in which land with a high marginal value is still available, the strategy of exit was the most rational way to manage conflict and any competition for resources. Over time, however, as land filled up, as migrating became more costly, the
relative value of looting increased. Under the latter circumstances, the condition of strict equality became the key to maintain the outcome of non-predation without any political institutions.

**The Nature of Cooperation**

So far I have employed the concept of cooperation (as opposed to predation) in a rather loose way. In fact, I have simply defined the production equilibrium by exclusion: as that set of social interactions where no systematic looting occurs. However, the state of cooperation may take several forms. In the most stripped-down interpretation of the model, it may consist of a set of isolated individuals, wandering alone in the forests, satisfying their needs on the spot, and, crucially, avoiding a state of war with each other – simply because, in line with the conditions spelled out above, economic resources are uniformly distributed, individuals are roughly similar in physical strength, and the costs of exit are low). On the other hand, it may imply a tighter conception of social interaction in the form of maintaining some kind of stable human community where men and women live together, engage in joint economic activities, and share the latter’s yield.

The first type of state of nature, akin to Rousseau’s vision of the first men as a savage men, “solitary, idle and always near danger”, does not square well with the evidence produced by anthropological research and, more recently, by evolutionary biology theory. Human beings appear to have lived in families and bands always, even though these groups tend to exhibit some considerable level of fluidity in primitive societies. Social units, from households to bands, are partly glued together by biological as well as emotional ties (Ingold 2000). Nonetheless, the existing literature explains human communities, especially at the band level, as the outcome of

---

5 On human sociality as a universal condition from an anthropological point of view, see Levi-Strauss (1961 [1955]), pp. 389-392, 397-398. For a review from the point of view of evolutionary biology, see Hanson (1992), Hawkes (1992) and Boone (1992).
two main types of two material calculations. On the one hand, it claims that individuals live together to minimize variance in income through sharing. On the other hand, it argues that they form a group to exploit the complementarities and economies of scale that come from joint production.

Because the model developed in this chapter is centrally concerned with the conditions that sustain a production equilibrium without any political enforcement, it says nothing about the specific factors that may lead to either a world of isolated individuals or one composed by stable human communities. Still, regardless of what may be the precise conditions that explain the formation of social groups (before the emergence of state institutions), it is important to emphasize that the condition of relative equality among individuals (necessary to have some spontaneous cooperation) underlies the two explanatory models of risk-minimization and joint-production gains.

**Equality in Risk-sharing**

In any models of group formation based on the idea of risk-sharing and risk-minimization, individuals join in a stable community and share their production over time to smooth their income over time and to shelter themselves from the possibility that some random shock leaves them below a survival threshold. This solution, which allows them to maximize the total expected value gained from exploiting the territory they live in, is only feasible if the variation in output is both stochastic and asynchronic across individuals. Notice that these two conditions imply the premise of interpersonal equality: if shocks are random and the independent probability of them occurring is similarly distributed across agents, then all individuals have the same expected output in the long run. It is then that long-run equality of income that allows them to sustain a reciprocal relationship of cooperation over time – at least without formal political institutions. In contrast, as soon as some individuals have a consistently higher yield than others,
the flow of resources among them becomes asymmetrical and conflict takes place (Smith 1987, Boone 1992).

Production Complementarities and Equality

The condition of equality is equally present in any model that relies on the gains that come from production complementarities or from economies of scales to explain group formation or complementarities. To see this, let us employ the model introduced in this chapter to examine the behavior of an individual j under two alternative scenarios: one where there are no gains from producing jointly with others (and which corresponds to the formal structure in the section on “Spontaneous Cooperation and the Condition of Equality”); a second one, where producing jointly with others leads to a higher output than just producing alone.

Consider first the case in which there are no complementarities. Under this scenario, j experiences some technological shock that allows her to expand her output by some amount \( \Delta \) from 1 (the baseline case we examined before) to 1+\( \Delta \). Under this shock, j gets 1+\( \Delta \) if she produces all the time. If she loots, she gets \((1-\lambda)(1+\Delta)+\lambda(1-\omega)A\) in the first period and \((1-\lambda)(1+\Delta)\). (To simplify the analysis, I assume, without any loss of generality, that \( \theta=1 \).) Hence, she will produce if:

\[
\frac{1+\Delta}{1-\delta} > (1-\lambda)(1+\Delta)+\lambda(1-\omega)A_i + \frac{\delta}{1-\delta}(1-\lambda)(1+\Delta)
\]

Simplifying:

\[
\frac{1+\Delta}{A_i} > (1-\delta)(1-\omega)
\] (1.5)

In the second case, the same technological shock only benefits j if both parties, i and j, produce. Formally, j obtains 1+\( \Delta \) if she produces all the time. However, if she now loots, there
are no joint production gains. She gets \((1-\lambda) + \lambda(1-\omega)A\) in the first period and \((1-\lambda)\) afterwards.

Hence, \(j\) will only avoid looting the other side if the following inequality holds:

\[
\frac{1 + \Delta}{1 - \delta} > (1 - \lambda) + \lambda(1 - \omega)\hat{A}_i + \frac{\delta}{1 - \delta}(1 - \lambda)
\]

Simplifying:

\[
\frac{1 + \Delta/\lambda}{\hat{A}_i} > (1 - \delta)(1 - \omega)
\]

Since \(\lambda \leq 1\), the numerator in the left-hand side of (1.6), \(1 + \Delta/\lambda\), is larger than the one in (1.5), \(1 + \Delta\), for any \(\lambda \neq 1\). The joint production gains, captured in the difference between (1.5) and (1.6), results in a more robust self-enforcing production equilibrium. In other words, if producing jointly with others increases the income of producers, the latter will have a smaller incentive to renege from cooperation. However, even under those circumstances, there still is a threshold beyond which inequality leads to the breakdown of a cooperation equilibrium.

\section*{A Nested Model of Cooperation}

Although the model has been employed to explain the behavior of individuals toward each other, it can be thought of as providing the building blocks of what may be labeled as a ‘nested’ theory of cooperation. At the individual level, peace has been shown to be feasible among a subset of actors provided they enjoy sufficiently similar payoffs (as well as a similar distributed of predation technologies). When they do, they can form a stable, orderly tribe, village or neighborhood. The same principle can then be applied to predict whether human bands or tribes may sustain cooperative relations among themselves: they will if their output and capacity to loot are similar and/or the possibility of exit has low costs.

Social cooperation, understood as a stable production equilibrium, breaks down at the point in the continuum of agents where inequality of conditions is too high and expression (1.4)
does not hold anymore. That continuum of individuals can be understood along two dimensions: first, going from individuals up to broader social groups; second, in the territorial disposition of humans.

**EMPIRICAL EVIDENCE ON STATELESS SOCIETIES**

To assess the relationship between equality of conditions and the mechanisms that sustain social cooperation, I focus here on those communities that the literature describes as “simple foraging societies”, that rely on the activities of hunting and gathering and that lack any clear, established authority structure and that have no stable organization (or group of individuals) holding the monopoly of violence in a given territory. According to existing archaeological material, from the emergence of modern humans in the late Pleistocene and until the Mesolithic, all mankind lived in foraging communities that presented little internal social and economic differentiation.6

Notice that the universe of cases to be examined does not contain two types of communities that are closely related to them but still rather different in their internal economic and social structure. On the one hand, it does not include “complex” foraging societies—mainly the fishing communities of the Northwest Pacific Coast, because they show considerable internal social differentiation and a vertical structure of political authority. As discussed in Chapter 2, those two traits were related to a pattern of strong spatial concentration of resources (such as salmon in the American Northwest) in very specific territories that encouraged the emergence of monopolies over those enclaves and that led, in line with the theory developed so far, to the

---

6 According to Hayden (2001), a possible exception is a few Paleolithic communities in Southern France about 30,000 years ago.
construction of coercive practices and institutions. On the other hand, it excludes those horticultural societies that retain some equal distribution of resources and have no political hierarchies for two reasons. First, it avoids examining cases that may be contaminated by the problem of control and intergenerational transmission of land plots inherent in communities that have started to use agricultural technologies. Above all, the exclusion makes good on the starting point of the theoretical model (a world where there is hardly any interpersonal differentiation) since horticultural societies are the result of some technological progress in previously more simple communities.

To examine the nature of simple foraging communities, we need to rely on two kinds of empirical evidence: archaeological material collected for the Paleolithic; and ethnographic studies of contemporary stateless societies. Both types of evidence have important limitations. The archeological data gives us information about the material conditions (nutrition, habitation, levels of storage and intracommunal distribution of resources) of mankind at a time when everyone lived in communities that were close to the initial conditions of the model. But it does not provide any direct evidence on the cultural practices and political interactions (directed to the creation of social cooperation) of those (arguably stateless) societies. Contemporary ethnographic reports, from travel diaries written by sixteenth and seventeenth century explorers and missionaries to contemporary physical anthropology, supply us with more direct and richer data.

---

7 It was correlated with the construction and ownership of sophisticated equipment, such as complex fish traps, which involved considerable labor (Rowley-Conwy 2001).
8 As a result of the influence of evolutionary biology, the behavior of nonhuman primates has been increasingly brought into the anthropological literature to shed light on the material and social conditions of “primitive” man (see, among many others, Janson 1992 and De Waal 2000). Although I will refer to some of that data (particularly to examine the problem of human violence), it cannot be employed systematically for two reasons: primates and human beings differ along several dimensions (such as linguistic ability,prehensility, etc.); comparisons across social states are impossible to make because primates have always lived in stateless communities.
on “primitive” societies. Still, using ethnographic evidence to make inferences about the nature of “original man” has been criticized on two main counts. The first critique points to the fact that today’s simple societies have been affected by their contact with more “complex” societies – minimally by the entry of researchers; more decisively, by trade flows, the importation of tools and ideas, etc. Although the influence of more developed neighbors is undeniable, two considerations seem to justify the analogical use of contemporary foraging societies. On the one hand, there are strong continuities in the behavioral patterns observed both at the very initial phase of contact (such as the dietary of Jesuit Paul Le Jeune, who lived with the Montagnais in 1633-34 (Le Jeune 1973)) and more recent anthropological studies. On the other hand, there are substantial similarities between the material evidence (on tool-making, storage or even proportion of violent deaths) collected from prehistoric sites and contemporary anthropological data. The second critique is that contemporary simple societies cannot be taken and examined as “original” communities in the sense of being direct descendants of pre-modern foraging bands. Instead, they are said to be once-complex collectivities that more successful or powerful neighbors pushed back into marginal habitats such as dense forests or the tundra. However, this claim is arguable. We know that some simple societies (such as the Eskimo (Rasing 1994) or the Aboriginal Australians (Lourandos 1985)) have occupied the same habitat for very long periods of times. In addition, the claim that simple foragers occupy marginal lands seems unsupported by the evidence. Once one excludes foragers living in very cold weather areas, the annual net primary productivity (or amount of new plant growth in grams/m2) is similar in foraging and agricultural habitats (Marlowe 2005: 60).

9 The debate on whether current ethnographic evidence may be employed to study past foraging communities was particularly intense in the eighties. See O’Connell, Hwakes and Jones (1988) and Headland and Reid (1989) and more recently Marlowe (2005) and Roscoe (2009).

10 For a careful analysis of all the sources on the Iglulingmiut Eskimos since contacted by the first European expeditions in the 1820s, see Rasing (1994).
Equality of Material Conditions

Table 1 provides some statistical information on foraging and non-foraging communities using data taken from the “Standard Cross-Cultural Sample” (SCCS), which is a subsample of 186 societies taken from the broader “Ethnographic Atlas”, which includes 1267 societies coded by George P. Murdock. After being developed by Murdoch and White (1979) and then progressively expanded in the following decades, the SCCS includes around two thousand variables. Simple foraging communities are small in size: the median size of foraging communities is between 50 and 99 individuals (and 44 percent of them have less than 50 individuals). Their median population density, at 0.2 individuals per square mile, is extremely low. They tend to follow some nomadic or semi-nomadic habitat pattern: less than 8 percent of them live in permanent settlements.

Foraging communities are overall rather equal. Material differences in food consumption or in the possession of material goods are either absent or very limited. Since food storage among foragers is very low and tools are rather primitive, asset accumulation (and, therefore, intergenerational transmission) is either absent or minimal.\textsuperscript{11} Individual contributions to production often differ: some individuals are more successful than others and middle-age individuals are net calorie contributors in the band. Indeed, band members track individual success in hunting (Hawkes 2000) and grant some social recognition to the good hunter (Turnbull 1965; Kaplan & Hill 1985). But foraging communities have strict social norms that prevent hunters from effectively owning the meat they have hunted. As Wiessner (1996) notes in a comparative study of twenty-seven foraging societies, even in those communities where the

\textsuperscript{11} The only exception takes place in communities living in very cold weathers, where food storage is feasible and substantial. However, the group tends to impose tight form of control over everything that has been processed and stored.
hunter may own the meat, ownership “means little more than having the right to distribute the meat or to decide who should do, for all foragers have rulers stipulating that meat must be widely shared” (177). Moreover, any status recognition given to the hunter is extremely transient (Erdal and Whitten 1994), partly because there is too much variability in hunting performance (Hawkes 2000), partly because, as detailed later, foragers spent a considerable part of their time leveling down any individuals that attempt to assert themselves over the others. As a result, food and, particularly, meat distribution is strictly egalitarian (except, in some communities, for very large preys) (Hawkes 2000) and differences in nutrition over the life cycle are minimal (at least within each gender) (Kelly 1995).

Foraging communities show moderate levels of polygyny. According to Table 1, about a third of simple foraging bands (as opposed to over half in non-foraging societies) have a significant fraction of men married to more than one wife. But even in those cases, the proportion of polygamous men in each foraging group is less than twenty percent. Having two wives is often the maximum with the exception of Australia, where polygyny is related to gerontocracy. In most cases, polygyny seems to be unrelated to wealth or hunting success (Table 1). Using individual data from the Hadza, Marlowe (2000) finds that better hunters do not have more wives and that there is no evidence that husbands and wives are assortitavely mated with respect to foraging returns. Good hunters seem to have more offspring but mostly because they have a higher chance of marrying young women upon divorce. Given that the yield of hunting is shared equally, it is unclear why women would enter into polygamous relationships: whereas Marlowe (2003) suggests that polygamy minimizes the amount of transfers (in the form of vegetables gathered) each woman will have to make to men, Divale and Harris (1976) and Otterbein (1994) relate it to sex imbalances due to inordinate levels of violence among males.

[Table 1 around here]
Recent work compares the distribution and intergenerational transmission of wealth in 21 foraging, pastoralist and agricultural communities employing individual data (Smith et al. 2010). The study distinguishes between three types of wealth: embodied, which includes such things as longevity, physical strength, hunting skills, etc.; material, which refers to tools, land, etc.; and relational, measured through reproductive success. Except for the first type of wealth, inequality is extremely low among the foraging communities of the Ache, Hadza and !Kung. The intergenerational transmission of wealth, including things such as the transmission of hunting skills from fathers to sons, is also low. The only partial exception takes place for traits that can be explained in strict genetic terms, such as body weight. By contrast, material wealth is unequally distributed in both pastoral and agricultural societies (with Gini indexes close to those of highly unequal modern countries) and its intergenerational transmission is also sticky (Borgerhoff-Mulder et al. 2010; Shenk et al 2010). Although those results are based on a rather limited number of societies (for example, they are only three simple foraging societies and the data is less than systematic), they confirm the existing conventional wisdom about the relative equality of conditions in foraging groups. According to the data in the SCCS, 73 percent of all simple foraging societies have an egalitarian social structure (versus 26 percent of the other communities) (Table 1). In addition, 89 percent and 37 percent have no inheritance rules on real property and movable property respectively. In non-foraging societies, the corresponding proportions are 27 and 9 percent.

The existing archaeological evidence for Paleolithic sites seems to confirm similar patterns of nutritional equality as well as little material differentiation for primitive foraging communities (Formicola 2007; Vanhaeren & D’Errico 2005). A similar pattern of equality is

---

12 For an opposite view, see Hayden (2001), in turn criticized by Kelly (2010).
systematically described in all the ethnographic reports written at the time of initial contact of European travelers with pre-agrarian societies (Le Jeune 1973; Rasing 1994).

**Political Life**

In line with the model, stable authority structures (with a set of individuals systematically controlling the use of violence) are absent in simple foraging communities. That does not mean, however, that there is no political life (Clastres 1974). Rather, the latter takes a rather loose, horizontal pattern. Decisions affecting the band are arrived at through discussions in which all adults participate. Those discussions do not necessarily take place through any formal mechanisms or at specific times: they are often embedded in the context of continuous conversations and gossip (Turney-High 1971: 61-73; Silberbauer 1982). Even when there are some patterns of leadership, leaders have no formal authority and act as mere referees among different individuals or families or spend substantial time negotiating with, cajoling or persuading other members of the tribe (Clastres 1972; Lee and leacock 1982; Lee 1982). And although threats and actual violence are certainly employed within the tribe, they happen in a non-institutionalized manner (Clastres 1972, 1974; Kelly 2000). Finally, the use of exit is quite widespread across foragers -- particularly in the form of individuals often rotating between different but nearby bands. As such, it is part and parcel of the overall system of collective decision-making within the band because it restricts the possibility that certain cliques form permanently within the band and it forces the latter to seek some wide consensus in making decisions that affect everyone.  

---

13 A large literature on the structure of political life among Australian Aborigines generally concludes that, at least in non-riverine or coastal communities, it was “characterized by a uniform distribution of rights, privileges, and duties throughout a social order based on kinship and suffused by an egalitarian ideology”
SHORING UP THE PRODUCTION EQUILIBRIUM: THE CREATION OF SOCIAL CONFORMITY

The material conditions and underlying dynamics of simple foraging societies match the main features of the model developed in the first section of the chapter. The situation of “spontaneous” social cooperation among individuals is underpinned by their equality in material conditions and social status. And the latter is turn generated by a relatively uniform distribution of resources (in conjunction with a primitive technology of production).

Still, the outcome of “non-state politics” (to employ Clastres’ terms) is also the result of deliberate and painstaking efforts made by everyone to maintain a state of peace among individuals. In the repeated prisoner’s dilemma game that underpins the result of cooperation discussed above, the production equilibrium is just one of two possible equilibria. If any of the individuals deviates from producing, the outcome becomes one of systematic violence, looting and exploitation – a possibility that can have catastrophic consequences for everyone, especially given the constraints often imposed by the kind of ecological environment in which foraging societies live. Thus, pre-state communities need to (and actually do) spend considerable efforts at making the production equilibrium robust. They engage in a vast array of social practices, including many ritualized actions, to equalize the material outcomes and social status of individuals. They develop a comprehensive set of sanctioning mechanisms, from gossiping to physical injury and execution, to secure the compliance of their partners. And they promote, through the upbringining of children and story-telling, a public ‘morality’ that emphasizes the primacy of equality and the subordination of individuals to the collective needs of the group.

(Hiatt 1987: 177). Although senior males enjoyed some specific authority in the sphere of religion, it was not translated into other domains. See, as well, Sharp (1958) and Meggitt (1966). For a dissenting position, Elkin (1964). Strehlow (1970) admitted the presence of hierarchies for ritual practices but did not not seem to have any spillover effects on any other social domains.
That overarching ideology of equality (and the practices it sustains) should not be interpreted as a set of ideals flowing from an ethical commitment to some principle of fairness but rather as the result of the choices made by self-interested individuals in a state of nature. Nineteenth-century scholars writing on hunters and gatherers, including Marx and Engels, who had been influenced by the speculative historical work of Lewis Morgan, tended to be sparse in their description and analysis of the “savage man”. Still, they believed that the collective ownership of land and tools characteristic of “primitive communism” made human exploitation impossible, at least outside the domain of the household, and hinted at the idea that, in living as equals, foragers could cultivate an ethic of mutual aid, conviviality and friendship. Twentieth-century cultural anthropology has instead described social life among foragers as one based on a pattern of “generalized reciprocity”, that is, as the giving and taking based on diffuse obligations to help others who may be in need, independently of the existing balance of transfers between individuals (Sahlins 1972).

That anthropological literature said little about the specific conditions that make the pattern of sharing over time feasible. But there should be little doubt that the system of generalized reciprocity is underpinned by a long-run tit-for-tat game conducted among self-interested individuals (Hawkes 1992, Boehm 1993). Reviewing multiple studies on foraging societies, Erdal and Whiten (1994) conclude that cooperation and order in pre-state communities result from the interaction of “self-interested individuals whose motivations include a strong desire to get enough for themselves coupled with a strong desire to make sure that no one else gets more than they do” (177). Those self-interested motivations are apparent in daily life.

obsequiously, and occasionally snatching pieces of meat”. Similar evidence on a conflictual process of the distribution of food, involving arguments and clear signs of jealousy and envy, has been reported among Zaire pygmies (Turnbull 1965), the Eskimo (Briggs 1970) and the Kalahari San (Tanaka 1980). Writing about the latter, Wiessner (1982) notes that “[it cannot] be argued that !Kung have little interest in material possessions. Most possessions … are highly desired; and many !Kung are really torn between the desire to accumulate goods and the desire to remain within a secure system of mutual help” (81). Turnbull (1965) indicates that “it would be a rare Mbuti woman who did not conceal a portion of the catch in case she was forced to share with others” (198). More recently, an experimental study conducted in fifteen small-scale societies shows that societies that are less integrated in modern market structures (with foraging groups falling into that category) systematically exhibit higher levels of selfish behavior and fewer instances of cooperative strategies (Henrich et al. 2005). Having to deal with a social setting where individuals have a strong temptation to free-ride, it is not surprising to learn that foragers spend inordinate numbers of hours (through conversations and gossiping) to track the precise obligations of every individual toward all the members of the group, as attested by Wiessner’s (1982) study on the xaro or pooling risk mechanism among the !Kung.

The countervailing forces of all those interests then result in a cooperative equilibrium. W.E. Parry, in command of an expedition to find the Northwest Passage to the Pacific, had to spend the fall and winter of 1822-23 living with the Inuit of Igloolik Island, located in the Foxe Basin. Parry reported that resources were shared openly and equally among the Inuit, who would never be “eating with closed doors” and would always welcome all visitors. Generosity and gratitude had nothing to with their behavior, however. “Any thing like a free gift”, wrote Parry, “is very little if at all known among them. If A gives B a part of his seal to-day, the latter soon returns when he is a successful fisherman (…) The regulation does credit to their wisdom, but has nothing to do with their generosity” (quoted in Rasing 1994: 20). Peace was ubiquitous: their
passions were completely restrained; conflict was extremely fleeting; and “in their transactions among themselves there is no doubt that (...) the strictest honesty prevails” (ibid. p. 23). In a word, the inhabitants of Igloolik conducted all the relationships under what Parry defined as “the tacitly-received law of mutual forbearance” (ibid. p.24). But that law strictly relied on an iterated exchange of favors among selfish individuals. That much seems clear from what happened to the old and weak, who could not comply with the demands the cooperative outcome imposed on every person. They were simply abandoned: “[to the] old people, whose age or infirmities render them useless and therefore burdensome to the community, the Esquimaux betray a degree of insensibility bordering on inhumanity, and ill-repaying the kindness of an indulgent parent” (ibid. p.20). Clastres (1972) found similar practices among the Guayaki. At one point, female widows were expelled from the community and left alone to fend for themselves in the deep forest. Their expectation was that they would be taken care of by the spirit of jaguars. And whenever they came back alive to the band’s camp, their sons-in-law would be in charge of leading them back into the forest and of taking care of their mothers-in-law directly, usually by hitting them from behind.  

15 Across all simple societies, from the Inuit Eskimos to the !Kung in the Kalahari, the ideology and practice of equality turns around avoiding and denouncing two capital sins: stinginess, that is, “to hoard one’s goods jealously and secretively, guarding them ‘like a hyena’” (Lee 1982: 52); and arrogance, which “is actually dangerous, since according to the !Kung

15 To account for cooperation in general, Fehr and Gächter proposed an alternative model, based on the idea of “strong reciprocity”, according to which individuals “repay gifts and take revenge even in interactions with complete strangers and even if it costly for them and yields neither present nor future material rewards” (2000: 159). In a detailed study of norm enforcement among the Ju’hoansi Bushmen, Wiessner (2005) shows that the tit-for-tat model prevails in foraging communities: the costs of punishing are low on a daily basis and well spread out among the members of the band. That result is not unsurprising given the small size of the group and the constant, very open set of interactions in which all of them engage with each other.
“[someone’s] pride will make him kill someone” (Lee 1982: 54). To do so, foragers engage in rather ritualized social practices around the processes of production and consumption directed at leveling down the successful hunter. Among the !Kung, for example, the latter rejoins the band announcing the yield of the hunt modestly, almost in a hush, since the community, which abhors any boasting, expects the hunter to belittle his own skills. Upon his entry, the band systematically downplays his merits through teasing and mocking, stresses the randomness of the success, and “emphasize[s] his duty to provide meat for others” (Wiessner 1996: 179). 16 Such social treatment extends in fact to all areas of collective life. Any individual deviation from sharing and, as a matter of fact, from the expected average behavior, is quickly noticed and then hammered out through mocking and shaming. Similarly, co-members show little patience toward any unilateral strategy of accumulation. As soon as someone becomes more successful, the rest engage in what some have labeled as “institutionalized stealing”: they beg the former, constantly and tirelessly, to follow the proper behavior of sharing. Among the !Kung, individuals who try to work hard and even attempt to hoard some material possessions “come under more and more pressure to give them away in hxaro, and finally give in and redistribute them” (Wiessner 1982: 82).

The strategies to create conformity do not stop at deriding and begging others. Foraging communities ostracized deviators in numerous ways. Some of them are of rather low consequence such as avoiding conversations, distributing the worst morsels of food, withdrawing specific attentions and spreading damaging gossip (Briggs 1970). But other forms include the expulsion of the individual from the group, corporal punishment and death. Among New Guinea’s Gebusi, individuals deemed to deviate from the boundaries of appropriate behavior are prone to accusations of employing black magic and witchcraft. Once the band accepts the charges

---

put forward by an individual accuser, their members are given free rein to kill the person that has
been accused (Knauft 1987). Balicki reports that excessively aggressive individuals and
“dangerous sorcerers [that] became old and bitter, and took to performing malevolent magic
against even their close relatives” (1970: 189), chapter 9) were killed with the consent or active
collaboration of the community at large among Netsilik Eskimos. Lee (1979) offers similar
evidence for the !Kung.

The suppression of individual differences is not done only through external mechanisms.
In some communities it also operates through the construction of the inner motivations of every
member of the group. Relationships between parents and children are hardly authoritarian and the
latter normally enjoy a very playful infancy. But children are brought up in the expectation that
they will uphold the tight social norms of equality and conformity in place. The Inuit invest
considerable time instilling ilira, that is, a feeling of uneasiness about (or fear of) others, among
their children. This sentiment, that encompasses both an “awareness that people could be kind
and caring but also malignant and aggressive” (Rasing 1994: 110), plays a key role in producing
order. As Briggs reports after living for almost two years among the Uthuhikhalingmiut Eskimos,
“‘a person who does not feel ilira’, people explained, ‘will impose himself on others’; he will
intrude their autonomy, be demanding, and may attack or even kill, because ‘he feels strong’”
(1978: 65-66). With ilira regulating social relationships, it follows that affection is too dangerous
because those who love and care (too much) can also aggress and so create fear” (Rasing 1994:
113). Emotions are suppressed and any external sign of agitation strongly condemned (Briggs
1970: 235-298). Rasmussen, traveling through Arctic Canada in the late 1920s, wrote that “the
mind of the Netsilingmiut is like the surface of the many lakes that are spread over their country;
it is quickly set in commotion but the waves never become deep swells and the water quickly
comes to rest again (...) it is a point of honor with them to preserve their equanimity (...) they
have the happy gift of being able to rest content with the knowledge of sorrow; they know that they have suffered a lot but do not become emotional.” (quoted by Rasing 1994: 114).

**Stability and Violence**

A situation of spontaneous or self-enforced cooperation should not be equated with the absence of violence. It simply implies that a particular group of individuals live in common in spite of the occurrence of violent acts between individuals and that they do so without resorting to the construction of a state.

As a matter of fact, violence, at least the one that is exercised with lethal consequences, is very high in pre-state societies. The annual homicide rate, which stands at around 20 to 30 individuals per 100,000 in American inner cities, has been estimated to approach around 300 individuals per 100,000 among Agta Pygmies, !Kung and Semai and over 400 per 100,000 among the Copper Eskimo and Papua-New Guinea’s Gebusi (Knauft 1987; Headland 1989; Gat 1999). About 25 percent of adult males die from a violent death on average in foraging societies. That proportion matches well the percentage of buried skeletal remains from the Paleolithic showing traces of violence (Kelly 2000: 155).

Lethal violence in foraging communities follows a very specific pattern. It is both irregular in its occurrence and apparently random in its causes. Ethnographers agree on the fact that daily life in foraging communities is marked by a strong level of familiarity and joviality. At the same time, however, they report that such peaceful existence becomes sometimes interrupted by sudden, unexpected and uncontrollable explosions of individual anger. The enraged individual, who instants before was sustaining an amicable conversation with other persons, loses his or her temper often to the point of exercising violence in an unrestrained manner and with lethal consequences. Moreover, that lack of control frequently makes the aggressor unable to punish the person that potentially led to the confrontation. Lee (1979: 397) reports that in over
half the cases of lethal violence among the !Kung the victim was not a part in the argument leading to the homicide.  

Those outbursts of rage frequently result from trivial, almost childish disputes, related to issues of minimal instrumental value to both parties, at least in light of the human costs that follow from the dispute (Knaupf 1991: 400). When trying to move beyond the immediate motivation of a particular conflict, the anthropological literature on violence has failed to develop a single, unified account of the reasons behind the occurrence and nature of lethal violence in primitive societies. There is some common agreement that it is mostly unrelated to any struggle for material goods or political control. Male competition for females explains violence partially. But its rate of explanatory success varies rather substantially across groups (Knauft 1991). It cannot account either for the killings of alleged sorcerers among the Gebusi, which, in fact, “seldom correspond[ed] to any outstanding grievance between the kin of the killers and the homicide victim” (Knauft 1987: 464-465).

The punctuated and relatively arbitrary nature of violence in foraging societies fits well, however, with the main features of a model of self-enforced cooperation. The production equilibrium is based, once again, on selfish individuals that are highly suspicious of the behavior and intentions of all others. Their solution to that external threat, which they recognize as taking place also within themselves, then consists of exerting an inordinate level of repression on everyone: from their material conditions to their external demeanor and from their social conversations to their inner thoughts. As Briggs writes pointedly, Inuits make sure that their “the expression of hostility is very stringently controlled”. But his makes it very likely that “people will project their own suppressed hostility onto others and, knowing that they themselves feel violent, will easily suspect others of murderous intents” (1982: 116). In this context of permanent

17 Turnbull (1965), Briggs (1982) and Dentan (1979) report similar behavior among the Mbuti, Inuit and Semai respectively.
restraint, collective vigilance and mordant humiliation before the community, personal freedom does not exist. Violence (at least the one that the group sanctions a priori) then becomes the unintended consequence of repression – the last resort to protest, even if unconsciously, against the lack of a private sphere of life. Without a third-party enforcer imposing sanctions in a predictable manner, the occurrence of violence is extremely high.

Foragers’ high homicide rate does not jeopardize social stability, however. A murder committed by a single individual is hardly interpreted as an injury to a whole group, it rarely escalates into a general conflict pitting one household against the other, and it does not lead to the collapse of social interaction within a given human band. As Kelly (2000) notes, “the violence that occurs is specific, not generalized, and it does not escalate beyond a sequence of events that encompasses homicide followed by execution of the killer. Typically, a murder is an isolated event with no sequel” (2000: 42-43). Since the cost of generalized violence may be too high, the community often intervenes to stop any revenge that could lead to a chain reaction – certainly the case of the Netsilik Eskimos, who knew very well that “every murder signified the loss of a highly needed seal hunter” (Balikci 1970: 182). If generalized violence becomes unavoidable, however, most foraging bands manage it through a process of fission: a section of the group secedes and migrates to a new location, where it self-organizes again along the standard patterns of relative equality and non-institutionalized politics.¹⁸ Naturally, the possibility of exit is circumscribed by the level of population density. In his study of foragers in the Andaman Islands during the nineteenth century, Kelly shows that warfare among bands appeared in a “circumscribed environment where maximum population density [had] been attained (as evidenced by periodic short shortages)” (2000: 105).

¹⁸ For an analysis of the exit option in the horticulturalist society of the Yanomanos, see Chagnon (1997). For a broader conceptual analysis of the effect of strong exit options on income distribution and political structures, see Hirschman (1981), chapter 10.
TABLE 1. SIMPLE FORAGING COMMUNITIES.

<table>
<thead>
<tr>
<th></th>
<th>Simple foragers</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population density -- persons per square mile (median)</td>
<td>0.2</td>
<td>1-26</td>
</tr>
<tr>
<td>Number of individuals in settlement (median)</td>
<td>50-99</td>
<td>100-199</td>
</tr>
<tr>
<td>Percent living in permanent settlements</td>
<td>8.8</td>
<td>65.1</td>
</tr>
<tr>
<td>Percent without social stratification (percent)</td>
<td>73.5</td>
<td>26.3</td>
</tr>
<tr>
<td>Percent with no inheritance rules on real property</td>
<td>89.3</td>
<td>26.8</td>
</tr>
<tr>
<td>Percent with no inheritance rules on movable property</td>
<td>37.0</td>
<td>8.8</td>
</tr>
<tr>
<td>Percent with monogamy or few cases of polygyny</td>
<td>50.0</td>
<td>30.1</td>
</tr>
<tr>
<td>Percent polygyny unrelated to status, rank</td>
<td>56.3</td>
<td>36.6</td>
</tr>
<tr>
<td>Number of observations</td>
<td>34</td>
<td>152</td>
</tr>
</tbody>
</table>

Own calculations based on the Standard Cross-Cultural Sample.

FIGURE 1.1

Agent j

<table>
<thead>
<tr>
<th>Produces</th>
<th>Loots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Produces</td>
<td>1</td>
</tr>
<tr>
<td>Agent i</td>
<td>$(1-\beta) (1-\lambda) \hat{A}_i \hat{A}_j \theta (1-\omega)$</td>
</tr>
<tr>
<td>Loots</td>
<td>$(1-\lambda) \hat{A}_i + \theta (1-\omega)$</td>
</tr>
<tr>
<td></td>
<td>$(1-\hat{A}_i)$</td>
</tr>
</tbody>
</table>

32
Chapter 1 showed that human cooperation is feasible without any need to establish an institutionalized authority to enforce peace – provided that individuals are roughly equal in both economic resources and fighting abilities. Such a state of social cooperation may take place through two alternative forms: in the form of individuals (or, at most, households) living separately and avoiding conflict among themselves; or in the form of a community of people living together and sharing in the processes of production and consumption to either minimize risks or exploit some complementarities in the process of production. Either way, however, life under a situation of spontaneous cooperation is poor, nasty, brutish and short. Men and women live at the margin of subsistence, almost always exclusively preoccupied with satisfying their most elemental material needs, and their private behavior and conversations are strictly fettered by a community that punishes any deviation from the socially expected standard of behavior with ostracism and sometimes even with death.

As explored in the first section of this chapter, the state of spontaneous cooperation tends to breakdown over time. In the process of production, individuals innovate: they perfect or invent new technologies to exploit the natural resources or territories they control. As soon as that new know-how is unequally distributed among individuals or affects the productivity of territories differently, the foundations of the “state of nature” break down. Inequality sorts agents out into different economic types or ‘classes’. Some individuals are more productive and therefore richer than others. That shift in economic fortunes makes them have opposite strategies. Say, for example, that the returns to production of agent $i$ increase relative to those of agent $j$. The
economically advantaged agent $i$ still has an interest in sustaining a production equilibrium. In fact, $i$’s cooperation incentives become, if anything, stronger. For a sufficient high level of inequality, agent $j$ will instead prefer plundering $i$’s output than producing. In a world populated by some individuals whose dominant strategy is looting, the outcome of self-enforcing cooperation breaks down, giving way to a Hobbesian world of systematic conflict.

The only solution to generalized disorder consists in the construction of an organization or structure that has both the incentives and the capacity to enforce order among both producers and looters. The second section of this chapter examines the two alternative paths through which political authority may be born. On the one hand, producers themselves can coordinate to defend themselves against plunderers. On the other hand, looters can set up a system in which they restrain themselves from looting the producers once and forever. Instead, they collect some permanent stream of income from the producers. To secure those rents, looters have to defend their producers (now turned into “taxpayers” of some sort) from other. The first type of solution leads to a republican compact. Republics generally preserve the existing distribution of wealth and income that results from the structure of ownership and the control of technology in the new political community in place. The second form of government is a monarchical regime. Monarchies lead to a substantial redistribution of income, from producers to looters, who have now become the lords of the former.

As examined in the third section of this chapter, warfare and war technology turn out to be the main driver behind the distribution of monarchies and republics across the world and over time, at least before the industrial revolution. The invention and spread of warfare techniques that give some comparative advantage to looters lead to the expansion of monarchical regimes. By contrast, those technologies of fighting that equalize resources across individuals or that even confer some additional strength to producers (naval power in commercial republics is one such example) increase the likelihood of republican survival. However, as section four shows, even
when military technologies were favorable to republican regimes, the management of warfare posed specific political risks and economic costs that tended to curb the expansionary capacity of republics. All in all, sections three and four explain why, at least until the twentieth century, republics tended to be territorially smaller, shorter in duration and much more infrequent than monarchies around the world.

**LEARNING-BY-DOING, INEQUALITY AND SORTING**

Before we explore the political consequences of the breakdown of cooperation in the state of nature, we should examine the sources of the rise of inequality. In fact, this section sketches a minimal theory of growth that we would come to in later chapters of the book.

Recall that in the initial world described in Chapter 1 all individuals, endowed with some time $L$, inhabited some territory $T$ which they exploited employing some technology $A$ – according a production function $r_i = A_i T_i^\alpha L_i^{1-\alpha}$. Since the marginal productivity of $T$ was assumed to be identical and the state of technology was common to all, all the individuals had the same income per capita. We then argued that such a state of affairs matched some initial stage of mankind’s history when human beings had the same primitive tools and occupied lands and forests of comparable quality. As a result, output per unit of land, which eventually determines the sustainable level of population, was equally low across the board. That hypothesis seems to be borne out by the evidence. Even though we lack any reliable data on production and income for that period, if we assume that in pre-industrial societies all production gains were absorbed by population growth instead of leading to a higher output per person (Kremer 1993, Clark 2007), we can use population density as indicative of the level and territorial distribution of production. Indeed, both archaeological and genetic evidence suggest that population densities were very low across the world before the agricultural revolution – at less than 0.01 persons per square kilometer in Late Paleolithic Europe (Bocquet-Appel et al. 2005) [REF Africa, Asia, America.]
Given that initial state of the world, inequality could only have come into place through two mechanisms: as a function of some variation in the marginal productivity of \( T \) such that some areas, which we will refer to as “core” territories, yielded a higher output than other, which we can call “periphery”; and as a result of a biased change in the kind of technology \( A \) in the hands of individuals.\(^{19}\)

Let us examine the latter channel first. In Chapter 1 there was no growth – the knowledge parameter \( A \) (as well as \( L \)) was treated as fixed. However, it seems plausible to think that, as individuals make tools and engage in their foraging activities continuously, they naturally develop new abilities and economic technologies to improve their production process. In other words, they accumulate new knowledge through a process of learning-by-doing (Arrow 1962, Lucas 1986). Then, as the state of technology \( A \) increases gradually, land (the other main factor of production besides labor) becomes more productive.

Initially, the effect of knowledge generation had no impact on the existing equality of conditions in prehistoric communities. The rate of knowledge accumulation was probably the same across all individuals for at least one of the two following reasons. First, given a roughly similar genetic endowment as well as the same level of effort and land input, the creation of new knowledge through learning-by-doing should have hardly differed across individuals. Second, even in the event that some individuals were more inventive than others, innovations probably spread across communities rapidly given that, one, there were no state institutions that could protect inventors and, two, exit and mobility across communities seemed to have been quite

\(^{19}\) Inequality could also result from differential population growth rates across territories. Given the same level of output per square mile, slower demographic growth should lead to higher per capita output. However, there is little evidence that population growth rates differed systematically across similarly endowed regions. Moreover, betting on lower rates of population growth does not seem rational: in the long run, and provided that military technologies were similar across the world, such a society would be at a disadvantage vis-à-vis its neighbors.
common in primitive societies. Existing archaeological evidence shows that the first
technological “revolution”, the making of high-quality stone tools, occurred about 40,000 years
ago. Genetic research seems to indicate a corresponding jump in population at around the same
period (Rogers and Harpending 1992, Marth et al. 2003) – a development that must be related to
the increased efficiencies in hunting and gathering that using those new tools. Still, those
transformations did not change the economic and social structure of human communities.
Technological differences across communities were very minor. All the existing archaeological
evidence culled from inhabited sites and burial remains points to the maintenance of very
egalitarian bands before the Mesolithic -- with the still debated exception of a few locations in
southwestern France (Hayden 2001: 234-240; ).

Over time, however, the same process of learning-by-doing resulted in the invention of
tools and the development of production techniques that altered the marginal productivity of land
differently across territories in a significant way. Economic resources, or more precisely, the
value of those resources became concentrated in dense and predictable patches and therefore the
returns to individuals varied depending on the area they occupy (Boone 1992). Such as a shift in
the resources gradient of territorial resources occurred for the first time among those foraging
communities, such as the populations on the Northwest Pacific Coast, that developed complex
fish traps which allowed them to exploit abundant riverine resources strongly concentrated in
some specific spots. Nevertheless, that process accelerated once plants and animals started to be
domesticated from 9,000 BC onward (Bellwood 2005, Barker 2006). The agricultural revolution,
which began in about six well-defined regions across the world, expanded slowly (and sometimes
not at all) to the rest of the planet. Population densities, which track output per unit of land,
would then began to diverge considerably: while tropical and tundra areas remained below 1
person per square mile, densities in some areas of the Middle East and China grew quickly to
over several dozens of individuals per square mile (Kuijt 2000).
In addition to being the consequence of the differential impact of A conditional on the underlying quality of land, the emergence of inequality derived mainly from a shift in the resource gradient of territories (rather than from a shift within a particular group) for the following additional reason. Since equal societies invest heavily in the maintenance of social conformity and of equality, any change away from the status quo (towards higher internal differentiation) can hardly happen within a well defined community. The goal of many social interactions in an equal community is to suppress any technological transformation that may jeopardize the equilibrium of self-sustaining cooperation. Therefore, change can only take place by either making sure everyone shares in the change equally or by overcoming the opposition of that community. The latter strategy requires that the innovation is applied in a separate location.

As pointed above, inequality may also result from variation in the quality of T (holding A constant, that is, abstracting from the effects of any process of technological innovation). A change in the value of land or, in other words, the emergence of some variation in the availability of game and edible plants across the world, would have happened in the following way. Initially, human populations, by definition small, limited themselves to occupy the most productive lands. As population grew, and absent any significant technological change, mankind had to occupy less productive territories (at least up to the point when fighting for the existing good land was less costly). The spread of the population would then lead to the emergence of a gap between the initial very productive core and the periphery.\(^{20}\)

---

\(^{20}\) Positive population growth may be the result of at least three factors: some natural drift; some fertility overshooting in a context of a volatile environment that encourages parents to have more children to compensate for sudden climate shocks (Jones 1981); the production of children in the context of a competitive scenario with other bands in which having more men gives an advantage in any future potential conflict.
There is no doubt that mankind went through a process of expansion and of occupation of lower-quality lands. That historical process may have explained, in turn, some of the variance in output and welfare across human communities. Still, the story of population growth (in interaction with a varying marginal productivity of $T$) cannot account for a central feature of human history – namely, technological progress and growth. Unless there is technological innovation, output per capita among those that control the best lands does not change. But the historical record shows that both output and population density expanded in absolute terms in certain parts of the world. In other words, a theory of growth through learning-by-doing is more convincing because it explains both average technological progress and its divergent effects across individuals and territories.

**FROM CONFLICT TO POLITICAL ORDER**

Overcoming a generalized situation of looting requires the construction of a “state”, that is, the creation of some organization or structure with the capacity to control the monopoly of violence and therefore to impose order in a given territory. Now, given the nature of the conflict and the two classes of individuals (producers and looters) involved, establishing a state may take place through two alternative paths—with each institutional solution embodying the interests of each type of individuals respectively: a monarchical settlement and a republican compact.

In a monarchical regime, the potential looters or bandits, that is, those individuals who have not benefited from the economic shock, govern ‘natural’ producers, that is, those that have benefited from growth. Under this political solution, the natural producers, who devote themselves entirely to a productive strategy, transfer some part of their output (generally in the form of direct labor, some tribute or lump-sum payment) to the agents that govern them. If that permanent transfer is sufficiently large (i.e. it exceeds the value of plundering the producers and
destroying their incentives to produce), the potential looters protect the natural producers from their own violence and against other potential bandits.

Alternatively, the ‘natural’ producers may decide to spend some fraction of their time to fighting, setting up a defensive structure and producing weapons to deter any potential plunderers. Under this scenario, producers double as rulers. To fund and manage their defensive structures, producers need to establish some governing institutions. Those institutional bodies may take different forms – an elected leader, a governing committee, a general assembly or a mix of all of them. These institutions monopolize the exercise of violence among natural producers (and any subjected looters) with two goals in mind: reducing the costs of coordinating all the citizens of the political community; and, more importantly, guaranteeing that no part of the members with the right to participate will exploit the rest of the polity. As a result, they constitute a state – in this case, of a republican kind.

The monarchical solution has received considerable attention in the literature of state formation to the point of dominating it – most recently, in the work Olson (1993, 2000), where the emergence of the state is traced back to the decision of “roving bandits” to become “stationary bandits” or landlords (Olson 1993, 2000). In contrast, the republican path to state formation has been hardly explored by the modern institutionalist literature. This disregard for a “self-government” solution is wrong on two counts. From an empirical point of view, it ignores the existence of republican states such as the Greek polis or medieval and modern Europe city-states. From a theoretical point of view, it makes any theory of state formation inconsistent. Olson (2000: 28 ff.) claims that liberal institutions appeared when there was a vacuum of power

---

21 Writing from an anthropological perspective, Carneiro (1970) also offered a theory of state formation that emphasizes the interaction of violence and exit options (the latter in turn deriving from population density). Exit strategies and density are, however, not endogenized in his paper. As in Olson, he focuses on the emergence of autocratic states. See Wright (1977) for a review of state formation theories in anthropology and archaeology.
in an already-formed state in which a set of notables had roughly equal claims to power – such as England in the second half of the seventeenth century. However, since this state of things, i.e. the existence of a balance of power, can be equally predicated for stateless societies, one can think of cases in which, at least hypothetically, a non-monarchical state is set up ab initio.

**Monarchical Solution**

Let us examine in more detail the monarchical solution in two steps: first by looking at the payoffs of the agents and then by solving the conditions under which both parties agree to live under a monarchical arrangement. Throughout the analysis, we will assume that the biased technological shock benefits individual $i$ to the point that individual $j$ has now incentives to become a looter.

Under a monarchy, the payoff of producer $i$ will be $(1-\varepsilon) \hat{A} / (1-\delta)$, where $\varepsilon$ denotes the transfer extracted by agent $j$, with $0 \leq \varepsilon \leq 1$, and all time is now devoted to production. In turn, agent $j$, who has shifted from looting to non-looting strategy (which now includes an effort to protect individuals of type $i$ against other potential bandits), receives a transfer from the producer in proportion to the time $\lambda$ spent governing or $(\varepsilon - \gamma) \hat{A}/(1-\delta)$, where $\gamma$ is the portion that the bandit-lord spends in governing and protecting the producer and where $\varepsilon \geq \gamma > 0$. To the extent that $j$ spends some time producing directly, his total payoff will be $[(1-\lambda)+(\varepsilon-\gamma)\lambda \hat{A}] / (1-\delta)$.

Given the structure of payoffs, in turn, $j$ will have an incentive to act as a monarch and not as a bandit if:

$$\frac{(1-\lambda) + (\varepsilon - \gamma) \lambda^{1/\theta} \hat{A}}{1-\delta} > (1-\lambda) + (1-\omega) \lambda^{1/\vartheta} \hat{A} + \frac{\delta}{1-\delta} (1-\lambda)$$

Simplifying:

$$\varepsilon > (1-\delta)(1-\omega) + \gamma$$

(2.1)
This implies that any potential looter will be more likely to become a monarch as soon as the extraction rate increases and the costs of governing decline.

In turn, natural producers will only accept to transfer some resources to a tyrant if that makes them better off than looting:

\[
\frac{(1-\varepsilon)\hat{A}}{1-\delta} > (1-\lambda)\hat{A} + \hat{\lambda}\theta (1-\omega) + \frac{\delta}{1-\delta} (1-\lambda)\hat{A}
\]

Rearranging the inequality:

\[
\lambda - \frac{(1-\delta)\hat{\lambda}\theta (1-\omega)}{\hat{A}} > \varepsilon
\]  
(2.2)

A monarchical regime becomes more attractive to producers under three circumstances: the level of extraction falls; their looting capacity (measured through \(\theta_L\)) declines; and their total output \(\hat{A}\) increases.

Putting together (2.1) and (2.2), a monarchical settlement will be possible whenever:

\[
\lambda - (1-\delta)\hat{\lambda}\theta (1-\omega) \frac{1}{\hat{A}} > \varepsilon > (1-\delta)(1-\omega) + \gamma
\]  
(2.3)

In addition to (2.3), a monarchical settlement will only be feasible if peace without transfers is not an equilibrium.\(^{22}\)

Contrary to a contractarian interpretation of that political solution (North 1973, Levy 1988), a monarchical regime should not be thought of as a contract made among equally free individuals who negotiate some extraction rate \(\varepsilon\) as well as some level of government \(\gamma\). Instead, it is a political equilibrium in the sense that no party has any incentive to deviate from it. Looters restrain themselves because governing makes them better off than pillaging. Natural producers subject themselves to a bandit because they cannot deter the latter from threatening them with

\(^{22}\) This condition takes place, following (1.3) in Chapter 1, when \(\hat{A} > 1/[(1-\delta)(1-\omega)]\). It is possible to show that there is a broad range of values for which both the latter inequality and inequality (2.3) in this Chapter hold at the same time.
plunder and death. In short, the monarchical equilibrium is based on the fact that looters have some comparative advantage in the production of violence. In such a political regime, where the bandit has the monopoly of force and those that are ruled do not retain any armed force to resist the ruler, the bandit’s self-interest will be the only mechanism that guarantees his ‘good’ behavior. The bandit-turned-into-lord will set the extraction rate at the level that maximizes his income and any promise he may make to reduce $\epsilon$ below his optimal choice will never be credible.\(^\text{23}\)

Although I have treated monarchies as a single ideal type, monarchical regimes can be thought of as varying at least along two dimensions: according to the distribution of power within the ruling elite and according to the kinds of coalitions rulers strike with particular sections of the non-ruling group. As was widely accepted in modern political thought (Anderson 1974: 397-400), monarchies can be broadly organized according to two different systems: a purely despotic, such as the one prevailing in the Ottoman empire and in most Asian kingdoms, where servants are vertically integrated below the monarch; and a feudal or aristocratic one, in which the allies of a ruler, even they were subservient or vassals to the king, conserve some autonomy of power (in the form of weapons or assets) and often participate in some institutional structure (such as an assembly of warriors or a parliament of notables) that ‘constrains’ the monarch (or, at least, monitors the execution of the pact kings and vassals have struck with each other).

**The Republican Compact**

Consider now the nature of the republican settlement. In a republic, the payoff of producers will be $(1-\lambda^0_D)\hat{A}_i$. Following the notation in Chapter 1, $\theta_D$ denotes the efficiency of the

---

\(^{23}\) The net extraction rate will increase with the ‘administrative’ efficiency of the state. Second, it will also vary with the tax elasticity of output and with asset specificity (Boix 2003): the higher the mobility of assets, the less punitive the level of extraction by the lord.
technology of defense (with \( \theta_D > 0 \)): the higher the value of the parameter, the more efficient the time allocated to defense. (Notationally, I distinguish \( \theta_D \) from \( \theta_L \) or the effectiveness of a looting technology. However, their values can be identical in any given individual.) The overall production of defense is then \( \lambda \theta_D \), where \( \lambda \) is the time devoted to defense instead of production. As \( \theta_D \) becomes larger, producers are able to reduce the time devoted to military activities, still meet the threat of looters successfully, and generate a higher direct output \( \hat{A}_i \). In turn, \( j \)'s payoff under a successful self-governing or republican regime becomes 1. The potential looter \( j \) devotes all his time to production and none to looting because producers have been able to deter \( j \) from plundering them.

Given that structure of payoffs, producers (agents \( i \)) will prefer a republican structure over looting (provided they would rather have a republic over a monarchy) if total output (net of defense expenditures) is larger than looting \( j \) the first time and then going into permanent war in the following periods:

\[
\frac{(1 - \lambda^\theta_D) \hat{A}_i}{1 - \delta} > (1 - \lambda) \hat{A}_i + \lambda^\theta_L (1 - \omega) + \frac{\delta}{1 - \delta} (1 - \lambda) \hat{A}_i
\]

Rearranging the terms of this inequality gives us:

\[
\frac{\hat{A}_i}{1 - \delta} > \frac{\theta^\theta_L (1 - \omega)}{(\lambda - \lambda^\theta_D)}
\]

(2.4)

The incentives of any agent \( i \) to move to a republican solution increase when the alternative option of looting becomes less attractive – either because the looting technology \( \theta_L \) declines in efficiency (making the numerator smaller), the defensive technology \( \theta_D \) improves

---

\( \theta_D \) must be always a positive number. If it were equal to 0, then government institutions would only appear when individuals of type \( j \) chose a looting strategy. Otherwise, if individuals of type \( j \) did not loot, everyone would simply cooperate spontaneously with no permanent structure governing them. This does not seem to be realistic.
(making the denominator smaller and the whole right-hand side expression smaller), or the production technology \( \hat{A}_i \) increases.

In turn, agent \( j \) will accept the self-governing structure set up by \( i \) if just producing leaves him better off than trying the looting strategy:

\[
\frac{1}{1-\delta} > (1-\lambda) + \theta^{\sigma} \hat{A}_i + \frac{\delta}{1-\delta}(1-\lambda)
\]

Rearranging this inequality leads to:

\[
\frac{\hat{A}_i}{\theta^{\sigma}} > (1-\delta) \hat{A}_i
\]

(2.5)

In short, \( j \) is more likely to accept the production strategy when the defensive capabilities of \( i \) increases and when the value of \( i \)'s production declines.

Putting (2.4) and (2.5) together, a republican settlement is feasible whenever:

\[
\frac{\hat{A}_i}{(1-\delta)\theta^{\sigma}} > \frac{\theta^{\sigma} (1-\omega)}{(1-\delta)} \frac{\lambda}{(1-\lambda)}
\]

(2.6)

As in the case of a monarchical outcome, a republican settlement will be only feasible if, in addition to (2.6), peace without transfers is not an equilibrium.\(^{25}\)

We should not think of bandits as ‘external’ enemies only, that is, as individuals residing in a territory separate from the area of producers. The term of looter may well refer to individuals who are surrounded by ‘natural producers’ but who somehow have not profited from growth in the same way – either because the specific territory or social role they occupy is not as productive under the new technology or because they do not apply the same level of effort in the understanding or application of the new technology. Under that instance, individuals \( i \) have also an incentive to set up some mechanism to dissuade individuals of type \( j \) from stealing, free

\(^{25}\) Again, this condition takes place, following (1.3) in Chapter 1, when \( \hat{A}_i > 1/[(1-\delta)(1-\omega)] \). As in the monarchical settlement, it is possible to show that there is a broad range of values for which both the latter inequality and inequality (2.6) in this Chapter hold at the same time.
riding, etc. As a result, a state also includes institutions directed to policing social order within a given community.

As in the monarchical case, a self-governing or ‘republican’ regime is a political equilibrium. In addition the formal conditions (on the incentives of producers and looters) spelled out above, that equilibrium rests on two underlying factors worth examining. First, it requires some relative equality among producers. Otherwise, that is, if their economic differences were not somewhat bounded, they would internally split between those interested in producing and those interested in looting the producers. This does not mean there should be full equality among its citizens: again, the degree of equality should be enough to elicit some spontaneous cooperation among them.

Second, it entails the existence of a threat that directly forces those producers to establish a state. In the absence of the danger of looting (either within or outside the territory inhabited by a set of producers), those producers would not set it up. They would simply carry out their productive activities knowing that everyone else would. In other words, they would have the incentives to sustain a cooperative equilibrium (of the kind modeled in Chapter 1) without having to resort to strong or formal institutions. As in the monarchical solution, there is no third party guaranteeing the republican settlement: producers make an implicit ‘contract’ that will dissolve when either the external threat subsides or their internal heterogeneity increases too much.

Notice that throughout the discussion I have used the term ‘republic’ in part as a conceptual short-cut. Today, a republic is generally associated with a regime in which the government is elected by a country’s people. Instead, I define as a republic a state under the control of ‘producers’ – as opposed to the government by looters turned into monarchs (and aristocrats). In some instances, a republic (as defined here) is coterminous with a democracy – today’s industrial democracies. In other cases, it is limited to the government of an oligarchy of traders – medieval and modern Venice. As developed shortly, the latter case takes place when the
members of a republican state govern themselves as equals and then invest some resources to
control other individuals as subordinates: either through the use of direct force to control certain
external territories (the case of imperial republics) or through the construction of patron-client
networks in which the wealthy employ their income to buy out poor individuals.

**The Possibility of an Expansionary (Imperial) Republic**

So far I have examined the nature of a republic that is strictly defensive where producers employ some time $\lambda$ to defend their territory from looters. However, there is the possibility that the producers develop an offensive strategy to both defend their land and attack and subjugate the looters. The offensive strategy costs an extra fraction of $i$’s production above the price paid to simply defend their own territory. Therefore, since it implies a lower direct output for $i$, we can denote the offensive technology parameter as $\theta_O$ with $0 \leq \theta_O \leq \theta_D$.\(^{26}\)

Since the use of an offensive strategy implies that agents $i$ will apply sufficient force to make individuals $j$ spend their time producing even when they experience some looting from $i$, the natural producers will follow that strategy (to establish an imperial republic) only if the benefits of looting $j$ (formally, $(1-\omega)\lambda$) are larger that the portion of output lost in upgrading the military strategy from a defensive into an offensive one or $(\lambda^{0O}-\lambda^{0D})A$. That will in turn depend on two factors. First, the incentives to set up an expansionary (or imperial) republic will decline whenever the offensive costs $\theta_O$ increase with respect to the defensive costs $\theta_D$. Second, they will also fall the more productive the agents of type $i$ become with respect to the agents of type $j$. The latter proposition derives from the fact that, as the $i$’s become wealthier, the opportunity costs of any expansion become higher. Or, to put it in slightly different terms, the more marginal lands

---

\(^{26}\) It seems plausible to think that even when the main endeavor of these agents is production, they still need to pay always some permanent positive costs of defense $\theta_D$ to be ready to move to the fully militarized strategy.
and persons (of type $j$) are, the less prone their (richer) neighbors $i$ will be to attack and subject them.\textsuperscript{27}

**A Note on Growth and Redistribution**

Besides affecting the distribution of political power across individuals, the type of regime in place has substantial implication for both economic growth and its distribution. As discussed in Chapter 4, monarchies result in a considerable transfer of resources from producers to rulers – to the point of reshuffling the initial distribution of output and income as determined by the factor endowments and know-how of individuals. In contrast, in a republic producers tend to maintain the underlying distribution of income as determined by the marginal productivity of their assets. Still, redistribution may take place in republican regimes, especially if the number and intensity of external threats increase. In that latter instance, richer producers may be willing to fund some equalization mechanisms and add new and poorer members to bolster the strength of a republic vis-à-vis a monarchy.\textsuperscript{28}

As examined in Chapters 5 and 6, monarchies will hamper any growth (resulting from a process of learning-by-doing) that could strengthen the position of producers and jeopardize their tenure in power. In a republican system, in contrast, its citizens, who, qua producers, have benefited from a process of knowledge accumulation, should generally favor its continuation. However, growth rates tend to be low in republics for two reasons. First, by definition technological innovation does not happen uniformly across individuals. The rise of economic

\textsuperscript{27} For an analysis of the conditions that lead to the choice of the optimal size of territory to control (and therefore of the drive to expand), see formally Findlay (1996) and Alesina and Spolaore (2000) and through the use of agent-based models, Boix et al (2011).

\textsuperscript{28} In a monarchy a higher level of threat has also its own set of implication for redistribution. As threats increase, the ruling class weakens and the extraction rate declines. Hence, the level of enforced redistribution on producers also falls.
heterogeneity within a human community, even if it is temporary, may lead, in the absence of compensatory mechanisms to equalize conditions, to an endogenous political breakdown. Hence, it is not uncommon to see republican systems that, for the sake of internal stability, impose limits to the process of endogenous growth, often to the point of stifling change completely. Second, even when those regulatory mechanisms are not in place, economic growth is minimal because republics lack economies of scale (due to a set of military and political constraints to be examined in the fourth section of this Chapter).

**MONARCHIES VERSUS REPUBLICS**

Having determined the conditions under which order will prevail, let us now consider the cases under which one of the two types of political regimes will prevail. A republican solution will be preferred by the natural producer $i$ whenever the discounted value of production under a republic (that is, excluding the time devoted to defense) is higher than discounted value of the output after paying some transfer $\varepsilon$ to the monarch:

$$\frac{(1 - \lambda^o_i)A_i}{1 - \delta} > \frac{(1 - \varepsilon)A_i}{1 - \delta}$$

(As discussed before, that equality requires too that the agents prefer some political order, either monarchical or republican, to a situation of anarchy.) Now, simplifying the expression, a republican will take place whenever:

$$\varepsilon > \lambda^{0i}_D$$

(2.7)

Notice that, if the producer sets up a republic, the potential looter $j$ will abide by the solution since, by definition, a republic is only possible if producers have the actual military capacity to deter looters from pillaging them (and, again, this deterrence strategy still is preferable to any other solution). If the producers are, instead, better off under a monarchy, agents of type $j$ will not accept the monarchical settlement if their payoff under a republic is
larger than the monarchical transfer \((1-e)A\) – with the looting alternative precluded by our assumption that inequality (2.7) holds. In that case, the producers will have to increase \(e\) to the point that monarchy becomes attractive to \(j\) (provided that, even after that increase in the extraction rate, a republic would still leave agents \(i\) worse off than the monarchical solution).

Two parameters determine the type of political regime that will prevail: the extraction rate \(e\) and the military technology of defense \(\theta_D\). On the one hand, the incentives of producers to establish a republic will rise with the extractive threat of the looter – i.e. the size of \(e\). On the other hand, the likelihood of a republican system will increase proportionally with the military capability of producers – formally, as \(\theta_D\) gets closer to 0, since the parameter captures the ratio between the military technology of looters and the military technology of producers.

The parameter \(\theta_D\) is driven by two factors. First, it simply depends depend on the fighting skills and military proficiency of producers and looters. Say that, in line with the sources of economic growth described in the first section of this Chapter, individuals generate new weapons or perfect some organizational techniques of war also through a process of learning-by-doing. To the extent that those innovations dovetail more closely with the resources and skills of each group, the military ratio \(\theta\) shifts accordingly. Whenever producers are better at waging war than plunderers, the former is able to deter the latter and to associate in human communities where decisions are made in a consensual manner. However, as soon as the producers’ technology of war declines relative to the looters’, the cost of a republican government is too steep and monarchical regimes become the prevalent form of government (Findlay 1996:44).

By the middle of the second millennium B.C., the populations living in the Middle East and in Central Asia had domesticated the horse, invented the flexible two-wheel chariots and developed the composite bow. From an economic point of view, the combination of those three technologies benefited the pastoralist peoples of the steppes more than the much richer agricultural populations of the Near East. They provided the nomad pastoralist with “a means of
herding his flocks at a pace faster than his feet could carry him, and also to put him on near, if not equal, terms of mobility with the predators, wolf and perhaps bear and the large cats too, that harried their flanks” (Keegan 2004: 165). But they could be also (and indeed were) seamlessly transferred to the military arena, where they gave their users an extraordinary fighting advantage over their neighbors. A chariot crew, with one driver and one shooter could kill about six men per minute “circling at a distance of 100 or 200 yards from the herds of unarmored foot soldiers” to the point that “ten minutes’ work by then chariots would cause 500 casualties or more …” (Keegan 2004: 165-166). Indeed, in the second half of the second millennium, and the brief span of about a few hundred years, the steppe people conducted one of the most significant expansions in history: Babylon fell under the attack of the Hurrites; the Aryan conquered Northern India, subjugating the existing Dravidian populations; the Acheans settled in Greece; and the Hyksos invaded Egypt (Anthony 2010; Andreski 1968). Yet technological advances could play as well in favor of producers. Creating and sustaining a powerful navy required an underlying set of skills (as well as capital) that could only be nurtured in trading and commercial cities. A comparative advantage in manufacturing and engineering could be translated into a stronger army as well. The expansion of crowbow manufacture in cities like Barcelona allowed its soldiers to destroy an army of French knights in Sicily in 1282 and to occupy half of the Greek Peninsula in the fourteenth century (McNeill 1982: 67-68). In the sixteenth century Italian engineers developed the trace italienne – fortification walls made of loosely compacted earth that could absorb cannot shots successfully. That invention, that required the use of enormous amounts of labor and monetary resources, could only be funded by the rich cities of the central belt of Europe and arguably introduced a barrier in the process of political unification in place in Europe since the introduction of gunpowder (McNeill 1982: 89-91).

The military ratio embodied by $\theta$ is a function too of the opportunity costs of fighting relative to the production of goods and services. That decision, which comes from An upward
shift in the marginal rate of transformation between the time spent in production and the time
directed to war reduces the producers’ incentives to wage war. If that shift is strong enough,
producers may prefer paying some third party to defend them while devoting themselves fully to
production, even if they retain some absolute advantage in the waging of war. The opportunity
costs of fighting are partly related to the technologies of war in place. But technological progress
in general may be one of the key factors leading to higher opportunity costs of fighting, to the
extent that it leads to more economic and military specialization. When both the production and
management of violence become harder and more complex, only those individuals that devote
their entire time to that task can become fully competent in any military endeavor. As noted by
Huntington, “the intellectual content of the military profession requires the modern officer to
devote about one third of his professional life to formal schooling, probably a higher ratio of
educational life to practice time than in any other profession” (1957: 13). Now, unless the
producers find a way to control the army they set or hire to defend them—a question I examine in
the next section, higher opportunity costs should result in the emergence of monarchical or
dictatorial states.

**WARFARE AND TERRITORIAL SIZE OF REPUBLICAN REGIMES**

Looking at the historical evolution of the state since its appearance a few thousand years
ago, two facts stand out. First, monarchies have been the overwhelmingly predominant system of
government across the world – at least until the last two hundred years. Second, republics have
been not only scarce but short lived and small in size as well. There is no doubt that the military
skills of looters and the opportunities costs of fighting have played a big role in explaining the
historical distribution of political regimes. There is, however, an additional factor, inherent to the
political structure of republics, that explains why the number and, more particularly, the
territorial extension and survival rate of republican regimes have been small.
To defend itself, a republic may follow two alternative strategies. On the one hand, it can raise a citizens’ militia. On the other hand, it may contract out the tasks of military defense (and offense) to a set of soldiers separate from the body politic. This latter solution reduces the opportunity costs of those producers that have to serve in a militia – an important consideration when mastering the technology of war requires a high investment in terms of time. However, it may come at a high political cost. A hired army cannot credibly commit to preserve the terms of the contract made with the republican regime. Once in control of the production and management of violence, it will be tempted to subject the producers and establish a monarchy. By the end of the fourteenth century, several Italian cities, awash with money, started contracting the service of mercenary captains and troops to replace their own town militias. Such a system, which was itself fraught with high transaction costs, often ended in violent coups around Italy – with the Sforza takeover of Milan being its most visible case (McNeill 1982: 74-76). As a result, already in the late fourteenth and early fifteenth century, several Italian humanists, such as Petrarch and Leonardo Bruni, would call for the restoration of an armed and independent citizenry (Skinner 1978: 75-77). A century later, Machiavelli would equally oppose the use of “mercenary captains” since one “cannot trust them because they always aspire to their own greatness, either by oppressing you, who are their patron, or by oppressing others contrary to your intention” (*The Prince*, chapter 12, p. 49).  

The city of Venice succeeded in employing mercenary troops while subjecting them to the civil power through two mechanisms: dividing the contracts among several, very small, units, which, out of jealousy, would not coordinate against the city’s authorities; and “bestowing civic honors and gifts upon loyal and successful condottieri and arranging suitable marriages for them with members of the Venetian aristocracy” (McNeill 1981: 76). But Venice was an exception  

---

29 For a similar position in his *Discourses*, see Book II, chapters 20 and 30.
among Italian cities. Until the creation of large standing armies in the late eighteenth century, its geographical location and its extraordinary advantage in commerce and naval power made it unassailable, probably reducing mercenary troops to performing a rather tangential in its defensive system. It is not surprising then that, in light of the Italian experience, both American and French revolutionaries, keenly aware that the survival of a republican form of government depended on the control of the army, tried to democratize, with rather unsuccessful results, the selection of officers (Huntington 1957:33-34). And that the American Continental Army was disbanded in 1784 on the grounds, expressed by Elbridge Gerry, that “standing armies in times of peace are inconsistent with the principles of republican governments, dangerous to the liberties of a free people, and generally converted into destructive engines for establishing despotism.”

To overcome the threats posed by condottieri, republican regimes may resort to the creation of a professional army, that is, a body of officers (and sometimes foot soldiers) who are permanently specialized in war-making and yet remain under civilian control. Historically, this strategy was only implemented very late in time and, as matter of fact, not by republican regimes. Prussia led the way by introducing a set of reforms to professionalize its officer corps between 1808 and 1812 in the aftermath of its defeat in Jena at the hands of the French army. Still, the German army was under the tight control of the Kaiser until World War One. In turn, France followed suit during the Restoration period, a few decades before the instauration of the Third Republic in the 1870s. Even after the stabilization of a republican system, however, France had to contend with the strong involvement of its officers from Boulanger through the Dreyfus affair up to Pétain’s regime in 1940. Among countries with more liberal constitutions, professionalization actually took longer to happen: the United Kingdom only abolished the purchase of officer positions in 1871 and all the institutions of American military professionalism

30 Quoted in Huntington (1957: 144).
developed after the Civil War (Huntington 1957). In any case, a professional army does not guarantee the internal survival of republics. Formally subjecting the officer corps to the state and establishing open recruitment mechanisms based on education and competence has not solved the republican dilemma of having an army that is both competent about subject to the civil authority. A large proportion of twentieth-century military coups have been staged by highly professionalized armies (Geddes 2011).  

The tension between republican governments and the organization of the army becomes particularly acute at the time of war and especially whenever the republics engage, consciously or by default, in some form of territorial expansion. War distorts the distribution of resources within the polity to the point of disrupting the political equilibrium that sustains republican institutions. A prolonged war effort generally requires the creation of a specialized or stable commanding structure. As Machiavelli stressed while discussing the causes behind the fall of the Roman republic in his Discourses, “the further the Romans went afield with their armies, the more necessary [the prolongation of military] commands seems to be, and the more use they made of it” and, in fact, “had the Romans not prolonged offices and military commands, they would not have attained such great power in so short a time” (Book III, chapter 24, p. 474). But, once a small number of men acquire experience as military commanders, they gain, if they are successful in the battlefield, a strong reputation among their troops and the population in general, which they may exploit to build up their own sets of followers or clienteles. That new form of

---

31 In those countries where a professional army does not lead to a military takeover, this result may be the result of well-crafted institutional procedures (in the sense that the army is just another agency within a specialized bureaucracy that is vertically integrated in the state) or to the presence of certain non-institutional conditions that dissuade the military from staging a coup. Among the latter, consider the possibility that land, which use to be the main asset that military commanders could distribute to their followers, has become much less important and valuable in the contemporary world.

32 For a relatively similar point, see Montesquieu, *L’esprit des lois*, book X, chapter 6.
political inequality, which often leads to civil war, brings about the fall of the republic and the construction of some dictatorial or monarchical structure – the case of Rome and arguably of Athens after the Peloponnesian war and, more recently, of Napoleonic France.

Faced with a trade-off between war-making and the maintenance of republican institutions, republican regimes have to minimize the extent to which they fight. They will only invest in defensive capabilities. They will shun away from territorial expansion. And they will engage in war mostly with polities of similar size. That trade-off or dilemma was then directly responsible for the territorial size and duration of republics, particularly before the rise of the professional army in the nineteenth century. First, other things being equal, the threat of a condottiero (or even a “professional”) army made republican regimes much smaller than monarchical states from a territorial point of view (provided there were the appropriate technologies of war that allowed monarchs to expand if they chose to). As indeed noticed by Montesquieu, “it is in the nature of a republic to have a small territory” (Spirit of Laws 1748, book VIII, chapter 16), for the territory of a monarchy to be of “un grandeur mediocre” (ibid. chapter 17) and for a despot to hold a large empire (ibid. chapter 19). Republican states only became large under two circumstances: as a result of the introduction of some technology of war gave them some offensive advantage – probably the case of the Italian armed infantry that defeated noble horsemen at Legnano in the late twelve century; and whenever war-making was tightly related to the kinds of productive endeavors of their citizens – the trading cities of Athens and Venice based their expansion on the systematic use of naval power.

Second, it made republics rather short-lived, spanning, at most, over a period of a few decades or, exceptionally, one or two hundred years – mainly because they were small and therefore easy to defeat. Until the twentieth century republican forms of government were extremely uncommon. They only prospered in classical Greece, among Etruscan and Latin cities before Rome’s expansion, in medieval and modern Europe and, arguably, in a few Swahili cities,
the Yoruba’s city of Ibadan and the Malayan town of Banda-Neira (Hansen 2000: 619). Since a single city could not survive against bigger principalities, it is not surprising that republican governments appeared clustered in a specific geographical area, such as Greece in the classical period or the North-South axis from Tuscany to the Netherlands in medieval and modern Europe. Those cities that survived the longest relied on the formation of a confederal structure of self-defense, or a “république federative” to employ Montesquieu’s expression, such as the Swiss and Dutch leagues (Spirit of Laws, Book IX, Chapter 1) or the Tuscans, Achaeans or Aetolians in pre-Roman times (Machiavelli, Discourses, Book II, Chapter 3). The trade-off between the weakness of small republics and the authoritarian tendencies of big countries constituted one of the main preoccupations of the Federalist debates. Indeed, the constitution of the United States may be seen as the first written charter that institutionalized a federal structure of self-governing states with the explicit goal of solving that dilemma.

A FUNCTIONAL THEORY OF STATE FORMATION

According to a rather extended theory of state formation, the state emerged as a wealth-optimizing tool to compel all individuals to coordinate on a cooperative equilibrium, mostly relying on a system of sanctions guaranteed by a third party (the state itself). Although that approach shares the same point of departure with the theory of the book, namely, the breakdown of spontaneous cooperation, it differs from the latter in how it understands the process through which the state came into being. In the most developed accounts of an optimizing theory of state formation, the players agree to establish a set of sanctioning mechanisms to sustain cooperation

33 There seemed to have been robust assemblies in otherwise monarchically ruled city-states in Sumer (Glassner 2000), some Phoenician towns such as Carthage (Hansen 2000), early medieval Ireland (Holm 2000), the Niger Delta (Princewill 2000) and the Chinese cities of the Spring-Autumn period (771-481 B.C.) (Lewis 2000).
in order to overcome an outcome of generalized defection that is suboptimal to all of them. In
other accounts, the state simply appears, through some process that generally remains
undescribed, to fulfill the function of coordinating everyone around certain tasks and of
providing some key public goods.

In the historical literature, Wittfogel’s theory of hydraulic societies comes the closest to
that theoretical approach. According to Wittfogel (1957), the state and its bureaucratic apparatus
were born to develop and maintain vast irrigation systems in the first agricultural societies that
flourished in Egypt, the Middle East, India and China. In the formal literature, in turn, the process
of state formation as an optimization problem has been generally modeled as a contract among
individuals who give up their capacity to fight on some party that then agrees to sanction them if
they deviate from cooperating with each other.34 Since the 1960s and 1970s, however, a host of
archaeological and historical studies have strongly questioned Wittfogel’s hypothesis: the process
of agricultural intensification and the construction of complex irrigation systems developed
without the presence of state institutions in areas such as the Titicaca Basin, Neolithic Turkey
and East Africa (Thurston and Fischer 2007). More to the point, framing the formation of the
state as a strict optimization problem bypasses the central theoretical problem at hand. A state is
necessary to coordinate everyone around a cooperative solution when all (or part of) the players
in the game have an incentive to defect. But, if they do, how do they overcome their temptation
to free ride and agree to be restrained? The fact that some particular outcome may be socially
optimal does not guarantee that it will be optimal for each individual as well. And, logically, if it
not individually optimal, there is no way everyone will surrender their capacity to coerce and
exploit others. In short, an optimizing theory of state formation is not internally consistent.

In a situation of generalized conflict, it will be only feasible to establish order if one can create a mechanism to change the payoffs (and therefore the preferences) of the players involved. That would require the transfer of some of the collective gains from having order to those that would otherwise defect. Or, alternatively, it would imply the use of sanctions to make free-riding worse than cooperating with others. The first solution corresponds to the monarchical solution in which the looters have an incentive to restrain themselves once they realize they can keep part of those collective gains for themselves. Naturally, this entails giving them the authority to coordinate everyone and to determine the allocation of total social gains. Now, since they come to control the state, nothing limits them except for the need to keep the non-looters producing. The second solution matches the republican solution. In this case, the producers pay the cost of sanctioning those that free ride. The only requirement here is that the benefits from order exceed the costs of punishing others.

In both cases, the final outcome is preferable to a situation of generalized looting. But, in contrast to a functionalist theory of the state, the state does not appear as a result of a collectivity solving an optimization problem. It emerges as one set of players finds a solution based on some application of force (or the threat of force) that is acceptable to everyone else. Similarly, sanctions and punishments are a feature of the state. But that does not mean that the state appears in order to establish them. Rather, they are put in place because they are the tool through which a particular set of individuals can overcome the collective action problem that follows from a biased technological shock or a sudden drop in the discount rate within a given human group.

**Conclusions**