Proof-reading by George Seel

COMPETITIO Books 12.

This book has benefited from a financial support of the Hungarian Scientific Research Fund (grant no. 67713).

ISSN 1785-5950
ISBN 978-963-473-550-2

Debreceni Egyetem Közgazdaság- és Gazdaságtudományi Kar
Felelős Kiadó: Dr. Kormos János dékán
Tördelés, nyomdai kivitelezés: Center-Print Nyomda Kft. Debrecen
Felelős vezető: Szabó Sándor

Institutions and the Industrial Revolution

EDITED BY

JUDIT KAPÁS
PÁL CZEGLÉDI

Debreceni Egyetem – Közgazdaság- és Gazdaságtudományi Kar
2012

CONTENTS

PREFACE

This volume is the product of our research entitled *Technological Advance and Institutional Change: A Coevolution and Developmental Paths* carried out between 2007 and 2012 and financed by the Hungarian Scientific Research Fund (grant no. 67713).

The aim of the research project was to explore the coevolution of technology and institutional structure. This comprised two basic elements. On the one hand, the question was how technological advance affects both the broad institutional structure and the microinstitutions of the firm. On the other hand, the research question also included the analysis of how this effect works in the opposite direction, that is, how the evolution of spontaneous institutions and the change in designed institutions affected technological change.

Since industrial revolutions represent major technological transformations, we devoted special attention to the British Industrial Revolution. We also see the Industrial Revolution as a special historical event that provides insights which help us derive a general hypothesis concerning the major question of our research project. Thus, although all four of the following papers are concerned with this event in some sense, some of them put more emphasis on this "inspirational" view of the Industrial Revolution, while others see it as an event to be explained.

The papers of this volume are not only connected through a common theme but also through their institutional approach, which takes it as given that institutions matter and tries to identify those institutions that matter and explain the ways they matter and the ways these institutions develop.

Despite the common broad theme and common approach the papers are not strictly interconnected, since they provide arguments in their own right. The first paper (*Mechanisms of Success: How Do Macroeconomic Models Explain the Industrial Revolution?*) is a review of the main mechanisms macroeconomic models provide to explain the industrial revolution in particular and the start of economic growth in general. Although these models are very diverse, from an institutional point view *Pál Czeglédi* identifies two main groups of models: those that consider institutions to be of a primary, and those that consider them to be of secondary, order of importance.

In the second paper (*Which Institutions Caused the British Industrial Revolution?*) *Judit Kapás* provides a review too, but with the aim of reviewing the institutional theories that aim to explain the British Industrial Revolution. She shows that the institutional theory of social orders can helpfully be applied as a meta-theory, or umbrella theory, of all the theories reviewed. This interpretation shows that there are no simple answers to the question raised in the title of the paper. It is not (only) because there are many institutions that are relevant but because what is to be explained is a process: the transformation of a limited access order to an open access one by the development of the institutional conditions of impersonality.

The third paper (*Larger Prey, More Predators: Culture as a Constraint on Expropriation*) focuses on those institutions that are usually grouped under the heading of "culture", and are often identified as important conditions of the industrial revolution and economic growth. In this paper *Pál Czeglédi* argues that the cultural factors that are relevant can be seen as elements of an "ideology of freedom" which can be interpreted as an attitude towards rent-seeking on the constitutional and sub-constitutional level. Incorporating this idea into a simple model of technology diffusion, it becomes possible to differentiate between two functions of this ideology, and between exogenous and endogenous changes in property rights security. The most important prediction is that these two kinds of change in property rights security

will have a different effect on the spread of technology. This prediction is tested by using data on culture and the spread of different technologies.

The term “Industrial Revolution” evokes images of the factories which made it possible to make use of the technology discovered at that time. Indeed, the evolution of factory system was an important part of the story of the Industrial Revolution, as is explained in the fourth paper (*The Factory: An Historical Theory of the Firm View*) by *Judit Kapás*. In it she combines two approaches to the factory system. One is the historical view which emphasizes technological changes as the causes of its development, and the other is the theory of the firm which, being concerned with understanding the nature of the factory system, usually does not pay much attention to an explanation of the origin of this particular institution. In her historical theory of the firm view the emergence of the factory system is an evolutionary theory which explains how the development in the division of labor caused by the growing extent of markets led to the emergence of firm-like monitoring and authority that lie at the heart of the nature of the factory.

We, of course, do not hope to give a full picture of the technology – institutions nexus with these four papers. We do hope, however, that we can contribute to the understanding of this complicated system of interrelationships by taking stock of, and organizing, the main arguments, and by shedding light on some mechanisms which have not been addressed in the literature so far.

Pál Czeglédi

Mechanisms of Success: How Do macroeconomic Models Explain the Industrial Revolution?

"Such exercises make it clear that the years since 1960 are part of a period of transition, but from what to what?" (Lucas 2002:117)

1. Introduction: economic growth as the spread of the industrial revolution

The question as to why there is economic growth is necessarily linked to the question as to why there was an industrial revolution. To ask the question "why some nations are rich while others are poor" (Olson 1996) is to ask the question why some nations have already joined the process of industrial revolution and why others have not. This view of world-wide economic growth is very well illustrated by Lucas (2000). In it the across-the-world income distribution is derived from a model in which the probability of a country's joining the club of economic growth is given.[1]

In the economist's mind the industrial revolution and economic growth are naturally interconnected; and the former marks the beginning of a new regime which makes it possible for the whole world to become developed (Easterlin 1981). Although some recent explorations in economic history (Goldstone 2002, Zanden 2009) show that the industrial revolution may not be such a clear-cut borderline between an era of economic growth and an era of its absence, even these agnostic economic historians admit that the industrial revolution was unique in one sense: it was the time when technology began to spread throughout Europe and economic growth was no longer confined to Britain and the Low Countries. The expression that the industrial revolution and modern economic growth was caused by a "wave of gadgets" (McCloskey 2008:248, Zanden 2009:2) still seems to be a good description of what has happened.

In realizing the importance of the industrial revolution in human history, growth theorists do not only face the challenge of explaining why per capita income is growing steadily, but to explain why per capita income may *begin* to grow after a long period of being stable and low. That is, the most important challenge is to understand the transition from a no-growth regime to a growth regime.

As a consequence of this somewhat ahistorical view of the industrial revolution, the theories of the industrial revolution must help explain contemporary differences in the wealth of nations. The task of this paper is to review the main theories of the emergence of a growth regime by focusing on the fundamental mechanisms that constitute these theories. Accordingly, I will group the theories into two branches. The models of the first are organised around the demographic transition and the ways parental choice matters in human capital accumulation. Into the second group I will put those theories that emphasize the barriers that can stand in the way of the spread of new technologies. The models in the second group also include those models that are centered around the idea that genetically or culturally transmitted traits can serve as barriers to technological change.

2. The Malthusian trap

One group of models that includes a mechanism to understand the transition to a new growth regime places population dynamics at the centre of the explanation. Trying to explain demographic change and economic growth simultaneously these models usually describe two

[1] Galor (2011:179-231) also draws conclusions on comparative development patterns from a model originally aimed at explaining the industrial revolution.

regimes in a "unified" analytical framework. The first regime is the Malthusian one, during which income per capita was stagnating, while the other is the age of modern economic growth. In addition they are able to model the transition from the stagnant Malthusian era to the abundant post-Malthusian one.

Malthusian logic can easily be given a formal description, which I will do here for the sake of illustration. The logic is explained in detail by, for example, Clark (2007:19-39) diagrammatically[2]. The following simple model is a formulation of his description. Assume that the production side of the economy can be described with a Cobb-Douglass production function with labor as a factor of production (N) and with land (T) being the factor that cannot be accumulated:

$$Y = AT^{1-\alpha}N^{\alpha},\ 0<\alpha<1, \tag{1}$$

where A is total factor productivity.

Assume that productivity is growing at a constant rate:

$$\gamma_A \equiv \frac{\dot{A}}{A} > 0. \tag{2}$$

The crucial assumption is that the number of births and deaths, and thus population growth are connected with per capita income: when per capita income is growing, the number of births goes up while the number of deaths goes down. To formalize this assumption, write:

$$\frac{\dot{N}}{N} = b(y) - d(y), \tag{3}$$

where b(y) is the birth rate and d(y) is the death rate, and we suppose that b'(y)>0 and d'(y)<0.

It is easy to show that this system has a steady state where population is growing steadily while income per capita is constant. Solving for the growth rate of income per capita (y) from the production function we get:

$$\frac{\dot{y}}{y} = \gamma_A - (1-\alpha)\frac{\dot{N}}{N}, \tag{4}$$

that is

$$\frac{\dot{y}}{y} = \gamma_A - (1-\alpha)(b(y) - d(y)) \tag{5}$$

which is a differential equation for y that has a steady state at y=y* with the latter being implicitly defined as

$$(1-\alpha)(b(y^*) - d(y^*)) - \gamma_A = 0. \tag{6}$$

It is clear that when per capita income is above this steady state level, the growth rate is negative, and vice versa. Thus in the long run, per capita income is constant and population is growing at a constant rate:

$$\frac{\dot{N}}{N} = \gamma_A/(1-\alpha). \tag{7}$$

According to this strange logic, even if there is a constant technological change, the growth rate of income per capita will remain zero at the steady state, and technological improvement will only raise population growth and the steady state level of income.[3] Within the framework of this model there is no way out of the Malthusian stagnation. Although these

[2] Every model discussed in section 3 includes a Malthusian economy as a special case. The model above is as simple as possible to highlight the main mechanisms.

[3] From the equation for the steady state income it follows that: $\frac{dy^*}{d\gamma_A} = \frac{1}{b'(y^*) - d'(y^*)} > 0$.

predictions may be in line with the historical facts before 1800 (Clark 2007, Galor and Weil 1999), they certainly do not fit the experience after that.

3. Non-Malthusian mechanisms

It is clear that that the main assumptions that make it impossible to leave stagnation behind in the Malthusian model are that (1) higher income leads to a higher birth rate; that (2) labor has a diminishing return, and that (3) technological change cannot accelerate. All these assumptions are doubted by the literature explaining the industrial revolution, although not all three appear to be equally invalid in each model. Observed through the lenses of these kinds of explanations the industrial revolution becomes the other side of the coin of the demographic transition. The explanation for economic growth must be the explanation for the change in population dynamics as shown by Lee's description (2003:167):
Before the start of the demographic transition life was short, births were many, growth was slow and the population was young. During the transition, first mortality then[4] *fertility declined, causing population growth rates first to accelerate and then to slow again, moving toward low fertility, long life and an old population.*

3.1. Fertility choice and the allocation of time

3.1.1. The quality-quantity trade-off

One reason why the prediction of the simple Malthusian model is obviously not true for the rich countries is its failure to model the parental choice to raise children. Consequently, one group of models that can explain the escape from the Malthusian traps includes those that made the choice of parents concerning raising children endogenous by associating the decisions on human capital investment with the decisions about the quantity and quality of children to be raised. These models are rooted in the economic theory of fertility proposed by Becker (1981) who models this choice by the standard tools of microeconomics.

One early economic model of the child raising decision is that of Becker and Lewis (1973), in which this decision is understood as a choice between the quantity and the quality of children, and other goods (consumption). The trade-off between quality and quantity is revealed by the fact that there are effectively two goods between which the players choose: consumption and the quality unit of children, that is, the product of their qualities and quantities. Becker and Lewis (1973) describe the situation by describing the budget constraint of parents in the following form:

$$\pi_y y + \pi_n n + \pi nq + \pi_q q = I, \tag{8}$$

where y is consumption, n is the quantity of children, q is their quality, and π's are the prices of these goods. This implies that the alternative costs of raising one more child are dependent on the quality chosen and vice versa: a better quality child is more difficult to raise. As a result, observed income elasticity of demand for children can easily be negative.

To understand the difference between observed and "real" income elasticities it is useful to follow Becker and Lewis (1973:S281) and rewrite the equation above as

$$p_y y + p_n n + p_q q = I + \pi nq \equiv R \text{ , where} \tag{9}$$

$$p_n = \pi_n + q\pi,\ p_q = \pi_q + n\pi. \tag{10}$$

[4] Interestingly enough, Guinnane (2011) in a similar review paper on the same topic is not so confident concerning the timing of the two sorts of decline. Using the "not esoteric" examples of France and the USA he argues that the timing is not easy to generalize. As he writes (ibid:593) "[m]ost scholars agree that France's fertility declined in the early nineteenth century at the latest", furthermore (ibid:599) "[f]ertility in the United States declined for decades before any noticeable decline in mortality"

Real income, R, is higher than observed income, I. That is why an increase in observed income will raise the real income by a smaller proportion, and observed income elasticity is lower than the real one. Calculating the real income elasticity of children requires us to see the real costs of raising a child as a constant, but this does not remain a constant when observed income grows. When the observed income rises, then, because of the positive real income elasticity, quality demanded will rise as well, increasing the cost of the quantity of children (p_n) and decreasing the quantity demanded. Simply put, when an income rise is observed, the price of the quantity of children is not constant. Assuming that the income elasticity of quality is higher than that of its quantity, observed income elasticity can be negative, that is, parents with higher incomes will choose to raise fewer children but with higher quality

When it comes to models of economic growth, this difference between income elasticities is important, because a rise in wages can reduce the demand for children (n). In this simple model a wage rise is equivalent to a rise in every price except for the price of consumption, since a wage rise is a rise in the alternative cost of time. This is equivalent to a decline in the price of consumption. In this case parents will substitute the quantity of children for quality, if the observed income elasticity of quantity is smaller. Although the substitution effect will decrease the demand for both goods, income effects will differ. If the income elasticity of quality is higher, then the real price of quantity ($p_n=\pi_n+q\pi$) will grow faster than that of quality ($p_q=\pi_q+n\pi$), thus quality will be increased at the expense of quantity. One cannot exclude the possibility that as an effect of the increase in wages not only demand for consumption, but demand for the quality of children will increase, while demand for the quantity of children decreases.

Becker and Lewis formalize this argument in a very generalized way, but it is worth examining a special case. It is quite common in growth theory to assume Cobb-Douglas utility (and production) functions. Let us suppose that the utility function of parents is given in the following form:

$$U(n, q, y) = \alpha\ln(n) + \beta\ln(q) + (1-\alpha-\beta)\ln(y), \tag{11}$$

where the notations are the same as above, and $0 < \alpha < \beta < 1$, $\alpha + \beta < 1$. Using the budget constraint with real prices we get the well known demand functions:

$$n(p_y, p_n, p_q, R) = \alpha\frac{R}{p_n}, \tag{12}$$

$$q(p_y, p_n, p_q, R) = \beta\frac{R}{p_q}, \tag{13}$$

$$y(p_y, p_n, p_q, R) = (1-\alpha-\beta)\frac{R}{p_y}. \tag{14}$$

For reasons explained before, observed demand functions are different:

$$n(\pi_y, \pi_n, \pi, \pi_q, I) = \gamma\frac{I}{\pi_n}, \tag{15}$$

$$q(\pi_y, \pi_n, \pi, \pi_q, I) = \delta\frac{\pi_n}{\pi}, \tag{16}$$

$$y(\pi_y, \pi_n, \pi, \pi_q, I) = \left[1-\gamma\left(1-\frac{\delta}{\pi}\right)\right]\frac{I}{\pi_n}, \tag{17}$$

where $\gamma = \frac{\alpha-\beta}{1-\beta}$ and $\delta = \frac{\beta}{\alpha-\beta}$, and to simplify further $\pi_q=0$ is supposed, which means that the alternative cost of quality as a function of quantity (p_q) grows faster than the reverse (p_n). As a result of the simplicity of this example, the income elasticity cannot be negative because that would make the quantity demanded negative ($\alpha < \beta$). Because of the Cobb-Douglas form

of the utility function, income elasticities are unity. It holds, however, that observed income elasticities are smaller than real ones, since the observed income elasticity of quality is zero. The effect of a wage rise is somewhat special as well. If π and π_n increase in the same proportion, demand for the quantity of children will decrease, while the demand for quality will not change.

3.1.2. Parental altruism

The static model cannot serve as a direct basis of a dynamic model of economic growth. This core model of parental choice has been developed further in two directions. The quality of children can either be associated with the human capital they accumulate or with their prospective utility level.

The first possibility[5] is followed by Becker and Barro (1988) and Barro and Becker (1989). In their models the quality of children is interpreted as their utility level, and parents' utility functions can be described as dynastic: parents' utility is the function of the utilities of their children. A classic formulation of this type of a model is that of Becker and Barro (1988). They start with the following general utility function:

$$U_0 = v(c_0, n_0) + \sum_{i=1}^{n_0} \Psi_i\left(U_{1,i}, n_0\right), \tag{18}$$

where U_0 is the utility of the parents (generation 0), while $U_{1,i}$ is the utility of member i of the next generation (that is the utility of the ith child), c_0 is the consumption of parents, and n_0 is the number of children. Ψ_i is thus a measure of the altruism toward child i. Making the assumption that this is the same toward each child, and specifying it further in the form:

$$\sum_{i=1}^{n_0} \Psi_i\left(U_{1,i}, n_0\right) = n_0 \Psi_i\left(U_1, n_0\right) = a\left(n_0\right) n_0 U_1, \tag{19}$$

the utility function will be simpler. Because of this simplification, the utility function of the parents can easily be written in a dynastic form which includes the utility of all future generations:

$$U_0 = \sum_{i=0}^{\infty} A_i N_i v\left(c_i, n_i\right), \text{ where} \tag{20}$$

$A_0 = 1, A_i = \prod_{j=1}^{i-1} a\left(n_j\right)$, $i > 1$, and $N_0 = 1, N_i = \prod_{j=1}^{i-1} n_j$, the number of all descendents.

Specifying it further they assume that[6]:

$$a\left(n_i\right) = \alpha\left(n_i\right)^{1-\varepsilon}, \ 0 < \alpha < 1, \ 0 < \varepsilon < 1. \tag{21}$$

To have a budget constraint they further assume that parents leave bequests to their children (k_i) possibly interpreted as human capital, and labor supply is perfectly inelastic (there is no labor-leisure choice, as is also shown by the form of the utility function). The budget constraint becomes:

$$w_i + (1 + r_i) k_i = c_i + n_i\left(\beta_i + k_{i+1}\right). \tag{22}$$

Earnings for labor and the future value of bequests should cover consumption, the cost of rearing children (β_i per child) and the bequest left to them. The budget constraint can also be written in a dynastic form:

---

[5] The second view emphasizing human capital investment will be dealt with in section 3.2.

[6] Note that this assumption is not as arbitrary as it would seem, because it is *implied* by the assumption that A_i is a function of N_i only.

$$k_0 + \sum_{i=0}^{\infty} d_i N_i w_i = \sum_{i=0}^{\infty} d_i \left(N_i c_i + N_{i+1} \beta_i \right), \tag{23}$$

where $d_i = \prod_{j=0}^{\infty} \left(1 + r_j\right)^{-1}$.

One interesting result of this framework is that the rate of growth of consumption is independent of the interest rate, although the level of consumption is not. This is because the consumption of the next generations depends only on the net cost of "creating" it ($\beta_{i-1}\left(1 + r_i\right) - w_i$). Somewhat surprisingly a higher interest rate implies higher fertility. The reason is that with a higher interest rate it is not only future consumption that becomes more attractive, but child rearing as well, since the higher consumption of future generations raises the marginal utility of rearing one more child by raising their utility.

Assuming a CES current time utility function ($v\left(c_i\right) = c_i^{\sigma}$) they show that the steady state with stationary β , r and w fertility and consumption is characterized by the equations:

$$n^* = \alpha^{1/\varepsilon} (1 + r)^{1/\varepsilon} \tag{24}$$

$$c^* = \left[\sigma / (1 - \varepsilon - \sigma)\right]\left[\beta(1 + r) - w\right] \tag{25}$$

Because production is assumed away, as well as technological change, steady state consumption per capita is constant. This is not a growth model yet.

One important implication of the model is that a decline in the death rate of children will affect the birth rate, but only temporarily. A fall in the probability of child death will reduce the cost of rearing a surviving child (β) which will induce a rise in per capita consumption, and, through lowering the marginal utility of consumption, it will induce a rise in fertility. But as it seems from the steady state values, fertility is not affected in the long run. Outside the steady state, however, the growth rate of consumption affects fertility negatively:

$$n_i = \alpha^{1/\varepsilon} (1 + r_{i+1})^{1/\varepsilon} \left(c_{i+1} / c_i\right)^{-(1-\sigma)/\varepsilon}. \tag{26}$$

Production is incorporated into this model in Barro and Becker (1989) by assuming a standard neoclassical production function with labor augmenting technological change. Further, they suppose that the cost of child rearing in terms of goods increases with the rate of technological change, while time cost is a given share of the wage:

$$\beta_i = a\left(1 + g\right)^i + b w_i, \tag{27}$$

where g is the rate of technological change, w is wages, and a and b are constants ($a \geq 0, 0 \leq b < 1$). Since production is modelled, the interest rate and wages are endogenous making it possible for several variables to have indirect effects on fertility, that is, through the interest rate. This is the way child mortality rates can have an effect on the steady state rate of fertility through increasing the interest rate (a lower cost of raising a child increases the steady state value of capital). Another important feature of the model is the effect of technological change: if the rate of technological change increases, fertility will decrease, provided that a higher income per capita, *ceteris paribus*, would increase the fertility rate.

This model of fertility is still not enough to explain the start of modern economic growth, but it provides a plausible alternative to the Malthusian model of population dynamics. This is why some form of this model has become a basis for many of those growth models trying to explain the rise of the West from a Malthusian state of affairs.[7]

[7] In addition, although a powerful argument the quality-quantity trade-off model does not persuade every economist. In his book on parenting Caplan (2011) proposes three "un-Beckerian explanations" for the decline in fertility (ibid:115-117). One is the change of values concerning marriage, religion and child-bearing. Another possible explanation is the changes in the rules modern parents think they should follow in parenting. These rules, Caplan claims, have become much stricter. And a third possibility is changes in foresight: parents are now better at predicting the costs of raising children than before.

Concerning one important model, Doepke (2005) makes the Barro-Becker model of fertility face the facts and also extends it in reasonable ways. His extension makes it possible to differentiate between a replacement effect and a hoarding effect when looking for ways in which child mortality influences (net) fertility. The replacement effect is an increase in the birth rate to replace a child who has died. The hoarding effect is the precautionary motive in the demand for children which reflects the parents' "preparing" for child mortality. He concludes that the basic Barro-Becker model performs no worse than the more sophisticated ones which are adjusted for uncertainty or sequential child birth. When calibrated with data from 1861 England the model fails to produce results with declining child (infant) mortality *and* a declining net fertility rate. Somewhat surprisingly (and to the contrary of the quotation at the head of this section) he finds by reviewing the empirical literature that this theoretical (lack of a) conclusion is right: there is no clear-cut timing of the decline of fertility and the decline of mortality. Including human capital investment in the model does not help either, "since mortality decline lowers the cost of every child equally, it favors quantity over quality and slows fertility decline" (ibid:360). These results rule out the possibility that declining child mortality can cause the demographic transition[8].

3.1.3. The allocation of time

The quality-quantity trade-off can be adjusted by another Beckerian idea – that of the productivity of time in consumption. Becker (1965) extends the model of consumer choice by incorporating the idea that every kind of consumption takes some time although consumption activities differ as regards their time intensitivity. Consequently, an improvement in the productivity of consumption time will make their time-intensive consumption relatively cheaper and will increase their demand through the income effect. This results in an effect that is quite the opposite of the effect of an increase in the productivity of working time. Insofar as child-rearing is a time-intensive activity, a (compensated) rise in consumption productivity will increase the demand for children.

To illustrate this mechanism with a simple Cobb-Douglas example, take the problem of the household that consumes two complex goods z_1 and z_2, which are produced by the household using commodities (c) and time (t) with different (but still Cobb-Douglas) production functions:

$$z_1 = c_1^{\beta}(e_1 t_1)^{1-\beta}, \tag{28}$$

$$z_2 = c_2^{\gamma}(e_2 t_2)^{1-\gamma}, \tag{29}$$

where c_i is the quantity of commodity used in the production of complex good i, while t_i is the time used. Further, $1 > \beta, \gamma > 0$. The household has a utility function on the form

$$U(z_1, z_2) = \alpha \ln z_1 + (1-\alpha)\ln z_2 \tag{30}$$

and faces the constraints

$$c_1 + c_2 = w t_3, \tag{31}$$

$$t_1 + t_2 + t_3 = T, \tag{32}$$

where w is the real wage measured in units of consumption, t_3 is the time devoted to working, e_1 and e_2 represents productivity of time in consumption, and T is the total time the household disposes of. The solution of this problem becomes

$$z_1 = \beta^{\beta}(1-\beta)^{1-\beta} w^{\beta} e_1^{1-\beta} \alpha T, \tag{33}$$

$$z_2 = \gamma^{\gamma}(1-\gamma)^{1-\gamma} w^{\gamma} e_1^{1-\gamma}(1-\gamma) T, \tag{34}$$

[8] Note that this is in line with the conclusions of Guinnane (2011) cited in footnote 4 doubting a pattern of clear-cut timing of fertility and mortality.

$$t_3 = [\alpha\beta + (1-\alpha)\gamma]T. \quad (35)$$

That is,

$$\frac{\partial \ln z_1}{\partial \ln w} = \beta, \ \frac{\partial \ln z_1}{\partial \ln e_1} = 1-\beta, \text{ and} \quad (36)$$

$$\frac{\partial \ln z_2}{\partial \ln w} = \gamma, \ \frac{\partial \ln z_2}{\partial \ln e_2} = 1-\gamma. \quad (37)$$

Two conclusions follow. First, as already stated above, the increase in the productivity of consumption time has just the opposite effect to an increase in the productivity of labor. Secondly, an increase in wages will exert a larger effect on the consumption of the complex good that has a less time intensive production function.

Galindev (2011) incorporates this idea into his model of demographic transition. In it child rearing is only one form of leisure time that can be substituted for by other forms of leisure time. Galindev models the demographic transition as a process in which parents substitute "ordinary" leisure time for child rearing as a consequence of the fact that leisure time becomes cheaper because technological change happens in this sector, too. That is, in addition to the quality-quantity trade-off Galindev's model emphasizes two other mechanisms. First, as technology advances child quantity becomes more expensive relative to leisure goods. Secondly, as in their model faster technological change leads to a faster deterioration of human capital, a higher technological change will decrease the marginal utility of the number of children, and parents will substitute them for more leisure time. These additional mechanisms speed up the transition to the growth regime compared to the basic model sketched above.

3.1.4. Human capital

Becker et al. (1990) incorporate two additional crucial elements to the model of fertility. One is the role of human capital by supposing that the stock of human capital of the next generation depends on the human capital of the present one. Secondly, they pay attention to the role of time allocation. Raising children consumes the time of parents, that is goods and time as well have to be allocated between two roles: child rearing, which is equivalent to human capital allocation:

$$H_{t+1} = Ah_t(H^0 + H_t), \quad (38)$$

and production:

$$c_t + fn_t = Dl_t(H^0 + H_t). \quad (39)$$

Here, the first equation is essentially the production function in the human capital sector, while the second is the production sector. H is for human capital for different generations, A and D is the productivity for the sectors; c is per adult consumption, and f is the amount of goods needed to raise a child, n is the number of children. As explained in section 3.1.3 the household faces a time constraint, too:

$$T = l_t + n_t(v + h_t), \quad (40)$$

where T is the total time that is available for adults for production (l) and human capital accumulation (h), and v is the minimum time needed to raise a child. The model has two (locally) stable steady states, one of which reproduces the dynamics of a Malthusian economy, while the other describes a modern steadily growing economy.

In the Malthusian steady state, there is no human capital accumulation ($H_t=H^0$). The local stability is assured by the fact that at a low level of human capital the income effect dominates the substitution effect in the demand for children. Thus, when there is an exogenous shock that is small enough, the family will raise more children but will not accumulate more human capital, and the stock of human capital will erode back to the initial level. However, when

initial human capital is large enough, the rate of returns on human capital will be high enough for the substitution effect to dominate the income effect, and parents begin to invest in children's human capital. In this case the economy is converging to a "developed" steady state in which human capital and consumption is rising steadily. Thus there is a need for an exogenous shock in technology or in the stock of human capital to kick-start the modern period of growth. This need for an exogenous shock is a crucial feature that makes the mechanism behind this model different from the "unified growth models" that will soon be discussed.

Tamura (1996) develops this model by further adjusting it for an external effect of human capital ("social capital"). His model assumes a "conditional external effect" of human capital by assuming that parents with human capital higher than a certain threshold level are able to use social capital in the production of human capital. Although the model leads to two stable equilibria there is a critical level of human capital that becomes lower as some countries are growing, since they are accumulating human capital which can be used in producing new human capital. As a result, convergence can happen without any change in the poor country's economic policy.

3.2. Virtuous circles between technology and fertility

Even if the effect of a growth in income on fertility is not obvious at first sight, the reverse effect seems to be much simpler: the more people we have, or the higher the growth rate of the population is, the less per capita income we end up with. It turns out, however that this effect is not that simple, either. An increasing population does not only have negative, but positive externalities. "One important example of external economies involves research and technological progress" (Phelps 1968:511). As Phelps (1968:511) famously summarized this effect, "[i]f I could re-do the history of the world, halving population size each year from the beginning of time on some random basis, I would not do it for fear of losing Mozart in the process".

The simple idea that more people means more innovation, hence a higher growth rate, was formalized in the seminal paper of Kremer (1993). He developed and tested a model in which the rate of population depends on the size of population, predicting a population growth that is faster than an exponential one. By conducting cross-section and time series tests of the model he shows that the long-run history of humanity gives support to the "more people more inventions" idea (see also in Romer 2001:126-132 or Jones 2005:1097-1101). The "scale effect" coming from this simple notion became the building block of the models of innovation-driven growth and also of the models reviewed here.

Models of the so called Unified Growth Theory include this effect of population together with a modified form of the parental choice mechanism described above. Such models "unify" the models of a Malthusian economy with that of a modern growth era; in addition they are able to explain the transition between these two. Fertility choice is not based on altruism; rather it is seen as investment in human capital: parents' utility does not include children's utility but children's human capital. The general structure of the model (Galor and Weil 2000, Galor 2005) is the following[9]:

$$U_t = U(c_t, n_t h_{t+1}) \text{ (households' utility function),} \quad (41)$$

$$y_t = f(h_t, x_t) \text{ (production function of the consumption sector),} \quad (42)$$

$$h_{t+1} = h(e_{t+1}, g_{t+1}) \text{ (production function of the human capital sector),} \quad (43)$$

$$g_{t+1} = g(N_t, e_t) \text{ (rate of technological change),} \quad (44)$$

[9] For a Hungarian summary see Földvári (2007).

In these equations variables refer to quantities per capita: h is human capital per capita, x is the effective amount (which means multiplied by the level of technology) of resources available in a fixed quantity ("land"); e is the time spent on rearing a child above a minimum level (τ). It is important that the production function of human capital includes the rate of technology as a decreasing factor ($h_g < 0$), which implies that technological change will increase the demand for human capital.

The rate of technological change depends on the size of population and per capita human capital. The first effect is an inherent feature of all innovation-based endogenous growth models (Jones 1999, 2005) as it is a consequence of the nonrivalry of ideas. The nonrivalry of the products of innovation implies that its usage is not constrained by the number of people. As an idea can be used in several different production processes at the same time, the production function will have increasing returns to scale on the macro level when "knowledge" or "ideas" are treated as a factor of production.

Households can spend their potential income on consumption or on raising a child:

$$w_t h_t n_t(\tau + e_{t+1}) + c_t \leq w_t h_t, \tag{45}$$

where w is the real wage, $w_t = f_h(h_t, x_t)$.

This means that the following optimality conditions must hold:

$$MU_c \frac{\partial c_t}{\partial e_{t+1}} = MU_{nh} n_t \frac{\partial h_{t+1}}{\partial e_{t+1}} = MU_{nh} n_t h_e, \tag{46}$$

That is, the marginal utility of time spent on raising a child should be equal to that of the consumption forgone resulting from child rearing. The same is true for the number of children:

$$MU_c \frac{\partial c_t}{\partial n_t} = MU_{nh} h_{t+1} \tag{47}$$

These two conditions are important to understand how an initially low rate of technological change will lead to sustained and ever faster economic growth. The key to this explanation is the interrelationships between technological change, investment in human capital and fertility choice. Technological change increases the marginal product of human capital ($h_{eg}>0$), that is the right-hand side of equation (46). Then the rational decision of the family is to increase the time spent on child raising, reducing the right-hand side of the equation, while the resulting reduction in consumption will increase the left-hand side. Whether demand for children is reduced by technological change depends on whether the sum of two terms on the right hand side (MU_{nh} and h_{t+1}) that change in the opposite direction as a reaction to technological change will increase or decrease. As marginal utility is decreasing, the larger the product nh, the smaller the change that will be induced in marginal utility by technological advancement. At the same time, the higher h is, the larger the proportion of human capital that will be made obsolete by the same technological advancement.[10] A reduction in the right hand side is thus more probable when nh is of high value, because h is high. After a certain point in time, investment in human capital will be identical with substituting the quantity of children with their human capital. More precisely above a certain level of human capital,

$$\left|\frac{\partial h_{t+1}}{\partial g_{t+1}}\right| \frac{1}{h_{t+1}} > \frac{\partial MU_{nh}}{\partial g_{t+1}} \frac{1}{MU_{nh}}. \tag{48}$$

In an age of low technological change, the population is increasing although at a slow pace, which raises the rate of technology, which further induces a rise in population and so on.

[10] Galor (2005:242) supposes that h_{t+1} is a "decreasing strictly convex function of the rate of technological progress".

As a result, however, human capital investment is initiated, which, in addition to the growing population, further increases the rate of technological improvement. When population reaches a certain threshold, it no longer grows, only the human capital per capita will be growing, but productivity is still increasing because of these investments in human capital. This is how the economy goes through a long lasting Malthusian era and reaches the state of balanced and sustained growth. Determinants of fertility and human capital investment are strongly connected here, putting the family at the centre of the theory of economic growth.

In this case the choice about rearing children is at the same time a choice about human capital investment. When parents decide about how educated their children should be, they decide at the same time how much resource they should allocate to their education. This decision is of course a function of the yield of human capital investment. Thinking back to the simple framework presented in section 3.1, this means that an increase in the return on investment in education reduces the opportunity costs of the quality of children (p_q) – or can even make them negative. Assuming further – as model builders usually do – that human capital investment raises the productivity of the economy as a whole, then this will bring about a rise in the demand for the quality, and a decline in the demand for the quantity, of children.

This implication can also be derived from the Cobb-Douglas representation of the Becker-Lewis-model (equations (15-17)). An increase in the return on investment in children's human capital reduces the alternative cost of quality, but at the same time the increase in productivity resulting from the increase in human capital raises the alternative cost of time, and thus that of rearing children. Thinking in this framework of comparative statics, this means that π_n increases while π decreases. As a result, the demand for n decreases, while the demand for q increases. In this way we get a self-enforcing process of human capital investment through generations, productivity improvement, and choice about rearing children. UGT authors model this nonlinear process in a dynamic general equilibrium model. These mechanisms, however, do not only characterize those models that see modern economic growth as a result of a long historical process.

Focusing on human capital and fertility is not the only way to construct models in which an industrial revolution is necessary. Jones (2001) constructs a Romerian model (Romer 1990) augmented with fertility choice to account for the long term dynamics of income and population. The virtuous cycles built into his model are somewhat different from those of the Galor style models. Here, it is not human capital accumulation that lies at the centre, but innovation understood as making ideas for profit. Thus the circular processes that cause the nonlinear dynamics of the model come from the interrelationships between population and ideas as in Kremer (1993), and between innovation and the share of income which goes to innovators. He simplifies fertility choice by assuming that the utility function includes the number of births not the number of children, and that the birth rate is a linear function of time not spent on working. This gives the possibility of a demographic transition: as the wage rate rises the substitution effect away from child rearing time to consumption will counterbalance the income effect that would otherwise increase the demand for children.

What is crucial here is the production side of the model, which is in effect the Jonesian version of the innovation growth model of Paul Romer (Jones 1995, 2002:96-123, Jones 2005). Production of consumption goods is a function of ideas (A), labor allocated to this sector (L_Y) and a factor available in a fixed amount ("land", T) [11]:

$$Y_t = A_t^{\sigma} L_{Yt}^{\beta} T_t^{1-\beta}, \ \sigma > 0, \ 0 < \beta < 1. \tag{49}$$

[11] I ignore the productivity shock which Jones supposes to exist, because that would not play any role in my analysis.

Ideas are also produced by labor and ideas:

$$\Delta A_{t+1} = \delta L_{At}^{\lambda} A_t^{\varphi},\ \lambda > 0,\ \varphi < 1. \quad (50)$$

Population growth is of course the difference between all births and deaths:

$$\Delta N_{t+1} = b_t N_t - d_t N_t, \quad (51)$$

where b is birth rate chosen by households, and death rate is a function of consumption:

$$d_t(c_t/\bar{c}) = f(c_t/\bar{c} - 1) + \bar{d}, \quad (52)$$

with f being a decreasing function and d a constant.

Finally, the labor market clears:

$$L_{Yt} + L_{At} = \ell_t N_t, \quad (53)$$

where ℓ_t is working time of the household

As the model assumes away capital and assumes that land is not owned by anyone income splits between labor and innovation. The income share that goes to innovation is crucial and is set exogenously. Given that the wage rate is set on the labor market this share will determine how much labor will be allocated to innovation which in turn will determine the growth rate of the stock of ideas. When income grows, the share of income going to innovators grows as well, thus making a virtuous circle. In addition, rising consumption will reduce the death rate leading to higher population and a higher labor force, although the birth rate will start to decline at a certain point in time and eventually reach a constant rate. Although in the steady state the growth of per capita income is not necessarily positive, the steady state here is just an asymptotic possibility which occurs only when consumption goes to infinity.

He finds that in order to make the model fit the data, the share of income that compensates innovators should be substantially high. The role of these factors becomes "dominant" in the 20th century but he is very unspecific about what should be meant by the institutions that created this change, despite mentioning the "development of intellectual property rights" (in ibid: 25).

According to their estimation (Jones 2001:32) if the innovators' share in income had remained at the level it was in the 19th century the industrial revolution would have been delayed 300 years. Thus the institutions that set the share of innovators' total income are crucial in this theory, where the latter is seen as a parameter. That is, in Jones's model the industrial revolution is not as inevitable as it is in the Galor and Weil-model. All in all, from the vantage point of this paper this is inevitable, because an exogenous shock is not needed to reach the period of economic growth. Or, to put it differently, the regime of growth is not fundamentally or qualitatively different from the regime of stagnation.

3.3. Emergence of a "new economy"

Another possibility to create a non-Malthusian economy is to assume a more complicated role for human resources than an undifferentiated mass of "labor" featuring diminishing returns like "land" or "capital". When this broad anti-Malthusian assumption is given a specific form and put into a model, the result is usually that the growth regime is described by a different technology than the Malthusian one.

A famous response of this kind is what Hayek (1992:122-123) gave to the Malthusian challenge:

With the intensification of exchange, and improving techniques of communication and transportation, an increase of numbers and density of occupation makes division of labour advantageous, leads to radical diversification, differentiation and specialisation, makes it possible to develop new factors of production, and heightens productivity ... Thus labour may yield increasing rather than decreasing returns. A denser population can also employ

techniques and technology that would have been useless in more thinly occupied regions; and if such technologies have already been developed elsewhere they may well be imported and adopted rapidly (provided the required capital can be obtained). Even the bare fact of living peacefully in constant contact with larger numbers makes it possible to utilise available resources more fully. When, in such a way, labour ceases to be a homogeneous factor of production, Malthus's conclusions cease to apply….As the market reveals ever new opportunities of specialisation, the two-factor model, with its Malthusian conclusions, becomes increasingly inapplicable.

One possible translation of these ideas into the language of a formal model is to suppose that there is another kind of technology in addition to the Malthusian one. One short but very influential example of a model of this kind is that of Hansen and Prescott (2002). Their former model is similar to those discussed above in the sense that it describes the transition between a Malthusian and a developed (Solovian) regime. But, of course it is not human capital accumulation which is seen as the key, nor fertility choice; rather it is the change in the relative profitability of the available technologies that brings the catching-up process about. Here, technology with constant returns to scale is available together with the Malthusian technology which uses a fixed supply resource from the beginning. The crucial difference between the two production functions is that the Malthusian technology features decreasing returns to scale in those factors (capital and labor) in which the Solovian technology has constant returns to scale. That is, there are two Cobb-Douglas technologies in the following form:

$$Y_{Mt} = A_{Mt} K_{Mt}^{\phi} N_{Mt}^{\mu} T_{mt}^{1-\phi-\mu}, \text{ and} \tag{54}$$

$$Y_{St} = A_{St} K_{St}^{\theta} N_{St}^{1-\theta} \tag{55}$$

where K is capital, N is labor, and T is land, and in addition $0 < \phi + \mu < 1,\ 0 < \theta < 1$.

Even if the Solovian technology is known from the beginning (Hansen and Prescott 2002:1211), it can only be applied efficiently when world technology has reached a certain limit.[12] That is the time when the industrial revolution starts.

The ideas in Becker et al. (1999) are much closer to the Hayekian quotation above since they suppose that population density alters the incentives for human capital accumulation. More precisely, they suppose altruistic parents as described in section 3.1.2 and that human capital does not necessarily have a decreasing return in the human capital ("urban") sector because population density here raises productivity. This is the crucial assumption that makes a transition possible.

Doepke's (2004) model is also based on the idea that there is a second sector in addition to the "Malthusian" or agricultural one that is described by constant returns to scale as far as labor is concerned. One of the novelties of Doepke's model is the introduction of a distinction between skilled and unskilled people and labor. Another novel feature is the possibility of child labor. Producing skilled children requires not only time but costs as forgone consumption. Together with the assumption that an unskilled child will work and will increase, not decrease, parental income, rearing an unskilled child is more time intensive than rearing a skilled one.

In Doepke's interpretation the economy begins with a quasi steady state showing features of a Malthusian economy. Skilled parents have skilled children, and some unskilled parents have unskilled children, too. This is the immediate consequence of the relative time

[12] In addition, when it is profitable for firms to apply modern technology, barriers of technology adoption can still prevent them doing so. That is, what can begin an industrial revolution - given the fact that an industrial revolution has already happened in the most developed countries - is abolishing certain barriers of technology adoption.

insensitivity of raising unskilled child. When the productivity of the industrial sector reaches a critical point, the industrial revolution begins: the skill premium is increasing and, as a result, unskilled parents begin to spend more on their children's education. Thus more and more labor will be allocated to the industrial sector which features constant returns to scale. As a result of this reallocation the economy settles in an industrial equilibrium with sustained economic growth and low fertility.

Desmet and Parente (2009) also builds a theory the main feature of which is a difference between two sectors, the agricultural one (rural areas) and industrial one (urban areas). No less important is the fact that while the agricultural sector is supposed to be perfectly competitive, the industrial one is monopolistically so. And this is a key feature: as the market is growing incrementally, the variety of products also increases and that makes the demand for each good more elastic. This leads to reducing mark-ups and means the quantity of production needed to break even will be greater. This will finally lead to firms becoming larger and able to bear ever increasing fixed costs of innovation. With this innovation proving profitable the industrial revolution begins. They do not model the quantity-quality trade-off regarding child rearing. Instead, raising a child in the urban area is supposed to cost more than in the rural one[13]. As a result, when the rate of innovation increases and the urban sector grows, people will tend to have fewer children.

Much attention was given to the model of Lucas (2002) which is somewhat similar to the Becker et al. (1990) model discussed in section 3.1.4. Similarly to many authors cited in this paper following this approach, Lucas (2002) sees an industrial revolution as a transition between two eras which can be modelled by a single general model framework. As is the case with the outcome of the same broad models, he describes the static economy (also analyzed by classical economists like Malthus or Ricardo) and a stagnating income per capita world, and the steadily growing income level of the modern economy. Switching from the first to the second is the (first) industrial revolution.

Lucas's model belongs to those approaches which do not (or do not only) see the essence of the industrial revolution as a change in the nature of technological change, but put great emphasis on the role of human capital and demography. Thus according to the model, the essence of the industrial revolution is not technological advancement, because technological advancement had not been absent even before that period[14]; rather the essence is the change in the return on human capital.

With the help of his unified model he shows that exogenous technological change (identified with exogenously assumed human capital growth, because the production of human capital is a function of human capital in the same way as it is in his other seminal paper (Lucas 1988)) cannot explain demographic transition together with a rise in per capita income. That is, Lucas (2002) argues, these "exogenous" models are inconsistent with the facts of industrial revolution. As opposed to this, his endogenous model, which models child rearing as investment in human capital, is not. If the return of human capital investment rises because of the expected improvement in technology, then the trade-off between human capital and the quantity of children shifts in favor of human capital, because the utility of the next generation can more easily be increased by one unit investment of time. Thus a rise in the return of human capital will force utility maximizing parents to invest in the quality instead of the quantity of children.

Lucas shows that income and population will be constant along the steady state path in an economy with one factor of production (land) having a fixed supply, and labor. The same is

[13] This assumption is in line with the empirical result reviewed by Guinnane (2011:603) that shows that urban areas were the first places where fertility began to decline and rural areas caught up later.
[14] This proposition is documented by e. g. Mokyr (2004).

valid, of course, if land is substituted for physical capital, since in this case we have virtually arrived at the neoclassical (Solow-) model.

Incorporating human capital takes a simple form: Lucas assumes that human capital is the single factor of production, and households should only decide about how much time to spend on human capital accumulation and how much on producing consumer goods, and how many children to rear. That is, parents' utility is given in the form:

$$W(c_t, n_t, u_{t+1}) = c_t^{1-\beta} n_t^{\eta} u_{t+1}^{\beta}, \quad (56)$$

While production depends on human capital and the time share allocated to production activity(u):

$$c = u_t h_t \quad (57)$$

In the "exogenous" case the rate of growth of human (γ) capital is exogenously given:

$$h_{t+1} = \gamma h_t, \quad (58)$$

Just as in Becker et al (1990) child rearing is time consuming. Assuming that rearing a child consumes k units of time and normalizing full time to 1, we get the time constraint as:

$$u_t + n_t k_t = 1. \quad (59)$$

The reason he rejects the exogenous case is the fact that in this case technological change does not affect population dynamics. More precisely, a change in the technology of the production of human capital does not affect decisions concerning child rearing.

In the endogenous case human capital accumulation becomes the function of the time devoted to children (r):

$$h_{t+1} = \phi(r_t) h_t, \quad (60)$$

thus the time constraint will be:

$$u_t + n_t(k_t + r_t) = 1. \quad (61)$$

Assuming a special form of human capital production technology:

$$\phi(r) = (Cr)^{\varepsilon}, \quad (62)$$

he shows that technology (captured by ε) will affect fertility negatively. This creates the possibility of the demographic transition which the exogenous case excludes. Thus demographic transition, and the industrial revolution, is seen as a change in the technology of human capital production, or child rearing if you like. As Lucas himself puts it (ibid: 160) "[t]he industrial revolution required a change in the way people viewed the possibilities for the lives of their children that was widespread enough to reduce fertility across economic classes, affecting propertied and propertyless people alike".

To be able to explain the transition, Lucas (2002) has to merge two models. The one just described, which focuses on the role of human capital and the one originating with Hansen and Prescott (2002) which supposes two different technologies existing together (see later for more detail on this). Households can change between two technologies when they decide about how to allocate their time: they can either choose a Malthusian technology with a fixed factor of production (just as I supposed in section 2 to illustrate Malthusian dynamics) or they can allocate some time to the "growth technology" just described. As the productivity of human capital is growing in the growth sector, while the Malthusian technology features decreasing returns in it, more and more time will be devoted to human capital accumulation. But this transition is not a necessary implication: given certain initial conditions, the economy will converge to the Malthusian equilibrium and stagnate for ever. To leave this point the economy needs a shock, and in some cases (Lucas 2002:166-168) one has to assume that people foresee the industrial revolution, that is they begin to accumulate human capital before it becomes clear that it is economical to do so. This may provide reasons to search for institutional explanations of the transition with special attention to informal institutions. If people do not have economic reasons to accumulate they must have other reasons that tell

them that doing so is morally right. The conclusion that can be drawn from Lucas (2002) may make one believe that changing ethical values might have something to do with the first industrial revolution.

3.4. Has the Malthusian logic ever been relevant?

The Hayekian quotation above does not obviously refer to the era of the industrial revolution. It describes the conditions of an "extended order", that is, the market economy, the development of which well predates that of the industrial revolution. This suggests that the escape from Malthusian poverty may not at all have been relevant when trying to explain the industrial revolution, because Malthusian conditions had already been left behind hundreds of years ago by the time the industrial revolution began.

One way to evaluate this possibility is to look at contemporary cross-country evidence. Provided that (1) joining the regime of economic growth is equivalent to joining the process of an industrial revolution and (2) that economic growth has been absent in the poor regions of the world in the past decades, poor countries must have been in their Malthusian regimes and the laws of a Malthusian economy should still be relevant for them.

A study of this kind is Norton (2010). The underlying message of his paper is that a higher population growth is not associated with worse welfare measures and more secure property rights are in step with lower fertility. Although the first half of this conclusion does not necessarily contradict the predictions of the Malthusian model[15], the second one does. Since a probable effect of better institutions is to enhance technological change (Hall and Jones 1999), the prediction that better institutions lead to slower population growth contradicts the Malthusian model[16]. This conclusion is strengthened by the fact that the countries Norton looks at are not developed ones, and that separating the poorest group does not change the conclusion. Another point that one can raise based on Norton's (2010) results is that the relationship between income and human capital and fertility is not independent of the institutional structure. The quantity-quality trade-off may work better in a better institutional environment

As for a Malthusian view, many argue that it is exactly the Malthusian logic that worked in Western Europe and made this part of the world the richest. This kind of strong argument is put forward by those authors who argue that the emergence of a European Marriage Pattern in Western Europe was crucial in making Western Europe somewhat richer than the rest of Europe and of the world. This unique marriage pattern meant (Zanden 2009:95-141) that people in England, and the Low Countries (1) got married by a mutual consensus, (2) created an independent household by their marriage, and (3) got married later in their life than people in other countries of Europe. In addition, the proportion of those who never married was larger. He claims that three factors can be held responsible for this pattern. First, the church and its ideology, that supported the idea that the marriage should be based on the consent of the partners. The second is the way intergenerational transfers were provided: there was no dowry as in the South, but intergenerational transfers were provided on the death of the parents. That is, the children had to make their own living to be able to live in an independent

[15] Although they contradict Malthusian alarmism, they do not contradict the predictions of the Malthusian model as, for example, put forth in its simplest form in section 2. Supposing that the economy is at its steady state, the population growth can be higher only if technological change is faster, or the share of the income received by land as a factor of production receives a smaller proportion of total income. Both will affect the steady state level of per capita income positively.

[16] Galor (2011:75, footnote 8) argues that "in the context of the Malthusian model the Neolithic Revolution should be viewed as a positive shock to the level of technology". Provided that the Neolithic Revolution can be interpreted as a transition to a regime characterized by a higher level of property rights protection than the regime of hunting and gathering, this is in line with my assumption.

household. The third was the development of the labor market, which made it possible to become wage earners and to lead an independent life. This emerging new pattern of marriage affected human capital accumulation by changing the incentives of adults to invest in their children's human capital and in their own. He also argues (pp. 130-131) that as intergenerational ties became looser capital markets also became more and more important. In addition there was a need for institutions outside the family to take care of the elderly and the poor. That is, the new pattern of marriage contributed to the development of institutions that are based on a high level of trust in the community. By the help of these mechanisms the new marriage pattern was a cause of North Western Europe's drifting away from the rest of the continent.[17]

Voigtländer and Voth (2011) develop a model and provide a different explanation. They emphasize the effect of the Black Death and the comparative advantage of women over men in the pastoral type of agricultural production. The Black Death, they argue, raised the land-labor ratio in Western Europe and made the production of pastoral goods (meat, milk, etc.) more profitable, and women had a comparative advantage in this sector (as opposed to grain production). Higher wages for women in this sector provided incentives to postpone marriage thus leading to lower fertility rates. Their model helps understand why this marriage pattern did not emerge in Eastern Europe: because of the relatively high productivity of the grain-growing sector and the relatively mild impact of the plague. These did not make it a valuable option for women to work in pastoral production.

Vollrath (2011) also puts great emphasis on Malthusian-era agriculture, or more precisely the difference in labor intensities in agricultural production across Europe and Asia. In a model with a manufacturing and an agricultural sector he argues that because of the lower labor intensity in European, as opposed to Asian, agricultural production, which is explained by the difference between rice and wheat production, food prices in Asia were lower. This implies that food consumption per head was higher in Asia (as opposed to the consumption of agricultural products) which, in turn, led to higher fertility. These differences made Europe a favorable place to start the industrial revolution.[18]

There seems to be no consensus when it comes to the question as to what makes Malthusian mechanisms obsolete: the accumulation of human capital, high income or institutions. The interpretation of the pre-modern age might be crucial in this debate, because this is the time when the institutions of the market economy developed and made the Malthusian explanation out of date. On the other hand, as the argument described in last two paragraphs shows, this era might be seen as the proof that Malthusian logic worked.

4. The barriers to technological change and the adoption of new technology

Questioning the validity of Malthusian logic to explain the industrial revolution may be a reason to develop models that do not put such a primal role on demographic changes. These models are usually "Olsonian" in the sense that instead of population they emphasize institutional barriers to technological advance. The models that I will review briefly in this section take technological change as given and the question they address is why there are

[17] Zanden (2009:233-266) also revises the GDP estimates before 1800. Using a Cobb-Douglas model of the economy that fits the real wage data of the era he shows that the industrial revolution was not a sharp break with the past, at least not in England. The constant rise of income began halfway through the 17th century and continued to do so during the industrial revolution. The contrast case is the rest of Europe, where per capita income had been stagnating until the early 19th century.

[18] Although Vollrath's (2011) model is a two-sector one, the basic argument can be seen even in the simple model of section 2. Since in Asia α was higher (in equation (1)), per capita incomes were roughly the same while population growth was higher in Asia (equation (7)).

nations that cannot get richer while others constantly are. That is, these models do not explain the process of growth but the failure to join the growth process.

The core argument of these models is the importance of (some kind of) institutions in providing incentives for innovation and technological adoption. The point of departure for these theories are that (1) since factor accumulation cannot account for income differences of a magnitude of ten or twenty, technology or - more precisely - total factor productivity, must be the key; (2) differences of total factor productivity lie in institutions and economic policy. As Prescott (1998:526), one of the most influential researchers in this line of research put it, "[t]he reason that Indian workers are less productive after correcting for stocks of tangible and intangible capital is that this useable knowledge is not as fully exploited there as it is in the United States. A successful theory of international income differences must explain why this is the case".

Although, these theories may overlap with the models described in Section 3, and may be an element of some unified models, factors that are emphasized here are different. First of all, it is not human capital that is at the centre of the explanation; it is technology or knowledge that is important. The difference between human capital and technology is crucial and is of an economic kind (Romer 1992): technology consists of "ideas" that are nonrivalrous and partially excludable, while human capital is like any private good. On a very abstract level there are only two kinds of goods that are important: "ideas" and "things", and human capital is a thing, just like physical capital or land. The approaches reviewed in this section put the emphasis on technology as ideas and take their development as given. The question is why some countries are able to apply them in an efficient way while others cannot. Thus this kind of a theory can be called as exogenous as opposed to those that have been reviewed so far, which are endogenous. Parente (2001) explicitly contrasts these two theories claiming that endogenous models failed because they do not enable us to explain the key development facts, only the growth in the world's technology frontier.

One important fact that highlights the importance of such a view is the relationship between the speed of catch up and the time elapsed since the beginning of the industrial revolution. This suggests that the crucial role is played by technology adaptation and diffusion and not by human capital accumulation or demography, because those are slow processes (Parente 2008).

4.1. Technology adoption at the extensive margin

The most important work of the approach described is the book by Parente and Prescott (2000) arguing step by step in favor of the two propositions mentioned above. First, they show that capital accumulation, including human capital, cannot account for the income differences experienced in the world. On the other hand, they work out an explicit model of the mechanism explaining how the institutions affecting the activities of firms can have effects on national income. The core of the story is the firm's decisions about investing in immaterial capital, or "technology capital" as they call it (Parente and Prescott 1994:302-303). To a reach a higher level of productivity, the firm must make an investment. A crucial assumption concerning this process is that the amount of investment required to reach a given level of productivity decreases at the level of world knowledge. In addition, countries are able to erect barriers which can prevent such investments. These barriers are modelled as costs that must be covered to make the investment. To put it differently, the higher these barriers are the greater the amount which needs to be invested in technology capital to reach a certain level of productivity. Using this decision problem they reconstruct the aggregated production function, which happens to be a well-known Cobb-Douglas production function with some modification (Parente and Prescott 2000:87-88):

$$y_t = (1+\gamma)^{(1-\theta_k-\theta_z)t} A(\pi_k, \pi_z, \pi_N) h_t k_t^{\theta_k} z_t^{\theta_z}, \quad (63)$$

where y is the income per capita, k is physical capital per capita, z is immaterial capital (the accumulated amount of technology capital made by firms) and h is the length of the working week. Furthermore

$$(1+\gamma)^{(1-\theta_k-\theta_z)t} = (1+\gamma_W)^{(1-\theta_z)t}, \quad (64)$$

where γ_W is the rate of growth of world knowledge interpreted as an exogenous parameter. This reflects the pace at which the production possibilities of the world are expanding. The total factor productivity is a function of three factors:

$$A(\pi_k, \pi_z, \pi_N) = (1-\pi_k)^{-\theta_k} (1-\pi_z)^{-\theta_z} (1-\pi_N)^{-1}, \quad (65)$$

where π's represent the barriers just mentioned. Thus they are essentially transaction costs measuring the costs firms face when planning to employ different factors of production. That is, π_k and π_z are the costs needed to carry out an investment in physical and immaterial capital. More precisely, to increase the stock of physical capital with one unit there is a need for $1+\pi_k$ units of net investment. The term π_N represents labor market regulation, since it measures those working hours a firm has to employ in addition to those it would employ in the absence of regulation.

The calibration of the model gives the important conclusion that the model fits the cross data on development well, provided that the role of technology capital or the immaterial capital accumulated by firms and unmeasured by national accounts is relatively high: the share of the two kinds of capital must be about 2/3 (Parente and Prescott 2000:77-80), or the share of the technology capital must be about 0.55 (Parente and Prescott 1994:308-313).

But how do the barriers preventing technology adaptation come into being and how are they able to persist? Parente and Prescott (1999, 2000) carry on by building up a complete model and find the answer in their theory of monopoly rights. In this model, they suppose that there is a coalition in each industry that is able to set (restrict) work practices and determine the wage rate paid by the coalition members and can choose the size of the coalition. In a game theoretic model they show that there is an equilibrium with plausible parameters, in which the members of the coalition employ an inferior technology inefficiently, while potential entrants with a superior technology will not enter the market. Eliminating the barriers that make this monopoly power possible will cause, they estimate, a threefold increase in income per capita.

What kind of barriers do they mean? They put great emphasis on those that restrict the application of better work practices (Parente and Prescott 1999:1218-1219). The state can simply prohibit the application of different work practices, it can restrict the layoff of workers or simply prohibit it at least for some time, or it can raise barriers for entry into the market, mainly through regulation. But these do not really work without constraint in international competition: to what extent the government restricts international trade affects the ability of the coalition in the domestic market to deter the application of new technologies. One example of the authors is that of India (ibid:1219-1220, Parente and Prescott 2000:92-97) in the early 19th century, where workers worked with a less developed technology than their fellows in Japan. This is because, the authors say, Indian workers were able to prevent the application of modern technology. What enabled them to do that was the fact that competition from foreign firms was severely restricted in India; as a result there were no competitors of the outdated technology.

When interpreting this model somewhat broadly as a neoclassical model, one has to realize that it is not only technology that can account for such huge differences in income levels as those reflected by the facts. It may be taxes on labor as well, at least in this Prescottian perfect market view. As he argues, in models where households make decisions about leisure and consumption as well, differences in labor supply can be blamed to a large

extent for differences in income per capita, while taxes on capital cannot. Based on these models he shows that differences between the US and Japan can be explained by differences in technology while the difference between the incomes of France and the US should rather be attributed to differences in taxes on labor and consumption. Prescott (2002:11-13) identifies three factors of the institutional mix as those which seriously affect productivity. One is a "trading club" consisting of countries that make an agreement not to support their exporters or importers, who can trade with each other freely. This kind of a trading club will be a sustainable arrangement because efficient exporters will prevent others breaching the rules of free trade. The second institutional factor is the financial system. Prescott declares that a centralized financial system retards economic growth because it artificially keeps unviable enterprises alive. The third is international competition. Citing micro-level studies he argues that competition raises the productivity level of firms.

The "barriers to riches" theory is incorporated into a unified one in Parente and Prescott (2005). This means that the theory of monopoly rights that accounts for the differences in total factor productivities across countries is helpful in explaining the criterion that determines the starting date of the catch-up process (the beginning of the application of Solovian technology). Since the criterion for the profitable application of modern technology is that the total factor productivity of the Solovian sector must be above a level set by the parameters of the production function and the prices of input, changes in TFP will affect the timing of the catch-up of a country. Thus, since the barriers to adapting a new technology affect the total factor productivity, they determine the relative efficiency of the production function of a certain country. That is they can set the time when the economy begins to grow on a modern growth path. The delay in catching up is thus explained by the same barriers which explain the differences in income per capita today. The key in catching up is abolishing the monopoly rights of domestic coalitions and freeing international competition.

Ngai (2004) adjusts this model so that it is able to account for the fact that the income ratios of rich and poor countries tend to follow an inverted U-shape path in the long run, while the time lags of poor countries entering the Solovian regime of growth are considerable compared to the income lag of the Malthusian state of the West. Ngai (2004) argues that the barriers should be modelled as higher prices for investment goods for the Solovian as well as the Malthusian sector. A reduction to this delay will cause a "miracle" (such as that of Japan) but will not cause the income ratio of the miraculous country to the wealthy one to reduce.

In sum these theories differ from those that were reviewed in the previous section, and not only because catch-up is not necessary, or because technological change is treated exogenously. They also differ in what their authors consider important facts for their models to be able to explain. In the Parente-Prescott-Hansen theory demography is not a crucial fact, or, at least not a fact that has to be explained in the model, rather the demographic transition is treated as exogenous. Much more emphasis is given to the application of technology which is completely ignored in those theories which put the demographic transition at their core. This emphasis on the barriers to riches gives way to institutional explanations, since clearly the next level of research is to find reasons why some countries adopt fewer barriers than others.

4.2. Technology adoption at the intensive margin

A separate branch of the most recent literature on the differences of technology adoption is the research conducted by Comin and Hobijn (2004, 2009a,b, 2010) and Comin, Hobijn and Rovito (2006, 2008). Their research is based on the large-scale database called CHAT (Cross-country Historical Adoption of Technology) which documents the technology adoption of the intensive margin back to 1800 "over 100 technologies in more than 150 countries" (Comin

and Hobijn 2009a:1)[19]. They came to different conclusions from this database, which in some cases seem to contradict what the models described in the previous section suggest.

One of their most important empirical conclusions from these data is that that technology diffusion, once one pays proper attention to the intensive margin, usually does not follow an S-shaped logistic pattern. They estimate (Comin, Hobijn and Rovito 2006, 2008) that a logistic curve cannot be fitted onto these data in almost 60 percent of all possible country-technology pairs they have data for.

Another important fact about technology diffusion Comin, Hobijn and Rovito (2006:22) uncover is that the "relative position of countries according to the degree of technology adoption is very highly correlated across technologies. This correlation declines significantly within the OECD". This claim suggests that country-specific factors can be responsible for the large differences in technology adoption.

The institutional explanation receives further support in Comin and Hobijn (2009b). Here the authors set up a simple Romerian model to make their argument precise. What they claim is that it is lobbies that retard technology diffusion. Being a monopoly, the intermediate good producer has the possibility to bribe the authorities to slow the introduction of new technologies by regulation. Since the political costs of regulation are supposed to depend on the country's institutions, institutions will affect the politicians' decision to raise or not to raise hurdles to the new technology: the higher the political cost, the more probable it is that they will not.

The pattern they show in the data to support their theoretical claim is that there is a difference in the ways institutions affect the diffusion of technologies with predecessor technologies and those without. Their results reveal that institutional variables such as democracy, a military regime, legislative flexibility, and judicial effectiveness have an effect on those technologies that have a predecessor. All of them, except for the legislative flexibility, have the effect one could expect, since more legislative flexibility (a lack of independence) has a significant negative effect on technology diffusion. The value of these results is that they do not only tell us something about the existence of the effect of institutions on technology diffusion but they also identify a mechanism for this effect: institutions can raise the cost of lobbying.

The case for institutional determinants is not so strong in every case, however. Examining post-war economic growth in Europe Comin and Hobijn (2010) came to an opposite conclusion, because they do not find the difference between technologies with and without predecessors in Europe as they do in Comin and Hobijn (2009b) on a world level with twice as long a time horizon. More precisely, they do not find that the speed with which a technology with a close predecessor diffused after World War II is higher than with a technology without such a predecessor. This is a falsification, they claim, of the Olsonian view of development once one supposes that the war disturbed the distributional coalitions of the societies involved.

Their results based on an older version of the database (Historical Cross-Country Technology Adoption Dataset (HCCTAD), Comin and Hobijn 2004) are not perfectly in line with the above. Analysing the effect of three institutional factors, openness, type of regime and legislative efficiency, and party legitimacy index they find that the (negative effect) of legislative efficiency and the (positive effect) of openness became stronger after World War II. Their results seem to suggest that formal institutions are important determinants of technology diffusion, but the reason they matter may not be the Olsonian one in every case. Countries may not only differ in terms of their level of development or historical background, but in terms of the ways bad institutions impede the spread of new technologies.

[19] See more on the CHAT database in Comin and Hobijn (2006, 2004) and in Comin, Hobijn and Rovito (2008) and another chapter of mine in this volume where I use the data in panel regressions.

In sum, institutions seem to affect technology adoption even at the intensive margin. There is also strong evidence that this effect arises because institutions are able to reduce the lobbying activities of those equipped with the old technologies. This may not, however, be the only mechanism which is important, especially if differences within developed countries are examined.

4.3. Culture and preferences as barriers to technology adoption

A relatively new branch of the literature proposes that technological change can be blocked by culture. Pioneering works have been produced by Spolaore and Wacziarg (2009, 2011) arguing that the similarity of cultural traits between nations makes technology adoption less costly. This claim receives strong empirical support in Spolaore and Wacziarg (2009). Spolaore and Wacziarg (2011) carry this research further and look at the effect genetic distance – conceived of as a proxy for cultural distance – has on different measures of technology adoption. They also incorporate their assumption into a Romerian model of economic growth by assuming that the cost of adoption of a certain technology, at the extensive as well as the intensive margin, is increased by a larger distance of traits between a country's population and that of the technological leader. In the empirical interpretation they capture this effect by using data on genetic distance, assuming that although the transmission of such traits is not of a genetic sort, the traits of two genetically similar populations are more similar. Consequently what their regressions results show is that something that goes in step with genetic distance and that is not included among the controls they use (including different aspects of geographic location) can be held responsible for the difference in technology.

When examining the role of culture on development, one can hardly avoid mentioning Max Weber, as he is considered to be the originator of the "culture matters" school of economic development (De Long 1989). Although McCloskey (2010:140-145) claims that Weber was wrong[20], his ideas still serve as bases for building models of the industrial revolution. Cavalcanti et al. (2007) examine quantitatively Weber's original idea which assumed that the difference between Protestants and Catholics breaks down to different answers to the question, whether worldly achievement increases the probability of going to Heaven, with the Protestants giving an affirmative answer. Inserting this assumption into a formal model of Hansen and Prescott (2002) they conclude that being Protestant instead of Catholic can explain a 35 to 70-year delay of the start of the industrial revolution. This can thus possibly account for the eminence of Britain.

Doepke and Zilibotti (2008) formalize the Weberian idea in a model in which time preference and the attitude toward work are important factors triggering the industrial revolution. In their model children's preferences are shaped by the parents, whose ideas about which values to transfer to their children are formed by the occupation they anticipate they will follow. When parents are allowed to devote time to influencing their children's preference, their choice will be affected by (1) their own preferences, (2) the expected income profile of the child that reflects their expected occupational choice or "class", and (3) the expected income of the child[21]. In the absence of efficient financial markets adults with a steeper income profile will make more effort to raise more patient children, and patient people will choose an occupation with a steeper income. On the other hand, "[t]he incentive to invest into the taste for leisure depends entirely on the amount of leisure enjoyed by future members of the dynasty" (ibid:763). Because of the steep income profile of "artisans" and the flat

[20] She proposes (McCloskey 2010:143,145) that Weber himself "dropped" his own hypothesis.

[21] It must be added, however, that according to research examining twins there is little reason to expect that parenting will have much effect on children's lives as explained by Caplan (2011). Among the many conclusions he draws from twin research is the proposition that "[p]arents have a little or no effect on how much money their kids make when they grow up" (ibid:56).

income profiles of "landowners" and "workers" the pre-industrial period leads to a stratification of society into classes with a middle class being relatively patient and hard working. With the advent of the industrial revolution bringing an improving technology for artisan production and capital accumulation, the members of the middle class will soon get rich and dominate society. Their model can be used to make the industrial revolution endogenous: higher productivity technologies may have existed even in the preindustrial era but no one intended to invest in them, because people were not patient enough. When preference formation reaches a critical level, the artisans begin to use this technology and the era of economic growth is set off.

The evolutionary (quasi)genetic explanation of economic growth seems to be a new niche in the research on the industrial revolution. The human capital based unified growth theory is further developed in Galor and Moav (2002) in this way. The underlying hypothesis still focuses on demography and human capital investment and the focus on choice between quantity and the quality of children becomes more important here. Here the authors suppose that there are two types of individuals in terms of the preferences over child quality and quantity. The interrelation between human capital investment, technological change, income and fertility choice let those dynasties with a higher valuation towards child quality survive. It is a crucial assumption that preferences are hereditary within a dynasty; that is, the child's preferences will be those of the parents. The evolutionary process is supposed to be Darwinian. At the same time a scale effect is assumed away in the determination of technological change, which is assumed to be determined only by human capital investment. Since the prediction of the model is not changed, they show that these kinds of Darwinian forces are sufficient to generate the dynamics of the industrial revolution.

To formulize the idea they specify the utility function for generation t as

$$U_t^i\left(c_t^i, n_t^i, h_{t+1}^i\right) = (1-\gamma)\ln c_t^i + \gamma\left(\ln n_t^i + \beta^i \ln h_{t+1}^i\right), \tag{66}$$

$0 < \gamma < 1$, and i refers to different types of dynasties. Actually there are two types of them characterized by a higher or a lower β^i. The letter determines the marginal rate of substitutions between the quality and the quantity of children (which is $\beta^i n_t^i / h_{t+1}^i$); that is it defines the intensity of the preference of quality over quantity.

When installed into the model described in section 3.2, the hypothesis generates an evolutionary dynamics where within Malthusian conditions those preferring quality are more fertile and more productive as well. More precisely, they are more fertile because they are more productive, since they invest more in human capital, and because of their higher income they can afford to have more children than those preferring child quantity. Thus, until income per capita reaches a certain level, the proportion of the "quality-biased" will grow, leading to more and more human capital investment and a higher rate of technological change.

This reasoning sheds light on an additional virtuous circle that could make the industrial revolution possible. People with a quality biased preference will invest more in the human capital of their children, who will be more productive and will have more children leading to an even higher level of human capital per head and in general, which makes technological change faster and human capital more productive. For higher levels of income the quality biased lose their evolutionary advantage, but by that time, human capital is productive enough to make even those inclined to child quantity invest more in quality. The reason for the timing of the industrial revolution lies in the mixture of the population. When the proportion of those preferring quantity over quality is too high, an increase in the rate of technological change will not begin the virtuous circle described above and the possible revolution dies off.

This kind of downward mobility of the rich is the underlying idea in Clark (2007) as well, although the trade-off between the quality and the quantity of children and the preference for one or the other is not at the centre of the argument. Instead he emphasizes that the rich used to have a different work ethic, and modern people are the children of the rich in the past, with

the values they inherited. He argues similarly to Galor and Moav (2002), basing his proposition on data gained from wills written in the 16th and 17th century England that rich people used to have more children than the poor, because of their higher income. Clark's main thesis is that because of their higher incomes, the rich have higher birth rates and lower death rates, which implies a "downward mobility" in society before 1800, as opposed to the upward mobility of modern times. Thus eventually the descendants of the rich will dominate society. What is crucial here is the assumption that they inherit the customs, values, and work ethic of their predecessors. At the end of the day, the values of the rich of the past will be shared by most people. As Clark (2007:166) puts it "[t]hrift, prudence, negotiation, and hard work were becoming values for communities that previously had been spendthrift, impulsive, violent, and leisure loving". His evidence is indirect. First, he can show that the children of the rich became rich irrespective of their inheritance, a fact suggesting that something else, not inherited wealth was the main factor in personal material success. Second, he argues that people in the distant past were different from us. "Modern men" are more patient, more hard-working, and less happy about public executions, for example. It is not more secure property rights that have made interest rates decline in England since the 13th century. It is the fact that the more patient attitude of modern men was becoming more and more widespread through Darwinian evolution.

Another attitude that is said to be hereditary and important for development is entrepreneurial traits (Galor and Michalopoulos 2011). Adjusting the Darwinian model further they suppose that people differ in their risk aversion type, which is responsible for innovation. Since the utility functions include consumption and the number of children, too, risk-aversion reflects the elasticity of substitution between these two goods. The cost of raising children which increases with the value of human time as described in Section 3 plays an important role: at low levels of income relatively risk-loving people will decide to have more children, while at high levels relatively risk-averse will have more. As a result, technological change will be fuelled by a higher proportion of people with entrepreneurial traits until a certain level of development. When their share begins to decline, high income will play its role in advancing innovation.

While the inclusion of some quasi-genetic explanation makes the models reviewed in this subsection similar, they reflect two different approaches. In some of these explanations human traits do not have direct consequences on economic growth. They work through the indirect channel of barriers on technology adoption. In others human traits have a direct impact which implies that the spread of traits within a certain community is crucial – and this occurs with the help of population dynamics.

5. Conclusion: Malthus versus Olson

One is ultimately left with two broad views of the industrial revolution in particular, and of the catch-up process in general: a *Malthusian* one and an *Olsonian* one. By the Malthusian one I mean those theories that accept Malthusian logic even if they deny Malthusian conclusions. In this view the Malthusian era is a necessary prelude to modern economic growth and is modelled in the same unified framework. There is no need for any "abrupt" change in the underlying parameters of the model: sooner or later the economy will evolve into the stage of economic growth. These models make one see the dynamics of per capita income and population dynamics as two sides of the same coin: the explanation of the evolution of the growth regime needs to be the explanation of the demographic transition. This was the method of Malthus, too, although he explained constant misery, not the escape of poverty.

Authors creating Olsonian explanations see the Malthusian economy as different from the modern growth regime and do not identify any evolutionary means of transition between the

two. In this view the growth regime is an institutional setting, the change of which will not be caused by development automatically. It is the elimination of different constraints that make the industrial revolution possible.

Another way to put this difference is to say that in the Malthusian models institutions are considered to be of a secondary order of importance while in the Olsonian ones they are considered to be primary. In the first group of models institutions are parameters of the production of technology, which cannot reverse the conclusion that the industrial revolution is either inevitable or a matter of pure chance. Institutions can change the timing of the transition away from the Malthusian regime, but cannot change the prediction that it will occur at some time. To put it differently, change in institutions for the better is not a necessary condition of economic growth.

Models of the second type are much smaller in number and much less coherent with each other. They argue in an Olsonian way, proposing that institutions *create* the environment in which the industrial revolution has the possibility to emerge. The era of growth is not a "stage" in the long-run path of development, but an institutional environment. In this view the Malthusian era was not left behind by accumulating (human) capital, but by adopting new kinds of institutions, which may have happened well before the industrial revolution.

Identifying different determining factors of fertility and population dynamics seems to be important in this debate. What I call the Olsonian view interprets the seemingly interconnected dynamics of income and fertility as caused by institutional change, and not necessarily in the traditional Beckerian (or Malthusian) way. Indeed, the theories reviewed in the last section show that the quantity-quality trade off may be contingent on culture.

All in all, the fundamental question of development has not changed in the past thirty years. In 1981 Easterlin (1981:3) argued that the question of development is virtually the same as the question as to "why rapid technological change has been limited to so few countries". But there is a more fundamental question: Why are "nations" or "countries" the units of analysis of economic growth? Seeing this through the lenses of the models described in this paper the answer is that because modellers implicitly think that actors within a nation are more homogenous than actors between nations as regards their preferences and technology. The question is, then, what national factor makes them so: national policies, national institutions, national culture, or something else?

References

Barro, R. J. – Becker, G. S. (1989) Fertility Choice in a Model of Economic Growth. Econometrica 57. 2:481-501.

Becker, G. S. – Barro, R. J. (1988) A Reformulation of the Theory of Economic Fertility. Quarterly Journal of Economics 103. 1:1-25.

Becker, G. S. – Gleaser, E. L. – Murphy, K. M. (1999) Population and Economic Growth. American Economic Review 89. 2:145-149.

Becker, G. S. – K. Murphy – R. Tamura (1990) Human Capital, Fertility and Economic Growth. Journal of Political Economy 98. 5:S12-S37.

Becker, G. S. – Lewis, H. G. (1973) On the Interaction Between the Quantity and Quality of Children. Journal of Political Economy 81. 2:S279-S288.

Becker, G. S. (1965) A Theory of the Allocation of Time. Economic Journal 75.199:493-517.

Becker, G. S. (1981) A Treatise on the Family. Cambridge, Massachusetts: Harvard University Press.

Caplan, B. (2011) Selfish Reasons to Have More Kids. Why Being a Great Parent Is Less Work and More Fun than You Think. New York: Basic Books.

Cavalcanti¸ T. V. – Parente, S. L. – Zhao, R. (2007) Religion in Macroeconomics: a Quantitative Analysis of Weber's Thesis. Economic Theory 32.1:105-123.

Clark, G. (2007) A Farewell To Alms. A Brief Economic History of the World. Princeton University Press, Princeton.

Comin, D. – Hobijn, B. – Rovito, E. (2006) Five Facts You Need to Know About Technology Diffusion. NBER Working Paper 11928.

Comin, D. – Hobijn, B. (2004) Cross-country technology adoption: making the theories face the facts. Journal of Monetary Economics 51.1:39-83.

Comin, D. – Hobijn, B. (2009a) The CHAT Dataset. NBER Working Paper 15319.

Comin, D. – Hobijn, B. (2009b) Lobbies and Technology Diffusion. Review of Economics and Statistics 91. 2:229-244.

Comin, D. – Hobijn, B. (2010) Technology Diffusion and Postwar Growth. NBER Working Paper 16378.

Comin, D. – Hobijn, B. – Rovito, E. (2008b) A New Approach To Measuring Technology with an Application to the Shape of Diffusion Curves. Journal of Technology Transfer 33. 2:187-207.

DeLong, J. B. (1989) The Protestant Ethic" Revisited: A Twentieth Century Look. Fletcher Forum 13. 2:229-242.

Desmet, K. – Parente, S. (2009) The Evolution of Markets and the Revolution of Industry: A Quantitative Model of England's Development, 1300-2000. Working Paper 2009/06, Instituto Madrileño de Estudios Avanzados (IMDEA) Ciencias Sociales, Madrid, Spain

Doepke, M. – Zilibotti, F. (2008) Occupational Choice and the Spirit of Capitalism. Quarterly Journal of Economics 123.2:747-793.

Doepke, M. (2004) Accounting for Fertility Decline During the Transition to Growth. Journal of Economic Growth 9. 3:347-383.

Doepke, M. (2005) Child Mortality and Fertility Decline: Does the Barro-Becker Model Fit the Facts? Journal of Population Economics 18. 2:337-366.

Easterlin, R. A. (1981) Why Isn't the Whole World Developed? The Journal of Economic History 41. 1:1-19

Földvári, P. (2007) A vályogkunyhótól a mikroprocesszorig: az Egyesített Növekedéselmélet. (From the Mud-Hut to the Microprocessor: The Unified Growth Theory) Competitio 6. 2:29-49.

Galindev, R. (2011) Leisure Goods, Education Attainment and Fertility Choice. Journal of Economic Growth 16. 2:157-181.

Galor, O. (2011) Unified Growth Theory. Princeton: Princeton University Press.
Galor, O. – Michalopoulus, S. (2011) Evolution and the Growth Process: Natural Selection of Entrepreneurial Traits. Working Paper 11-07, Massachusetts Institute of Technology, Department of Economics.
Galor, O. – Moav, O. (2002) Natural Selection and the Origin of Economic Growth. Quarterly Journal of Economics 117. 4:1133-1191.
Galor, O. – Weil, D. N. (1999) From Malthusian Stagnation to Modern Growth. American Economic review 89. 2:150-154.
Galor, O. – Weil, D. N. (2000) Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond. American Economic Review 90. 4:806-828
Galor, O. (2005) From Stagnation to Growth: Unified Growth Theory, In: Aghion, Ph. – Durlauf, S. N. (eds) Handbook of Economic Growth, Volume 1A. Elsevier, The Netherlands, pp. 171-293.
Goldstone, J. A. (2002) Efflorescenes and Economic Growth in World History: Rethinking the "Rise of the West" and the Industrial Revolution. Journal of World History 13.4:323-389.
Guinnane, T. W. (2011) The Historical Fertility Transition: A Guide for Economists. Journal of Economic Literature 49. 3:589-614.
Hall, R. E. – Jones, Ch. I. (1999) Why Do Some Countries Produce So Much More Output than Others? Quarterly Journal of Economics 114. 1. 83-116.
Hansen, G. D. – Pescott, E. C. (2002) Malthus to Solow. American Economic Review 92. 4:1205-1217.
Hayek, F. A. (1992) The Fatal Conceit. The Errors of Socialism. In: Bartley, W. W. (ed) The Collected Works of F. A. Hayek. Volume I. London: Routledge.
Jones, Ch. I. (1995) Time Series Tests of Endogenous Growth Models. Quarterly Journal of Economics 110. 2:495-525.
Jones, Ch. I. (1999) Growth with or without Scale Effects? American Economic Review 89. 2:139-144.
Jones, Ch. I. (2001) Was the Industrial Revolution Inevitable? Economic Growth Over the Very Long Run. Advances in Macroeconomics 1. 2:Article 1.
Jones, Ch. I. (2002) Introduction to Economic Growth. Second edition. W. W. Norton & Company, New York.
Jones, Ch. I. (2005) Growth and Ideas. Aghion, Ph. – Steven, N. Durlauf (eds) Handbook of Economic Growth, Volume 1B. Amsterdam: Elsevier. pp. 1063-1112.
Kremer, M. (1993) Population Growth and Technological Change: One Million B. C. to 1990. Quarterly Journal of Economics 108. 3:681-716.
Lee, R. (2003) The Demographic Transition: Three Centuries of Fundamental Change. Journal of Economic Perspectives 17. 4:167-190.
Lucas, Jr., R. E. (1988) On the Mechanics of Economic Development. Journal of Monetary Economics 22. 1:3-42.
Lucas, R. E. (2000) Some Macroeconomics for the 21st Century. Journal of Economic Perspectives 14.1:159-168.
Lucas, R. E. (2002) The Industrial Revolution: Past and Future. In: Lectures on Economic Growth. Cambridge, Massachusetts: Harvard University Press. pp. 109-188.
McCloskey, D. N. (2008) Industrial Revolution. In: Hamowy, R. (ed.) The Encyclopedia of Libertarianism. Los Angeles: SAGE Publications, Inc. pp. 247-249.
McCloskey, D. N. (2010) Bourgeois Dignity. Why Economics Can't Explain the Modern World. Chicago: The University of Chicago Press.

Mokyr, J. (2004) A gazdagság gépezete: technológiai haladás és gazdasági haladás. (The Lever of Riches: Technological Creativity and Economic Progress) Budapest: Nemzeti Tankönyvkiadó.
Ngai, L. R. (2004) Barriers and the Transition to Modern Growth. Journal of Monetary Economics 51.7:1353-1383.
Norton, S. W. (2010) Population, Institutions, and Human Well-Being. Paper presented at the 14th Annual Conference of the International Society for New Institutional Economics, University of Stirling, Scotland, UK. http://extranet.isnie.org/uploads/isnie2010/norton.pdf
Olson, M. (1996) Big Bills Left On the Sidewalk: Why Some Nations Are Rich, and Others Poor. Journal of Economic Perspectives 10. 2:3-24.
Parente, S. – Prescott, E. C. (2005) A Unified Theory of the Evolution of International Income Levels. In: Aghion, Ph. – Durlauf, S. N. (eds) Handbook of Economic Growth, Volume 1B. Amsterdam: Elsevier. pp. 1371-1416.
Parente, S. L. – Prescott, E. C. (1994) Barriers to Technology Adoption and Development. Journal of Political Economy 102. 2:298-321.
Parente, S. L. – Prescott, E. C. (1999) Monopoly Rights: A Barrier to Riches. American Economic Review 89. 5:1216-1233.
Parente, S. L. – Prescott, E. C. (2000) Barriers to Riches. The MIT Press, Cambridge, Massachusetts.
Parente, S. L. (2001) The Failure of Endogenous Growth. Knowledge, Technology and Policy 13.4:49-58.
Parente, S. L. (2008) Narrowing the Economic Gap in the 21st Century. In: Holmes, K. R. – Fuelner, E. J. – O'Grady, M. A.: 2008 Index of Economic Freedom. Washington: The Heritage Foundation and the Dow Jones Company Inc.
Phelps, E. S. (1968) Population Increase. Canadian Journal of Economics 1.3:497-518.
Prescott, E. C. (2002) Prosperity and Depression. Richard T. Ely Lecture. American Economic Review 92. 2:1-15.
Prescott, E. C. (1998) Lawrence R. Klein Lecture 1997: Needed: A Theory of Total Factor Productivity. International Economic Review 39. 3:525-551.
Romer, D. (2001) Advanced Macroeconomics. Second Edition. McGraw Hill New York.
Romer, P. (1992) Two Strategies for Economic Development: Using Ideas and Producing Ideas. In: Summers, L. H. – Shah, Sh. (eds.) Proceedings of the World Bank Annual Conference on Development Economics. Washington, D.C.: The World Bank. pp. 63-91.
Romer, P. M. (1990) Endogenous Technological Change. Journal of Political Economy 98. 5. S71-S102.
Spolaore, E. – Wacziarg, R. (2009) The Diffusion of Development. Quarterly Journal of Economics, 124. 2:469-529.
Spolaore, E. – Wacziarg, R. (2011) Long-term Barriers to the International Diffusion of Innovations. NBER WP 17271.
Tamura, R. (1996) From Decay to Growth: A Demographic Transition to Economic Growth. Journal of Economic Dynamics and Control 20. 6-7:1237-1261.
Voigtländer, N. – Voth, H.-J. (2011) How The West 'Invented' Fertility Restriction. NBER Working Paper 17314.
Vollrath, D. (2011) The Agricultural Basis of Comparative Development. Journal of Economic Growth 16 .4:343-370.
Zanden, J. L. van (2009) The Long Road to the Industrial Revolution. The European Economy in a Global Perspective, 1000-1800. Leiden, The Netherlands: Brill.

Judit Kapás

Which Institutions Caused the British Industrial Revolution?

1. Introduction

The period 1760-1850, known as the British Industrial Revolution (BIR) had an enormous long term impact on Western Europe: it prepared the ground for the economic transformation that made the difference between the West and the Rest of the World (Mokyr 2005a). As argued by many (e.g., Mokyr 2005a, Lucas 2002), the major novelty brought about by the BIR was sustained growth (also called modern growth). Growth before 1750 was, if not totally absent, different in nature from what was to occur in the 19th century and later. Despite the absence of growth itself, the BIR represents the transition from the slowly-growing economy of the early modern period to the faster growth of the post 1830 period (Mokyr and Voth 2010).

Table 1 shows various scholars' estimates concerning output growth and TFP growth for Britain for the period around the BIR. What is remarkable about the period after 1750 in Britain is not output growth or TFP performance as such; these measures grew slowly as compared to their modern counterparts, but rather the change in the "quality" of the economic processes not shown by these data.

	Feinstein (1981)	Crafts (1985)	Crafts and Harley (1992)	Antras and Voth (2003)
Output				
1760-1800	1,1	1	1	
1801-1831	2,7	2	1,9	
TFP				
1760-1800	0,2	0,2	0,1	0,27
1801-1831	1,3	0,7	0,35	0,54

Table 1: Output and TFP growth rates for Britain for 1760-1860
(source: Voightländer and Voth 2006:323)

Until 1750 the slow and reversible economic growth can be explained in terms of the negative feedback effects in which economic growth created the causes of its own demise (Mokyr 2002b). Three mechanisms accounted for these effects: (1) Population dynamics. When income per capita rises, the Malthusian theory predicts a population increase, which leads to higher fertility. Such a population increase will at some point run up against some fixed resource, often believed to be food supply or farmland. (2) The limitations on human knowledge. Before 1750 those people who engaged in a systematic search for better techniques made few advances because they did not know why things worked. (3) Institutions. When economic progress took place in a society, it frequently generated a variety of social and political institutions that ended up terminating it (i.e., rent-seeking coalitions such as guilds or government-enforced monopolies).

In modern growth, according to Mokyr (2002a, 2002b), these three negative effects have been turned around and have become positive.[22] The question of what caused the BIR, i.e.,

[22] The above three negative feedbacks were replaced by the following, respectively. (1) In modern growth rich and industrialized countries have reduced demographic growth, preferring well-educated people, while poor countries are still subject to growing population. (2) The limitations on the knowledge base no longer impose as much constraint on the development of the economy as used to be the case; instead, science and technology affect one other and evolve in a mutually reinforcing way. (3) After 1750, due to the emergence of open access orders in terms of North et al. (2009), the institutional framework supported markets and the rule of law, and accordingly reduced rent-seeking and other institutional biases.

modern growth is, in the words of Clark (2003), one of history's great mysteries and is a crucial one in economics.

Various theories offer an explanation for the "why England" question. One influential group of these theories is institutional explanations. Institutional theories on the BIR in general do not deny that the essence of the BIR was macroinventions[23] which were subject to the greatest extent to exogenous factors, as argued by many (Helpman 1998, Mokyr 1990, 1999, 2002a, Lipsey et al. 2005). To a non-negligible extent macroinventions were due to talented inventors whose activities cannot be regarded as consequences only of the prevailing social, economic and demographic factors; that is, the inventions were the results of individual genius, rather than the outcome of a conscious social process (Freeman and Louça 2001). Put differently, macroinventions arose partly from outside the economy; British inventors were on numerous occasions simply lucky (Mokyr 1990) and macroinventions came simply "out of the blue". But this is not to say that institutions could not play a role; on the contrary. The uniqueness of Britain was precisely its extremely favorable institutional background for technological advances, which constituted Britain's advantage over the Continent when it comes to the "why in Britain?" question. In fact in Britain there was a congruence of favorable developments in all subsystems of society as well as the positive mutual interconnection of these developments (Freeman and Louça 2001).[24] So, macroinventions could not have come partly "out of the blue" if the institutional background had not supported this, which undermines the significance of the institutional explanations.[25]

Today the view that institutions matter for sustained growth is commonplace. The question is rather "which institutions" and "how" they matter. The question of how institutions account for the BIR is a difficult one and what is even more difficult is to answer the question of to what extent institutional changes were necessary for the BIR.

Institutional economics and economic history has provided us with different institutional explanations of the BIR.[26] These views are sometimes in harmony with one other, sometimes they conflict with or contradict one other, and sometimes one view even refutes another. In what follows I will provide a detailed overview of the institutional explanations of the BIR, by highlighting the eventual conflicts in the views. The lessons one can draw from the shortcomings of these explanations, together with some *a priori* requirements vis-à-vis such an explanation makes me turn to a more general theory of institutions, namely the theory of

[23] Mokyr (1990) proposes calling major technological advances macroinventions, which create essentially new techniques and tend to be abrupt and discontinuous. They represent a break compared to previous techniques. As Mokyr (1999) suggests, the idea of macroinventions is akin to the notion of speciation in biology: speciation is the emergence of a new category of life that is distinct from everything that existed before. By analogy, macroinventions are inventions that start the emergence of a new "technological species". They are usually followed by a large number of microinventions that improve and refine them or make them workable without changing the context of the macroinventions. Mokyr's macroinventions are in fact General Purpose Technologies (GPTs), as is also suggested by Lipsey et al. (2005).

[24] In this sense the BIR was not a sudden event; instead, it was a contingent culmination of evolutionary paths that had been in place for centuries (Lipsey et al. 2005:258). In fact, as far as the theories explaining the timing and location of the Industrial Revolution are concerned, it is possible to distinguish two kinds of explanations. One (e.g., Jones 1981, Crafts 1977) sees the evolution of the Western countries as a highly unlikely event, the result of a fortunate concatenation of circumstances. In this respect, it differs dramatically from unified growth theory (e.g., Galor and Weil 2000) where the seeds of economic development of the West were sown centuries before, and once they were there, growth was unavoidable. This latter can be paralleled with Landes (1994) who argues that both the Industrial Revolution and Britain's role in it were determined by that country's starting conditions.

[25] Of course, Britain had both a technological and an institutional "advantage" which can explain the country's development.

[26] What is more, over the last few decades a much more complete and accurate picture of the BIR has emerged on account of detailed data-oriented work by economic historians.

social order[27] of North et al. (2009) to understand why England had the industrial revolution first.

The structure of the paper is the following. In section 2 I will discuss the explanations centered on various political, legal and property rights institutions. In section 3 I will show institutional theories emphasizing the role of culture, or informal institutions, in general. In section 4 I will move to a more abstract level and turn to the theory of social order (North et al. 200) to propose this theory as an umbrella explanation for the BIR.

2. Formal institutions

One way for institutions to induce innovations and modern growth is through a direct encouragement of technological progress, that is, through the patent system. Another is through secure property rights, enforceable contracts and constraints on the powers of the executive. Institutional economics does not cease to emphasize the importance of both for development. In what follows I will summarize the literature on both, including both the pros and cons present in the literature.

2.1. Political (state) institutions

2.1.1. The Glorious Revolution: secure property rights and constrained government

North and Weingast (1989) in their influential paper identified the institutional breakthroughs in Britain with the Glorious Revolution of 1688 and its aftermath. They emphasized the importance of the fact that the Crown and the Parliament accepted complementary roles, that is, they saw each check the power of the other while building a stable and non-arbitrary state. For them, and for North (1981), constraints on the executive were paramount from the viewpoint of modern economic growth.

Three elements of governmental organization were problematic before the Glorious Revolution (North and Weingast 1989:813). First, the royal prerogative allowed the King to ignore legislation. "Second, the Star Chamber, combining legislative, executive, and judicial powers, played a key role" … "sometimes having the final word on the prerogative" (ibid p. 813). Finally, the Crown paid the judges, who served at its pleasure. The most important changes emphasized by North and Weingast are the reversal of these three practices as a result of the Glorious Revolution.

Thus the political history of England before the BIR, in their sense, reflects two propositions: (1) the establishment of secure and stable property rights for private persons is a necessary and sufficient condition for economic growth, (2) the establishment of such rights depended on the creation of representative democracy. Therefore they believe that there was an inanimate relationship between the Glorious Revolution and the BIR in the sense that the Glorious Revolution created the preconditions for the BIR.

The Glorious Revolution, in their understanding, seems to be a turning point from the viewpoint of the appearance of modern growth, for the following reasons. First, by the beginning of the eighteenth century, the English government was sufficiently constrained in its powers that private initiative and enterprise flowered. Second, the relevant constraints on the state were primarily legal and were embodied in the highest levels of law, i.e., the constitution. Third, the decisive moments of constitutional change were in the years immediately following the Glorious Revolution of 1688, with the passage of the Bill of Rights of 1689 and the Act of Settlement of 1701. Fourth, these constitutional developments were the

[27] By social order they mean the complex of military, political, economic and religious institutions of social organization.

product of design by forward-looking individuals. Fifth, the state was not predatory, due to the control of Parliament. The importance of this fact is that the profits of the technological breakthroughs generated for entrepreneurs were not expropriated by the state.

In North and Weingast's account, by changing the "rules of the game" which determined the costs and benefits of different actions taken by the king, the Glorious Revolution solved the problem of credibility because it was either not feasible or not desirable for the king to renege on commitments after 1688. At the same time, North and Weingast emphasize that the new rules were self-enforcing because of a credible threat of removal of any Monarch who violated them. These new institutions served to "limit economic intervention and allow private rights and markets to prevail in large segments of the economy" (ibid p. 808).

In sum, North and Weingast characterize the Glorious Revolution as a change in the *de jure* institutions, alternatively formal institutions, specifically emphasizing how this constrained the future actions of the king. What they suggest is that the security of private property rights and, in parallel with this, the commitment of the state not to infringe these rights, were the *sine qua non* preconditions for the BIR.

Epstein (2000) is largely in harmony with the above views by offering a more sophisticated view of the impact of political constitutions on solving coordination problems and permitting Smithian growth, that is, growth dependent on efficiency gains from spatial specialization and division of labor. He suggests that economic freedom and limited government, due to the Glorious Revolution, are the keys to economic growth. In his opinion, the essential element for growth is undisputed jurisdictional sovereignty over the realm both in economic and political spheres. This behavior of the government probably rested on the notion of free trade, an idea which was introduced by Adam Smith's book: profit-seeking activities were seen as promoting social welfare.

Restraints on government initiated by the Glorious Revolution are important from the viewpoint of the improvements in public and private finance in England, too. Klerman and Mahoney (2005), by adhering to the argument of North and Weingast (1989)[28], emphasize the crucial role of judicial independence in promoting financial development. Judicial independence plays a central role in constraining the government as it makes it more difficult for the government to engage in opportunistic behavior. Judicial independence is clearly an 18th-century phenomenon in England: judges gained formal independence in a series of steps starting in 1701. In particular, Parliament enacted statues granting judges security of tenure and increasing salaries. So, the role of the Glorious Revolution is quite clear in this process, particularly in assuring *de jure* independence.[29]

Klerman and Mahoney (2005), in their empirical analysis, investigate the effects of two aspects of judicial independence – namely the security of tenure for judges and judicial salaries[30] – on abnormal returns of securities in London. They find that increases in judicial independence increased the value of financial assets.

But the relationship between the Glorious Revolution and the private economy is rather controversial in the literature. While North and Weingast argue that due to the security of property rights, the cost of capital to the British government declined substantially, a phenomenon which they interpret as a fall in the required risk premium, Sussman and Yafeh (2006) provide evidence that the risk premium on English government bonds remained high until the mid-19th century. Accordingly, their analysis contradicts the view that the government's credibility improved after 1688. Sussman and Yafeh (2006) also show that the volume of British government debt remained low for nearly a century after the institutional

[28] The Revolution of 1688 led to a "fundamental redesign of the fiscal and governmental institutions" (North and Weingast 1989:804).

[29] According to Klerman and Mahoney (2005), *de facto* judicial independence existed even before the 18th century, while *de jure* independence was established by the Act of Settlement.

[30] Note, however, that the concept of judicial independence is much broader than is understood by them.

changes described by North and Weingast. And what is more, they show that British interest rates moved in tandem with Dutch interest rates, suggesting that Britain did not embark on a different path following the institutional changes of the late 17th century; instead, some of the trends in interest rates in Britain were actually shared by the Netherlands as well.

In contrast to the above, Quinn (2001) argues that the risk premium on government debt declined in the 1690s, and interest rates on private debts increased. To provide evidence for this claim he collected rates of return on loans held by a London goldsmith-banker named Sir Francis Child. Due to an analysis of Child's portfolio, he rejects the hypothesis that an increased supply of loanable funds was the dominant result of the Glorious Revolution. Quinn (2001) also finds that in the 1690s Child and his customers began to own government debt, East India Company bonds, and other corporate assets. This evidence of advances in financial technique confirms that private-sector behavior was being altered by the revolution in public finance.

De Long and Shleifer (1993) also deal with the effects of a limited government on economic prosperity. They empirically analyze the relationship between the type of the government (either constrained or absolutist) and the growth of urban populations as a measure of economic growth for medieval Europe for the period circa 1050-1800. They find evidence for the view, put forward by many scholars (e.g., Olson 1991), that limited governments are more concerned with economic development than absolutist ones. The reason behind this, in brief, is that a constrained government bound by rules tends to impose lower and less-destructive taxes. However, this well-known fact alone cannot explain why England led the Industrial Revolution, since England was not the only country to have a constrained government. While De Long and Shleifer (1993) do not focus on the period of 1650-1800, I think their results can point to the fact that the English case was substantially different from those in other countries with limited government. This is the finding that 17th and 18th century England exerted the only significant shift on the regression coefficient: the removal of England cuts the estimated effect of an absolutist government on city growth by almost 30 percent. This suggests that Britain's institutions other than limited government and the rule of law also mattered when it comes to the causes of BIR.

Thus, the question is not that whether secure property rights (economic freedom) can promote development or not – we know that they can – but whether they really were at the heart of the factors leading to the Industrial Revolution. Fortunately, important historical databases have recently become available, making it possible to empirically test the above hypothesis.

One approach to examining whether the Glorious Revolution was crucial to the future development of Britain is Murrell's (2009). Murrell examines empirical evidence regarding when change came to England. He uses the econometrics of unknown structural breaks to estimate the years in which breaks occur in many data series related to various socioeconomic factors[31]. He found 58 break dates, 29 of which fall before 1688, with 13 of the significant ones doing so. In sum, his results establish that there is nothing in usable data sources to suggest a structural break in development as a result of the 1689-1701 measures; in other words, improvements were under way before 1688.

Murrell also analyses the content of two critical laws, namely the Bill of Rights and the Act of Settlement to see whether their clauses really did define the nature, power and duties of the government. The results are interesting: of the fifteen measures in the Bill of Rights, only two were unarguably new, and many of them did not survive as viable constitutional measures, meaning that it is simply impossible to characterize the Bill of Rights as providing either new legal protection of property rights or new defense against prerogative taxation or

[31] Murrel (2009) has in total 58 variables, including various price indexes, product prices, growth rates, GDP data, data related to patents, data related to the severity of punishments. See Table 1 in Murrell (2009).

new Parliamentary rights on taxation. Figure 1 shows that the proportion of severe punishments for property crimes did not increase at the times of the Glorious Revolution. As for the Act of Settlement, of the nine distinct measures five were new, two were old, and two reflected much historical precedent. Of the five truly new, four did not survive. So Murrell's (2009) analyses provide strong evidence that the Bill of Rights and the Act of Settlement comprised mostly old measures that survived and new measures that did not. To sum up, Murrell (2009) argues that the constitutional changes of 1689 and 1701 largely summarized what was already in existence in Britain.

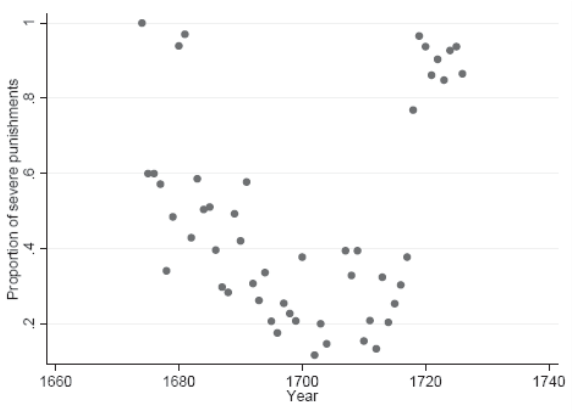


Figure 1: The behavior of judges: severe punishments as a proportion of all punishments for property crimes (source: Murell 2009:87)

The view that the enforcement of property rights by the state was crucial to the BIR is strongly contested by Clark (1996), too. In accordance with Murrell (2009), based on his empirical analysis, he argues that nothing special happened in 1688 from this point of view: secure private property rights existed in England almost as early as 1600, or probably earlier (see also Figure 1). He also disputes the view put forth by North and Weingast (1989) which claims that economic growth needs a stable and non-autocratic political system. North and Weingast's argument runs in the following way. Government interest rates declined after the Glorious Revolution (from 10 percent to 3 percent), which is a sign that the government operated differently after 1688: a new stable government was established with private capital markets.

To test whether North and Weingast's insight is true, Clark analyzed whether important political events of the 16th and 17th centuries affected rates of return in the private capital market. He found that the Glorious Revolution seems to have had no effect on rates of return in the English economy between 1660 and 1730[32]: the rates of return on capital fell in the 100 years prior to the BIR, which thus shows there was no connection with political events.[33] That is, the financial revolution started before the BIR; data show that capital assets were traded in an integrated market even before the BIR. All this means that the private economy largely before 1688 was basically insulated from political events (Clark 1996).

[32] In a formal test of three series (real property: land, houses, tithes; rent charges; and bonds and mortgages) Clark (1996) demonstrated that none of the political or military convulsions of the 17th century seems to have had any significant effect on private capital markets in the predicted direction.

[33] Note that Murrell (2009) also finds, based on an analysis of institutional and administrative innovations, that many key developments affecting government finance were a product of the era before 1688.

So the view that Britain's advantage in leading the Industrial Revolution was due to its efficient enforcement of property rights after 1688 needs to be revisited. Of course, this is not to say that secure property rights may not be necessary conditions for growth, but, based on the above-cited empirical analyses, they are not sufficient, and an adequate explanation for the BIR requires factors other than the emergence of stable private property rights. This argument also suggests that the major role of Parliament at that time may not have been to secure property rights[34], but was something different.

Clearly, Parliament seems to have had a crucial importance in inducing favorable changes both in technology and institutions. It was a meta-institution that had the legitimacy to change other institutions. As explained by Mokyr (2008) Britain was almost unique in Europe to have developed a parliament after 1650, which acquired a position of legitimacy and power. Mokyr and Voth (2010) emphasize another aspect of the British political constellation which seems to be central, namely that *de jure* and *de facto* power coincided to a great extent: both were in Parliament's hands.[35]

In what follows I will summarize the theories ascribing importance to the Glorious Revolution on other grounds than securing property rights.

2.1.2. The Glorious Revolution: a new political equilibrium, lower rent-seeking and reorganized property rights

Pincus and Robinson (2011) revisit North and Weingast's (1989) argument and the evidence supporting it. They argue that North and Weingast were correct in their belief that the Glorious Revolution was a decisive turning point in the political and economic history of England. However, they suggest that the causal account provided by North and Weingast is not substantiated by what actually happened in the wake of the Glorious Revolution.

As opposed to North and Weingast's argument that the Glorious Revolution established new *de jure* rules, Pincus and Robinson (2011) argue that rather than being *de jure*, the most significant of these were *de facto*, alternatively "informal", in the sense that they emerged from the context of the change in the English political equilibrium.

After a profound analysis of English history, they come to the conclusion that nothing in the Declaration of Rights, or in the Revolution Settlement of 1689, created a new method for Parliament to audit royal spending, provided new guarantees for the supremacy of common law courts, or provided new credible threats of removal against miscreant rulers, nor did the Settlement introduce more stable or predictable governments. Pincus and Robinson (2011) also claim that 1688 did not change the security of property rights.

What the Glorious Revolution brought about is that the Whigs came to power and they imposed their particular vision on the state. The Whigs were increasingly becoming the political party of the manufacturing sector, the export-driven long distance trades, and the newly dynamic cities and towns. The newly dynamic economy shifted the social balance. Manufacturers, urban dwellers, and colonial traders became much wealthier.

These arguments suggest that the right way to think of the Glorious Revolution is as part of a change in the political equilibrium. In Pincus and Robinson's account the Glorious Revolution was not significant because it was a change in the *de jure* rules, but it was important in helping to cement a change in the distribution of *de facto* power in the country in favor of the newly dynamic manufacturing middle classes. This consolidated a change that had already been under way. Moreover, the importance of this change for future economic

[34] It is worth noting that Olson (1982) argued that the security of property rights might have ambiguous effects: bad property rights (a privilege) could be damaging to economic development even if they were well-secured.

[35] Bearing in mind the model of Acemoglu et al. (2005a) explaining how political institutions affect economic performance, the significance of the above is hardly questionable.

growth did not stem from the fact that it established a credible commitment to property rights. Rather, the Glorious Revolution was important because in the new political equilibrium Parliament was dominated by the Whig Party which had a specific program of economic modernization.[36]

The Glorious Revolution is given a different significance by Mokyr and Nye (2007) who argue that Parliament contributed to reduce rent-seeking activity. In their view, the success of Britain was the result of the emergence of a progressive oligarchic regime that divided the surpluses generated by the new economy between the large landlords and the newly rising businessmen, and that tied both groups to a centralized government structure. The government provided uniform rules and regulation. Wealth (inherited or earned) remained the source of political power, but as its base broadened, its political objectives shifted.

This process was facilitated by the existence of Parliament, a meta-institution that wrote the rules according to which other institutions changed.[37] Parliament changed British laws in accordance with what its members viewed as their interests. There is no question that Parliament was a mechanism by which the richest and most powerful families in England manipulated the system to advance their interests. Clearly, in the decades after the Glorious Revolution rent-seeking activity was the norm. But at some point, a gradual change in the culture of legislation occurred: purely redistributional actions began to lose ground.

Parliament became the arbitrator of disputes between special interest groups. So, basically elites allowed processes to unfold that ultimately destroyed some of their entrenched positions. The results were that production shifted from agriculture to industry, from local to national markets. But why did elites create democracy when in fact political power is the key to the distribution of income?

Aceomoglu and Robinson (2000) suggest a convincing answer to this question. They propose a "political loser hypothesis" (as opposed to an "economic loser hypothesis") which argues that it is groups whose political power – not economic rents – are eroded that will block technological advances.[38] If agents have and maintain political power, i.e., are not political losers, then they have no incentive to block progress. And this is precisely what happened in Britain after the Glorious Revolution: the landlord class retained its political power.[39] Accordingly, landlords did not use their political power to seek more rents – because the cost of transformation was not on landlords, but on the consuming middle classes – instead, as they were part of the ruling elites, they passively assisted economic and technical transformations.

So, as opposed to North and Weingast (1989), according to whom the major role of Parliament was to serve as constraint on the executive, Mokyr and Nye (2007) see its role in reducing rent-seeking redistribution. The result was that in Britain there was an environment in which the Olsonian "stationary bandits" did not create obstacles high enough to suppress the technological potential of the country, which was critical for the BIR.

Another major function of Parliament is depicted by Bogart and Richardson (2011) when focusing on its role in altering property rights. The fact that property rights were secure in Britain largely before the Glorious Revolution (see Figure 1 or Clark 1996) does not mean

[36] Pincus and Robinson (2011) show that Whig institutions such as the Bank of England, the Land Tax, and the new East India Company, which favored economic development, were largely due to a provision of essential infrastructure for their war effort.

[37] Olson also (1982:78-83) pointed to the Glorious Revolution as a watershed. According to him, the Glorious Revolution weakened most distributional coalitions.

[38] In the same spirit Mokyr (1990:243) notes about Britain, "… the landowning elite, which controlled political power before 1850, contributed little to the Industrial Revolution in terms of technology or entrepreneurship. It did not, however, resist it."

[39] Despite the franchise reforms of 1832, 1867 and 1884, the House of Lords guaranteed the security of landed interests until the Liberal government of Asquith after 1906.

that there could not have been problems with them: Britain's property rights system, inherited from the past, was inflexible. The role of Parliament, according to Bogart and Richardson (2011) consisted in reorganizing rights to land and resources, which enabled landholders and communities to exploit opportunities that could not be accommodated otherwise; entrepreneurs, landowners, and localities would have forgone investment opportunities without alterations in their property rights.

As analyzed by these two scholars, holders of equitable estates could neither mortgage, nor lease, nor sell much of the land under their control; holders under many types of tenures could transfer property only to particular persons or members of a local community; and residents in common-field villages often had to keep land in traditional uses. To overcome these problems, Parliament established procedures for processing petitions from groups hoping to reorganize rights to land and resources.

Bogart and Richardson (2011) focus on three kinds of acts, namely estate, statutory authority, and enclosure acts, from 1700 to 1830. Figure 2 shows the activity of Parliament in this respect. These acts loosened constraints on investment inherent in Britain's medieval landholding system. Estate acts enabled holders of property to take certain actions prohibited by the rules under which they had inherited their land. They were necessary because the inheritance system limited estate holders' power over their property, particularly the ability to sell or lease land. They facilitated the enforcement of contracts by clarifying permissible transactions and the rights of pertinent parties. Acts establishing statutory authorities created new organizations that built, operated, and maintained infrastructure and public services. Enclosure acts disbanded collectively managed common-field villages and assigned to individuals rights to particular pieces of property.

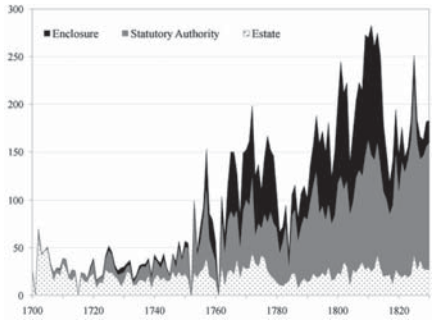


Figure 2: Number of acts reorganizing property rights, 1700–1830
(source: Bogart and Richardson 2011:250)

To account for the significance of these acts, Bogart and Richarson (2011) empirically examined the relationship between certain economic variables (the real interest rate and the volume of foreign trade) and legislation reorganizing property rights. Their major finding is that causation runs from changes in economic conditions to changes in the quantity of legislation. So, Parliament responded elastically to changes in public demands for reorganizing property rights. According to Bogart and Richardson (2011), relaxing these constraints was probably a necessary condition for English economic development.

2.2. Patents

Among the institutions affecting the BIR intellectual property rights are traditionally thought of as being extremely important: "Innovation will be encouraged by modifying the institutional environment, so that the private rate of return approaches the social rate of return. Prizes and awards provide incentives for specific inventions, but do not provide a legal basis for ownership of intellectual property. The development of patent laws provides such protection. … [B]y 1700 ... England had begun to protect private property in knowledge with its patent law. The stage was now set for the industrial revolution" (North and Thomas 1973:155-156).

North (1981:164-166) provided the canonical statement that the rate of technological change depended on the inventor's ability to capture a larger share of the benefits of his invention. Patents are seen as one of factors inducing innovation through providing incentives to innovators. But was this really the case during the BIR? Did intellectual property rights really matter for inventions? As we will see below, the answer given by the literature to the question of whether intellectual property rights afforded to inventors during the BIR levered technological and industrial progress is generally negative.

Dutton (1984) was the first to consider in a systematic way the connection between the patent system and inventive activities in the BIR. He argues that a group of "quasi professional inventors" emerged during the BIR who took their profits through the sale or licensing of their intellectual property rights. Sullivan (1989) confirms this view by showing the existence of a structural break in 1757 in the time series of total British patents: after 1757 there was acceleration in the pace of invention, which is demonstrated in Figure 3.[40]

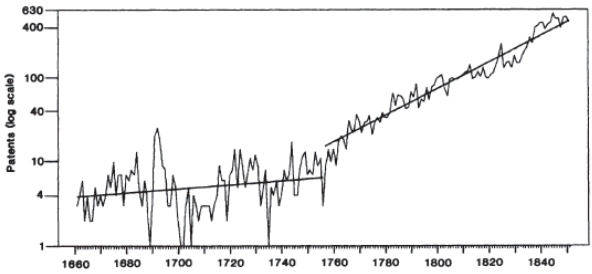


Figure 3: English patents with estimated trend lines, 1661-1851
(Source: Sullivan 1989:430)

Now the question is whether the increase in the number of patents from 1757 onwards can be interpreted such that patents were the cause of the BIR? The answer is basically negative. Two kinds of counter-argument occur in the literature.

One line of argument proceeds by shedding light on the bad characteristics of the British patent law and patent system. MacLeod's (1988) evaluation of the British patent system is very cautious when she draws attention to the unorthodox use of patents, the most typical case of which was where the patent was used to obtain support through specific government concessions. Mokyr (2010b, 2010a) also stresses that the idea that technological progress depended on inventors' incentives through the patent system is dubious for both historical and theoretical reasons.

[40] However, one has to be cautious when evaluating the increase in patents. Sullivan (1989) argues that the increase in patenting may be a sign of the increase in patentable inventions, and not in inventions in general, because many inventions are not patentable.

To a large degree, patent institutions in Britain, created in 1624, offered rather limited incentives to investors (Khan and Sokoloff 2004). The British patent institutions had many defects. The fees were extremely high, the system was too complicated, there was no examination, and as the "first and true inventor" included importers of inventions that had been created abroad, the interpretation had to be proved (Khan and Sokoloff 2004). No patent was fully valid till it had been tested by the courts, but judges were on the whole hostile to patentees, and people rarely sued: between 1770 and 1850 only 257 patent cases came before the courts, out of 11.962 patents granted (Dutton 1984:71). In addition, the patent system was riddled by the widely-condemned practice of so-called caveats, which were an expression of the intent to file a particular patent later on, and by the acquisition of a block on any application before warning the filer. Finally, many patents were infringed upon, and judges before 1825 or so were often hostile to patentees, considering them monopolists (Mokyr 2010a).[41]

Patents laws were revised only in 1852, but the process continued to discourage technological creativity. In addition, Britain's advantage over its neighbors was only limited in this respect since many European countries adopted a patent law similar to Britain's.

Another line of the counter-argument against the strategic importance of patents in the BIR argues the reverse interpretation, namely, that the growth of patenting after 1760 followed industrial development. In a series of papers, Greasley and Oxley (1997a, 1997b), and Oxley and Greasley (1997) consider possible causal linkages between industrial production and other aggregate level data that have traditionally been identified in the economic history literature as potential candidates for "drivers of growth", including patents. Here, typically bidirectional causality between patents (levels or growth rates) and industrial production (levels or growth rates) was identified.

Greasley and Oxley (2007) add to these debates by investigating the causal links between patenting activity and industrial output at the sector level during the period 1780–1851. Using time series methods they consider the existence of bi- and multi-variate causality between patents and 16 sectors of the British economy. The two scholars, based on their empirical results, conclude that the rise in patented inventions after 1780 was a consequence, not a cause, of the BIR.

Since patenting procedures and institutions did not change materially in the period to 1851 (see also Sullivan 1989), the simple implication is that the value of protecting the intellectual property embodied in technical inventions rose sharply during the Industrial Revolution. These findings offer support to those historians, including MacLeod (1988), who argue that inventors ''rediscovered'' the patent system after 1760 and learned to use it to best effect.[42] Greasley and Oxley's (2007) results show that patenting activity was particularly associated with the ''new'' fast growth sectors of the BIR, notably cotton and iron. Probably this increased propensity to patenting was caused by an increased awareness of the benefits of patenting (Sullivan 1989).

Allen (1983) also emphasizes that patents should not be seen as key factors in British technological progress. He draws attention to the role of collective invention[43] before the BIR, akin to modern open-source technology (Nuvolari 2004).

---

[41] MacLeod's (1986) analysis of the 1690s boom in patenting is a good example of how the mere number of patents is highly misleading in evaluating inventions. She argues that the spectacular increase in the number of patents does not indicate an increase in inventions, but the fact that the ready availability of capital promoted many worthless new projects.

[42] The fact that patents did not cause industrial growth together with the fact that not all inventions are patentable indicates that patents are a not a satisfactory measure of technological progress during the BIR.

[43] In collective invention settings, inventors freely release to one another pertinent technical information on the construction details and the performance of the technologies they have just introduced. This represents knowledge spillovers. As an example see Nuvolari's (2004) steam pumping engine case.

More recently, Mokyr (2010a) draws attention to cultural factors when arguing that patents were not crucial in Britain (see also section 3). Originating in the Baconian program[44], most of the people who generated useful knowledge during the BIR did not do so primarily to generate income directly. Their primary aim was not to maximize profit, but rather to signal and demonstrate to their peers their intellectual and technical capabilities. There was an intuitive sense that knowledge should be free-access because anything that limited access to useful knowledge was bad for the Baconian program. There was also a moral sense that inventors, like scientists, were serving the public good, and should be rewarded by honors, not necessarily financial rewards.

So, the above arguments place serious doubt on the strategic importance of the patent system in advancing technology. Just to give one additional support for this claim, remember that the key-technologies that lay at the heart of the BIR, such as high pressure steam engines, steamboats, iron production techniques, etc. were also developed in a collective invention fashion, and consequently they were never patented.[45]

2.3. Private-order institutions

Institutions that created bridges between prescriptive and propositional knowledge, in the spirit of the Baconian program, such as universities, polytechnic schools, research institutions, museums and agricultural research stations were also important in facilitating economic progress in Britain. These together with various other institutions (professional journals, technical encyclopedias) made the country uniquely suited to induce technological progress.

In addition, at that time technical seminars and scientific associations were commonplace in England. As Mokyr (2010b) argues, Britain created private organizations that encouraged innovation and the dissemination of knowledge beyond the patent system. A notable example is the Royal Society of Arts, founded in 1754, which aimed explicitly at disseminating existing technical knowledge, at augmenting it through an award program[46], encouraging networking, and the publication of periodicals. Another institution was the Royal Institution which was founded in 1799, devoted to research and charged with providing public lectures on scientific and technical issues. These private institutions together with The Mechanics Institute were adequate for the creation of a stimulating environment for most British inventors. Figure 4 shows the mushrooming of scientific organizations in the 18th century.

Despite these institutions aimed at disseminating scientific ideas, a unique characteristic of the BIR was that before 1850 the contribution of formal science to technology remained modest (Mokyr 2002a). Much of the technological progress came from the semi-formal and pragmatic knowledge generated by great engineers, or in other words, by a technological elite of inventors, engineers, mechanics and skilled craftsmen, whose dexterity and ingenuity was critical (Mokyr and Voth 2010). This seems to be true when thinking of the direct effect of science. However, examples of the importance of science and mathematics to some of the inventions of the BIR can certainly be found. It is equally true that many of the most prominent breakthroughs in manufacturing, especially in the mechanical processing of textiles, were not based on science, and that in other areas of progress, such as steam power,

[44] Mokyr (2005b) argues that the intellectual origins of the Industrial Revolution are traced back to the Baconian program of the seventeenth century, which aimed at expanding the set of useful knowledge. The eighteenth-century Enlightenment in the West carried out this program through a series of institutional developments that both increased the amount of knowledge and its accessibility to those who could make best use of it. The Industrial Enlightenment in Mokyr's (2005b) terms was about the expansion of useful knowledge and consisted of the emergence of institutions devoted to the flow of ideas.

[45] Moser (2007) also shows that only a small portion of the significant inventions made in Britain were ever patented by the middle of the 19th century.

[46] Note that only inventions which had not been patented were eligible for the Society's prizes (Mokyr 2010b).

progress occurred on the basis of trial and error, not a deep understanding of the underlying physical processes. As argued by Mokyr and Voth (2010) trial and error, serendipity, and sheer intuition never quite disappeared from the scene.[47]

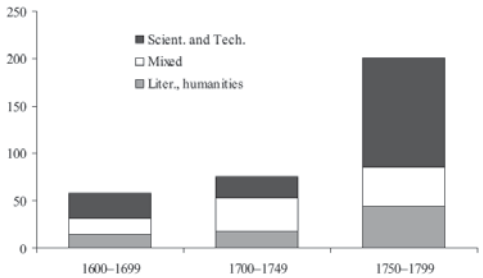


Figure 4: Scientific societies by period and main purpose
(source: Mokyr 2005b:335)

Due to the practice-oriented character of Britain, many inventions were imported, further developed and utilized in Britain.[48] Technical training through master-apprentice relationships was at a relatively high level, favoring learning by doing and creating a favorable climate towards inventions and experimentations. Apprenticeship was an ideal way to transmit the kind of tacit artisanal knowledge that was essential to competence.

To sum up, private-order institutions – mainly those that can be associated with scientific dissemination – were mushrooming in England, and clearly, these institutions were embedded in the Enlightenment and informal institutions (see also section 3).

2.4. Markets as the cause of the industrial revolution

Adam Smith (1776) was probably the first economist to emphasize the role of the market in understanding why England was the first country to experience an industrial revolution. He stressed the importance of specialization through which markets, and particularly the size of the market, can induce development. In his view, the best way to improve productivity was division of labor: one is more productive if one concentrates on one thing than if one tries to do several. One can acquire whatever one does not produce from others who concentrate on other things. The degree of the division of labor is proportional to the size of the market, which makes all factors limiting the extension of the market – such as transportation costs, or the likes of non-tariff barriers to trade – obstacles to development.[49]

The Smithian story of development can easily be continued: urbanization created a feedback process in the sense that large markets where wealthy merchants were located attracted artisans and fueled a second wave of specialization and division of labor. Craftsmen were attracted by other craftsmen, they could share some costs of production, provide

[47] In contrast, according to Lipsey et al. (2005), the development of science, mainly Newtonian mechanics was a necessary precondition for the BIR: "Indeed, it does not seem an overstatement to say that Newtonian mechanics provided the intellectual basis for the First Industrial Revolution, which in its two stages, was almost wholly mechanical" (Lipsey et al. 2005:241).

[48] In many cases the first successful applications of the new techniques appeared in Britain. Among these the most remarkable were gas-lighting, chlorine bleaching, the Jacquard loom and the Robert continuous paper-making machine. See Mokyr (2005a).

[49] Note that Smith was well acquainted with some of the institutional foundations of development, too: "... commerce and manufactures gradually introduced order and good government, and with them, the liberty and security of the individuals" (Smith 1776, Book 3, Chapter 4, http://www.marxists.org/reference/archive/smith-adam/works/wealth-of-nations/book03/ch04.htm).

protection to each other via the creation of guilds, and they could also complement each other. So as the urban population increased, the possibilities for division of labor increased as well.

The division of labor was not the only cause of growth in the pre-modern era, of course. What is clearly missing from the above account of pre-modern progress is an explanation of innovation. On the other hand, the virtuous cycle described above seldom worked so well; there were numerous bottlenecks such as barriers to trade or contractual insecurities.

Acemoglu et al. (2005b) also placed emphasis on markets in inducing growth by offering an explanation for Europe's rise based on the interaction between Atlantic trade and medieval political institutions. They suggest that Atlantic trade – the opening of the sea routes to the New World, Africa, and Asia and the building of colonial empires – contributed to European growth between 1500 and 1850 through an indirect institutional channel as well as via direct effects. Their hypothesis is that Atlantic trade generated large profits for commercial interests in favor of institutional change in countries that met two crucial preconditions: easy access to the Atlantic and non-absolutist initial institutions. (England and the Netherlands were the two examples of such countries.) Here Atlantic trade provided substantial profits and altered the balance of political power by enriching and strengthening commercial interests outside the royal circle, which enabled them to demand and obtain the institutional changes necessary for economic growth. This group could then demand and obtain significant institutional reforms protecting their property rights. These merchants also received strong support from Whigs who sought to constrain the king (see also section 2.1.2).[50]

Opposing the Atlantic trade argument to a certain extent, Greasle and Oxley (1998), utilizing two types of robust cointegration-based causality tests, argue that domestic forces, notably technological progress, shaped the industrial revolution, whereas overseas trade expansion was mainly a consequence of industrial growth. They investigate Granger-causality between industrial production, and population, real wages, overseas trade, and technological activity for Britain during the period 1780-1851. Basically, they find that the origins of the BRI seem to lie within the domestic market: what was distinctive about the British marketplace in the period 1780 to 1851 was a conjunction of critical real wage, population, and technological creativity levels. To the extent that the first industrial revolution offers a template, exports appear not to provide a simple pathway to industrialization.

Another weak point of Acemoglu et al. (2005b) is emphasized by Wagener (2009:312) who says that "the restriction to Atlantic long distance trade leaves unmentioned the much earlier, also sea-bound development of Italy with Genoa, Venice, and Pisa and of Flanders with Brugge, Antwerp, and Gent and the rise of the Hanse league of towns". Also unmentioned is the catching-up of the West European interior regions that was only delayed by the higher transportation costs. So probably access to the Atlantic alone is not enough to explain the British development.

Markets are also at the heart of an explanation for the BIR in Allen (2001, 2009), but for completely different reasons to the above. Allen argues that the success of Britain originated from markets, or to be more precise, from commercial gains. As he emphasizes, the success of markets created a structure of wages and prices that differentiated Britain from the Continent. More specifically, in Britain wages were high, and energy was cheap, which were the fundamental reasons for technological breakthroughs in the 18th century.

The underlying assumption of Allen is that technology was invented by people to make money, and inventions were investments where future profits had to offset current cost. So

[50] Ferreira, Pessôa, and Santos (2010) show in their model that without trade one cannot fully explain the Industrial Revolution, although their model is not about providing an explanation for why the Industrial Revolution happened in Britain. The reason for this in their two-sector model is very simple: without international commerce England would not be able to shift resources to the production of manufacturing goods at the rate one observes in the data.

inventors in Britain were led to invent machines that substituted energy and capital for labor. The market was important for that because the balance between the profits and costs of an invention depended on the size of its market. Briefly, British inventions were biased: they were labor saving and capital using. Accordingly, cost reductions were greater in Britain than in the Continent, so the new technologies were adopted in Britain and not in the Continent.

So, in Allen's (2009) framework invention is considered an economic activity, the character and pace of which depended on factors that affected profits and prices. The conclusion is that the famous inventions of the industrial revolution were made in Britain because they were profitable only in Britain (under British conditions). In his account, the favorable legal framework and culture were also attributable to commercial roots.

The market is a key institution in Zanden (2008, 2009) as well, but he considers another aspect of the market critical, namely the extent of market integration. While he does not focus on England, but on Western Europe, when analyzing efficient institutions which developed in the Middle Ages[51], his results shed some light on the British case, too. What he emphasizes is the extent of market integration in an economy and the depth and breadth of factor markets, while bearing in mind the assumption that efficient institutions reduce transaction costs, and therefore lead to high levels of market integration and dense markets. As direct measures of market integration he uses the variability of (annual) prices and the convergence of prices. The variability of prices reflects the extent to which markets are able to cushion shocks via trade: generally, there is low variability in market systems with low transaction costs and high volumes of trade. He finds empirical evidence for Europe's advantage in terms of both market integration and the prevalence of dense factor markets. Since England was not the only case for such evidence, Zanden (2008, 2009) cannot explain the Little Divergence.

More recently, Desmet and Parente (2009) by bearing in mind different mechanisms also pointed to the role of the markets in inducing the Industrial Revolution. While their model is a formal one, and their emphasis is not purely institutional, they clearly attach significance to the market mechanism as such. The novelty of their paper lies in the mechanism by which larger markets bring about the BIR, rather than in the idea that an expansion of markets is critical.

In their theory a gradual expansion of the market, coupled with an increasing variety of consumer goods and growing firm size, sows the seeds for process innovation, which allows the economy to move from Malthusian stagnation to modern growth. They show that their theory is empirically plausible by deriving its quantitative implications in a model calibrated to the historical record of England over the period 1300-2000.

The model works as follows. The subsistence constraint, together with low initial agricultural productivity, implies that the economy starts off with most of its population employed in agriculture. Given that so few people live and work in the city and given the fixed operating cost, only a small number of industrial varieties are produced, implying that goods are not particularly substitutable. Mark-ups are high, and hence, firms are small. As a result, firms do not find it profitable to incur the fixed costs of innovation. However, during this Malthusian phase with stagnant living standards, exogenous increases in agricultural TFP allow for increases in the population and a larger urban base. Eventually, the population reaches a critical size, making industrial firms sufficiently large to warrant process innovation. At this point, firms endogenously lower their marginal costs, and hence, an industrial revolution ensues. While the size of the market depends on a country's total population, it is also affected by transportation costs, internal and external trade barriers, and other institutions.

[51] Zanden joins those scholars who argue that the BIR could be interpreted as the culmination of a process of economic expansion begun in the Middle Ages (see footnote 3).

Basically, Desmet and Parente's (2009) theory on how markets lead to modern growth is a Smithian one in the sense that the extent of the market plays a critical role in inducing an industrial revolution, although the mechanism behind the extension of the market is different from the division of labor.

3. Informal institutions, culture

A conclusion from the above is that the formal institutions were favorable for inventors and entrepreneurs, but secure property rights, the rule of law, the constraints on the executive, and patents in themselves were not sufficient to induce major technological changes. Informal institutions and more particularly, culture played a crucial role. Despite the significant role attributed to norms and ideology in early work on institutions (e.g., North 1981), the economics literature has only recently come to view culture[52] as of similar importance (e.g., Guiso et al. 2006, Tabellini 2008).

The latest research by Mokyr (2010b, 2008) sheds particular light on the overwhelming role of informal institutions, including culture, in which Britain's configuration was unique. According to him, at the level of embededdness[53], "cultural beliefs" created an environment in which inventors and entrepreneurs could operate. This is about recognizing the importance of accepted codes of behavior, patterns of beliefs, trust, etc., that is, informal institutions that channel creativity into productive activities.[54] In fact, according to him, what was unique in Britain was the growth of a set of these social norms beyond the rule of law and explicit penalties for opportunistic behavior. The development of such behavioral rules can be to a large extent attributable to the Enlightenment which made productive activity as such more attractive relative to rent-seeking.[55] Mokyr (2008) argues that in eighteenth century Britain such institutions played a major role in allowing markets to operate and also helped Britain take the technological lead: in Britain, more than anywhere else, informal institutions were becoming more favorably disposed toward technologically innovative entrepreneurship.

It may come as a surprise, but formal law enforcement was a last resort in Britain; markets functioned well because of the above-mentioned informal rules (Mokyr 2008). The key to successful economic exchanges was not necessarily impartial and efficient third-party enforcement, but precisely the existence of a level of trust or other self-enforcing institutions that supported free-market activities. Within a circle of commerce, finance and manufacturing, trust relations and private settlement of disputes prevailed over third party enforcement. Most business was conducted on informal codes and relied on reputation; voluntary compliance, respect for property (private-order institutions) was important in Britain. These norms involved a variety of devices associated with "gentlemanly" behavior.[56]

The idea of being a "gentleman" has acquired a meaning of behavioral codes that signaled that a person was trustworthy. People who felt constrained by the gentlemanly code of behavior behaved honorably, kept their word and did not renege on promises. This behavior made it possible to overcome the kind of free riding and opportunistic behavior that seem to

[52] The notion of culture here is the one now common in the economics literature and consistent with that accepted in cultural anthropology: values or beliefs that are socially transmitted through teaching or imitation, within a pre-defined group of individuals.

[53] See Williamson (2000) to identify the hierarchy of institutions.

[54] The origin of the view that culture, or broadly speaking, informal institutions play an important role in development can be found in Weber's (1930) theory: he traced back the transition to rapid modern economic growth to a transformation of the motivation structure triggered by the Reformation. According to him, the spirit of capitalism follows from the protestant ethic.

[55] Mokyr (2005b, 2006) provides a detailed analysis of the role of the Enlightenment in sustained economic growth.

[56] Note, however, that these norms applied only to the "middle class" that emerged before 1760 and included intelligent and well-educated people.

require coercion by formal state institutions. As shown by Mokyr (2010b) a primary example of the operation of gentlemanly codes was the 18th century credit market in Britain. Credit markets depended on a set of self-enforcing codes framed by norms of gentlemanly conduct. This credit market was primarily enforced by reputational mechanisms; accordingly commercial disputes rarely came before the courts and were often settled through private arbitration.[57]

Briefly, based on Mokyr's analyses, informal rules were even more important than formal rules. What mattered was that within the merchant and artisan classes there existed a level of trust that made it possible to transact with non-kin.[58] Thus it can be argued that such informal institutions led to the emergence of a small, but significant economic elite that carried the BIR.

Having said that, the question of how the middle classes gained ground vis-à-vis aristocrats still remains. In an innovative paper, Doepke and Zilibotti (2007) argue that the rise of a bourgeois elite in industrializing Britain may be regarded as a surprise. Before the transformation got under way, aristocrats had all the odds stacked in their favor – available funds, political connections, access to education. Despite this fact only a few members of the old political elite actually got rich through manufacturing after 1750. Doepke and Zilibotti argue that this is because the middle classes had accumulated a larger stock of "patience capital", that is, a host of cultural practices and norms that make the delay of immediate gratification accepted and expected. Over centuries, the middle class built up both financial capital and valuable cultural traits. As the new technologies of the Industrial Revolution suddenly offered greater returns to patience, the groups best-placed to exploit them were not the elite but the middle classes. Those people who acquired "patience capital" - which was the kind of culture that played a central role in the subsequent development of capitalist industrialism - became key figures in British society.

The above insights are given emphasis in McCloskey (2006) as well, although she focuses on another aspect of culture, which she refers to as 'bourgeois virtues' that developed in the West and are the following: Hope (optimism, entrepreneurship), Faith (identity, integrity, loyalty, honesty), Love (benevolence, friendship, agape), Justice (social balance and honesty), Courage (autonomy, daring, endurance), Temperance (individual balance and restraint, humility), Prudence (know-how, foresight, phronesis). As she carefully explains, all these virtues are beneficial for the development of capitalism. Of course, McCloskey's focus is not only on England, but – bearing in mind the close connection between Weber's protestant ethic and the bourgeois virtues – England was probably a pioneering country in exhibiting these virtues.

The view that culture was crucial in England's development is given empirical evidence in a recent paper by Murrell and Schmidt (2011). They investigate the relationship between culture and formal institutions in 17th-century England. For the institutional variables, they use reports on 17th-century court decisions. Their cultural variables reflect data on word usage in a catalog of publications (books, pamphlets, etc) from the seventeenth century, the *English Short Title Catalogue*. They try to capture the diffusion of a "Whig" political culture, which emphasized the virtues of freedom and the necessity of constraints on the monarchy.

They find a gradual cultural development over the whole time period (1559-1714). They also find that until 1640 the diffusion of Whig culture is limited, but then there is dramatic change with over half of the cultural diffusion completed by 1660 (see Figure 5). The process

[57] Zanden (2009) also argues that low interest rates are the proof of trust in markets. He also shows that the interest rate in Western Europe was low as compared to other regions of the world, which was an essential precondition for the dynamic economic development of Western Europe after the BIR.

[58] What also mattered from this point of view was the fact that the British nation witnessed a blossoming of voluntary organizations (e.g., clubs) that created linkages supporting market activity. This was a kind of social network. For a detailed overview see Mokyr (2010b, 2008).

of cultural change was therefore largely completed in the years before the Bill of Rights of 1689 and the Act of Settlement of 1701, the two major pieces of constitutional legislation.

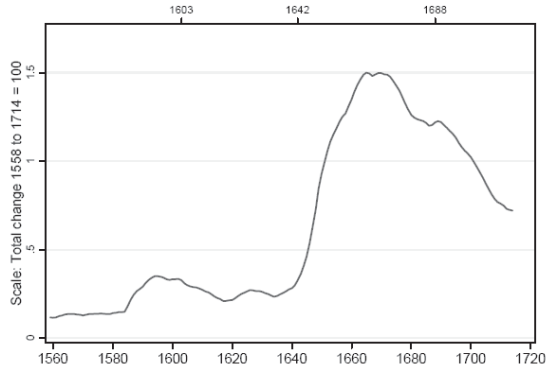


Figure 5: Yearly changes in the importance of "Whig" culture in England, 1558-1714
(source: Murrell and Schmidt 2011:44)

Since Murrel and Schmidt (2011) have yearly data from 1559 to 1714, they apply standard time-series methods to analyze interactions between cultural diffusion and institutional development. They use a vector error correction model, which relates changes in culture and institutions to each other and to deviations of each from their long-run relationships. The results suggest that culture and case-law institutions co-evolve but that statute law is a product of the other two. This co-evolutionary process is shown in Figure 6.

In sum, in Murrell and Schmidt's (2011) empirical analysis culture seems fundamental to the development of formal institutions, spurring direct changes in case law and indirect changes in statute law (in the longer term).

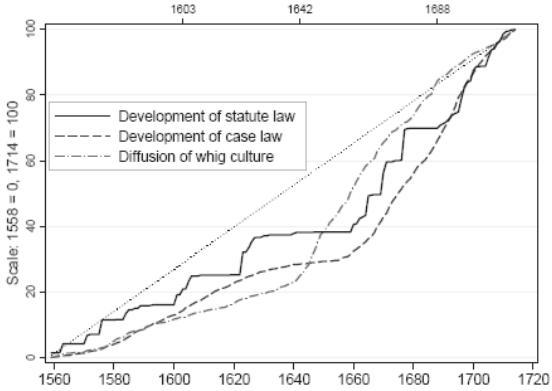


Figure 6: Institutional development and cultural diffusion, England 1559-1714
(source: Murrell and Schmidt 2011:46)

A conclusion of this section is that England's uniqueness in informal institutions was favorable to an industrial revolution; that is, modern growth is almost unanimously accepted and proved by various scholars, although England's uniqueness in industrial revolution-favoring formal institutions is much more doubtful. This may suggest that a perspective which takes into account the co-evolution of formal and informal institutions would be more fruitful is answering the "why in England" question.

4. An umbrella view: England's shift from limited access order to open access order

The above overview of the literature on the institutional causes of the BIR has summarized its major positions, and clearly shows that each focuses only on one particular institution. The view that the ruler was efficiently constrained in England, signifying the protection of private property from the predatory tendencies of a monarch is commonly held in the literature. As discussed above, North and Weingast (1989) argued that this was due to the Glorious Revolution, more precisely to the Parliament after 1688, while others (e.g., Mokyr and Voth 2010, Murrell 2009, Clark 1996) doubt the significance of 1688 and argue that the constraints on monarchs were real enough before 1688.

The culture view can also be weakened by those who seek to make culture endogenous to economic institutions (e.g., Aoki 2007). According to them, beliefs and values are simply elements of institutions; there is no analytical distinction between formal legal rules, informal social customs, and inward beliefs and values. These perspectives, which derive from a certain game-theoretical approach, define institutions as endogenous and self-enforcing.

Without recalling all the theories mentioned above, one can argue that the above-discussed theories centering on a given institution in the explanation for such a complex problem as the "why in Britain" question can be questioned on at least two grounds. One is whether England was really unique in the particular institution emphasized by the theory. Bearing in mind for instance the Atlantic trade argument, it is clear that other countries also opened markets towards new territories, not only England. As for patent law, as mentioned above, the Continent also had a similar patent law to England. The other ground on which one can question certain theories is whether the particular institution in which England was unique really did cause the industrial revolution. Here let me recall once again the debate on the importance of Parliament after the Glorious Revolution: many researchers have provided evidence that the commitment of the government was not enough to induce the industrial revolution.

To arrive at a more accurate institutional account of the BIR one has to take into consideration certain requirements. In my view, there are at least three requirements that any theory of institutions must meet. The first is the recognition that any institution does many things and that it is doubtful that we can really separate one function of an institution from the others (see Ogilvie 2007). The second is the recognition of the importance of the self-sustaining character of the institutional framework. Thirdly, we have to admit that there are inherent complementarities between certain institutions, and that a whole cluster of institutions may be mutually reinforcing, so we cannot study one institution in isolation. As explained by Williamson (2000) various institutions are related to and depend on each other, where the direction and the concrete form of the dependence are determined by a hierarchy of institutions. For our concerns here, it means that the norms, beliefs and culture, that is, informal institutions at the level of embededdness, pose a constraint on the political and legal (formal) institutions at the level below, but some feedback mechanisms operate as well, allowing in this way a kind of co-evolutionary process to take place in institutional changes (see also Hayek 1960).

I believe that the theory of social order developed by North et al. (2009), emphasizing the co-evolution of institutions, offers a new and convincing framework for understanding how institutional changes led to an industrial revolution, and accordingly sustained growth in England. The pre-industrial-revolution economy in particular was characterized by "limited-access" institutions that coercively limited economic entry to valuable resources and organizations in order to create rents for the powerful elites, while excluding the vast mass of economic agents.

In the light of the theory of North et al. (2009) what really happened in Britain in the 18th century – and this was precisely the uniqueness of Britain – was the transition from a limited access order (called also a natural state) to an open access order that relies on competition and open access to form organizations both in the political and economic systems. England's transition in the 18th century from a limited access order to an open access order represented a fundamental change in the broad institutional setting, embodying a particular mix of political, legal and economic institutions. This means that only a particular constellation of various institutions (social order in terms of North et al.) was appropriate for an industrial revolution, and here England was unique as compared to the Continental countries. Put differently, the institutional uniqueness of Britain consisted in the emergence of open access social order institutions, which, in turn, led to modern growth.

Thus, the major question is how and why a transition to the open access order was initiated in the limited access order-England. In this respect, North at al. (2009) argue that the transition was (and had to be) consistent with the logic of the natural state. So, basically the crucial question is why elites transformed their universal privileges into impersonal rights shared equally among elites?

To answer this question, one has to understand the process that evolved in England over several centuries in which the rule of law, and particularly its major characteristic, *impersonality*, solidified by numerous "good" institutions became the standard. This evolutionary process is brilliantly described, among others, by Hayek (1960) and Glaeser and Shleifer (2002).

The rule of law requires laws that apply equally to all citizens and judicial systems that apply the laws impartially (Hayek 1960). *Impersonality*, i.e., treating everyone the same without regard to their individual identity, is the underlying institution in this (Wallis 2011). Wallis (2011) argues that impersonal relationships occur when two individuals interact in a way that does not depend on their personal identity, irrespective of whether they are personally known to each other or not. He clearly separates it from an *anonymous* relationship. Anonymous exchange refers to situations where people who are not personally known to each other interact, although the actors know the social identity (or the group, organization, tribe, city, etc.) of the other in the relationship. Anonymous exchange explicitly does not require people be treated the same, as the actors are constrained by kinship ties, etc.

As explained by Wallis (2011) in detail, natural states create organizations that make anonymous relationships sustainable on a large scale, but impersonality exclusively underpins open access societies.

The evolution of impersonality in England is very much related to the evolution of land ownership, as is shown in North et al. (2009). Land ownership was special in England, and it played a major role in the transition to open access society. English land law is one example of how the rule of law for elites may develop.

In English land law, the freeholder was an impersonal category based on land tenure that granted all freeholders the right to use the king's court and the right to vote (if they had enough land). Once all elite landowners possessed the same rights (inheritance rights and the right to devise by will), the elite had the interest to protect these rights. By the end of the 16th century ownership rights in land were relatively secure and impersonal in England, and by the end of the 17th century organizations associated with land and landownership had been moved outside the immediate control of the state. So, the evolution of land ownership clearly points to how elites found it in their interest to support and obey the rules impartially.

A second factor that contributed to the rise of impersonal rules was free entry to form organizations. England began chartering joint-stocks companies in the mid 16th century, the largest of which engaged in overseas trade and colonization: The Russia Company (1553), The Virginia Company (1606), The East India Company, The Africa Company, The Massachusetts Bay Company, The Hudson Bay Company. These were all natural state

creations, i.e., organizations controlled by the elites. However, in parallel with the rise of commercial and trading interests many of the new rules sought by the commercial constituency were in the form of greater rights and impersonality. In this way, elites found it in their interest to support free entry because they always had greater fear of each other than of the rest of the population. That is, at a certain point elites found themselves in a position of supporting free trade and entrepreneurship. This process, including mutually reinforcing incremental institutional changes[59], prepared England to be the first to meet the doorstep conditions[60] that created the possibility for impersonal elite relationships, and accordingly transformed England from a limited access order into an open access order.

To sum up, what was special and unique in England was an evolutionary process which, for the first time in history, transformed a limited access to an open access society. This complex institutional change led, in turn, to the BIR which was feasible only in the institutional context present at that time in England. In this sense no single institution was responsible for the BIR, but rather the integrity of formal and informal institutions. The evolutionary view that modern growth in England is due to processes started earlier than 1688 and decades before the Glorious Revolution is given empirical evidence as well, for instance in Kishtainy (2011). All this suggests that the BIR, i.e., the emergence of modern economic growth in England may have had deep roots.

[59] Open access to organizations transformed the nature of political and economic competition.
[60] These are as follows: (1) the rule of law for elites, (2) perpetually lived organizations in the public and private spheres, and (3) consolidated control of the military.

References

Acemoglu, D. – Johnson, S. – Robinson, J. (2005a) Institutions as the Fundamental Cause of Long-Run Growth. In: Aghion, Ph. – Durlauf, S. N. (eds), Handbook of Economic Growth, Volume 1A. Amsterdam: North Holland, pp. 385-472.

Acemoglu, D. – Johnson, S. – Robinson, J. (2005b) The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth. American Economic Review 95:546-79.

Acemoglu, D. – Robinson, J. (2000) Political Losers as a Barriers to Economic Development. American Economic Review 90. 2:126-130.

Allen, R. C. (1983) Collective Invention. Journal of Economic Behavior and Organization 4:1-24.

Allen, R. C. (2001) The Great Divergence in European Wages and Prices from the Middle Ages to the First World War. Explorations in Economic History 38:411-447.

Allen, R. C. (2009) The British Industrial Revolution in Global Perspective. Cambridge: Cambridge University Press.

Antras, P. – Voth, H.-J. (2003) Factor Prices and Productivity Growth during the British Industrial Revolution. Explorations in Economic History 40. 1: 52-77.

Aoki, M. (2007) Endogenizing Institutions and Institutional Change. Journal of Institutional Economics 3. 1:1-31.

Bogart, D. – Richardson, G. (2011) Property Rights and Parliament in Industrializing Britain. Journal of Law and Economics 54. 2:241-274

Clark, G. (1996) The Political Foundations of Modern Economic Growth: England, 1540-1800. Journal of Interdisciplinary History 26. 4:563-588.

Clark, G. (2003) The Great Escape: The Industrial Revolution in Theory and in History. Unpublished manuscript. UC Davis. Available: http://www.econ.ucdavis.edu/faculty/gclark/papers/ir2003.pdf

Crafts, N. F. R. (1977) Industrial Revolution in England and France: Some Thoughts on the Question 'Why Was England First'? Economic History Review 30:429-441.

Crafts, N. F. R. (1985) British Economic Growth during the Industrial Revolution. Oxford: Oxford University Press.

Crafts, N. F. R. – Harley, C. K. (1992) Output Growth and the Industrial Revolution: A Resatatement of the Crafts-Harley View. Economic History Review 45. 4:703-730.

De Long, J. B. – Schleifer, A. (1993) Princes and Merchants: European City Growth Before the Industrial Revolution. Journal of Law and Economics 36:671-702.

Desmet, K. – Parente, S. L. (2009) The Evolution of Markets and the Revolution of Industry: A Quantitative Model of England's Development, 1300-2000. CEPR Discussion Papers No. 7290.

Doepke, M. – Zilibotti, F. (2007) Occupational Choice and the Spirit of Capitalism. NBER Working Paper. No. 12917.

Dutton, H. I. (1984) The Patent System and Inventive Activity During the Industrial Revolution, 1750-1852. Manchester: Manchester University Press.

Epstein, S. R. (2000) Freedom and Growth: The Rise of States and Markets in Europe, 1300-1750, London: Routledge.

Feinstein, C. H. (1981) Capital Accumulation and the Industrial Revolution. In: Floud R. – McCloskey D. N. (eds.), The Economic History of Britain since 1700, vol. 1. Cambridge. pp. 128-142.

Ferreira, P. C. – Pessôa, S. – Santos, M. R. (2010) Globalization and the Industrial Revolution. Economics Working Papers (Ensaios Economicos da EPGE) No. 708. Available: http://ideas.repec.org/p/fgv/epgewp/708.html

Freeman, C. – Louça, F. (2001) As Time Goes By. From the Industrial Revolutions to the Information Revolution. Oxford: Oxford University Press.

Galor, O. – Weil, D. (2000) Population, Technology, and Growth. American Economic Review 90. 4:806-828.
Glaeser, E. L. – Shleifer, A. (2002) Legal Origin. Quarterly Journal of Economics 117. 4:1193-1230.
Greasley, D. – Oxley, L. (1997) Causality and the First Industrial Revolution. Industrial and Corporate Change 7:33–47.
Greasley, D. – Oxley, L. (1998) Vector Autoregression, Cointegration and Causality: Testing for Causes of the British Industrial Revolution. Applied Economics 30:1387–1397.
Greasley, D. – Oxley, L. (2007) Patenting, Intellectual Property Rights and Sectoral Outputs in Industrial Revolution Britain, 1780–1851. Journal of Econometrics 139:340–354
Guiso, L. – Sapienza, P. – Zingales, L. (2006) Does Culture Affect Economic Outcomes? Journal of Economic Perspectives 20:2:23-48.
Hayek, F. A. (1960) The Constitution of Liberty. Chicago: The University of Chicago Press.
Helpman, E. (1998) General Purpose Technologies and Economic Growth. Cambridge: The MIT Press.
Jones, E. L. (1981) The European Miracle. New York: Cambridge University Press.
Khan, B. Z. – Sokoloff, K. L. (2004) Institutions and Democratic Invention in 19th Century America: Evidence From "Great Inventors" 1790-1930. American Economic Review 94:395-401.
Kishtainy, N. (2011) The Glorious Revolution and Long Run Growth in Britain. Available: http://www2.warwick.ac.uk/study/csde/gsp/eportfolio/directory/pg/live/ecrgag/papers/gbgth_nk_jun.pdf
Klerman, D. – Mahoney, P. (2005) The Value of Judicial Independence: Evidence from Eighteenth Century England. American Law and Economics Review 7. 1:1-27.
Landes, D. S. (1994) What Room for Accident in History?: Explaining Big Changes By Small Events. Economic History Review 47:637-656.
Lipsey, R. – Carlaw, K. I. – Bekar C. T. (2005) Economic Transformations: General Purpose Technologies and Long-term Economic Growth. Oxford: Oxford University Press.
Lucas, R. E. (2002) The Industrial Revolution: Past and Future. In: Lucas, R. E. (ed), Lectures on Economic Growth. Cambridge: Harvard University Press.
MacLeod, C. (1986) The 1690s Patents Boom: Invention or Stock-Jobbing? Economic History Review 39. 4:549-571.
MacLeod, C. (1988) Inventing the Industrial Revolution. The English Patent System, 1660-1800. Cambridge: Cambridge University Press.
McCloskey, D. (2006) The Bourgeois Virtues: Ethics for an Age of Commerce. Chicago: University of Chicago Press.
Mokyr, J. (1990) The Lever of Riches: Technological Creativity and Economic Progress. New York, London: Oxford University Press.
Mokyr, J. (1999) Invention and Rebellion: Why Do Innovations Occur at All? An Evolutionary Approach. In: Brezis, E. – Temin, P. (eds), Minorities and Economic Growth. Amsterdam: Elsevier Publishers. pp. 179-203.
Mokyr, J. (2002a) The Gifts of Athena. Historical Origins of the Knowledge Economy. Princeton University Press. First paperback, 2005.
Mokyr, J. (2002b) Thinking about Technology and Institutions. Paper for the Macalester College International Roundtable 'Prometheus's Bequest: Technology and Change' October 10-12, 2002.
Mokyr, J. (2005a) Long-term Economic Growth and the History of Technology. In: Aghion, Ph. – Durlauf, S. N. (eds), Handbook of Economic Growth. Amsterdam: North Holland. pp. 1113-1180.
Mokyr, J. (2005b) The Intellectual Origins of Modern Economic Growth. Journal of Economic History 65. 2:285-351.

Mokyr, J. (2006) The Great Synergy: the European Enlightenment as a Factor in Modern Economic Growth. In: Dolfsma, W. – Soete, L. (eds), Understanding the Dynamics of a Knowledge Economy. Cheltenham: Edward Elgar. pp. 7-41.
Mokyr, J. (2008) The Institutional Origins of the Industrial Revolution. In: Helpman, E. (ed), Institutions and Economic Performance. Cambridge: Harvard University Press. pp. 64-119.
Mokyr, J. (2010a) The Enlightened Economy. London and New Haven: Yale University Press.
Mokyr, J. (2010b) Entrepreneurship and the Industrial Revolution in Britain. In: Baumol, W. – Landes, D. S. – Mokyr, J. (eds), Entrepreneurship in History. Princeton: Princeton University Press. pp. 183-210.
Mokyr, J. – Nye, J. (2007) Distributional Coalitions, the Industrial Revolution, and the Origins of Economic Growth in Britain. Southern Economic Journal 74. 1:50-70.
Mokyr, J. – Voth. H.-J. (2010) Understanding Growth in Europe, 1700-1870: Theory and Evidence. In: Broadberry, S. – O'Rourke, K. H. (eds), The Cambridge Economic History of Modern Europe Volume 1:1700-1870. Cambridge: Cambridge University Press. pp. 7-42.
Moser, P. (2007) Why Dont't Inventors Patent? Searle Center on Law, Regulation, and Economic Growth. Northwestern University. Available: https://anthonydamato.law.northwestern.edu/searlecenter/papers/Moser_Inventors.pdf
Murrell, P. (2009) Design and Evolution in Institutional Development: The Insignificance of the English Bill of Rights. Available: http://ssrn.com/abstract=1522864
Murrell, P. – Schmidt, M. (2011) The Coevolution of Culture and Institutions in Seventeenth Century England. Available: http://ssrn.com/abstract=1880957
North, D. C. (1981) Structure and Change in Economic History. New York: Norton.
North, D. C. – Thomas, R. P. (1973) The Rise of the Western World. Cambridge: Cambridge University Press.
North, D. C. – Wallis, J. J. – Weingast, B. R. (2009) Violence and Social Orders: A Conceptual Framework for Understanding Recorded Human History. Cambridge: Cambridge University Press.
North, D. C. – Weingast, B. R. (1989) Constitutions and Commitment: Evolution of Institutions Governing Public Choice in Seventeenth-Century England. Journal of Economic History XLIX. 4:803-32.
Nuvolari, A. (2004) Collective Invention during the British Industrial Revolution: the Case of the Cornish Pumping Engine. Cambridge Journal of Economics 28:347-363.
Ogilvie, S. (2007) Whatever Is, Is Right? Economic Institutions in Preindustrial Europe. Economic History Review 60. 4:649.
Olson, M. (1982) The Rise and Decline of Nations. New Haven: Yale University Press.
Olson, M. (1991) Autocracy, Democracy, and Prosperity. In: Zeckhauser, R. (ed), Strategy and Choice. New York: MIT Press.
Oxley, L. – Greasley, D. (1997) Vector Autoregression, Cointegration and Causality: Testing for Causes of the British Industrial Revolution. Applied Economics 30:1387-1397.
Pincus, S. – Robinson, J. (2011) What Really Happened During the Glorious Revolution? NBER Working Paper No. 17206.
Quinn, S. (2001) The Glorious Revolution's Effect on British Private Finance: A Microhistory, 1680-1705. Journal of Economic History 61. 3:593-615.
Smith, A. (1776) An Inquiry into the Nature And Causes of the Wealth of Nations. London: The Adam Smith Institute. Electronic Edition, 2001. Available: http://www.adamsmith.org/smith/won-index.htm.
Sullivan, R. J. (1989) England's "Age of Invention": The Acceleration of Patents and Patentable Invention during the Industrial Revolution. Explorations in Economic History 26:424-452.

Sussman, N. – Yafeh, Y. (2006) Institutional Reforms, Financial Development and Sovereign Debt: Britain 1690 – 1790. Journal of Economic History 66. 5:1-30.
Tabellini, G. (2008) Presidential Address: Institutions and Culture. Journal of the European Economic Association 6. 2-3:255-294.
Voightländer, N. – Voth, H.-J. (2006) Why England? Demographic Factors, Structural Change and Physical Capital Accumulation during the Industrial Revolution. Journal of Economic Growth 11. 4:319-361.
Wagener, H.-J. (2009) The European Journal of Comparative Economics 6. 2:287-323.
Wallis, J. J. (2011) Institutions, Organizations, Impersonality, and Interests: The Dynamics of Institutions. Journal of Economic Behavior and Organization 79. 1-2:48-64.
Weber, M. (1930) The Protestant Ethic and the Spirit of Capitalism. London, New York: Routledge. 2001.
Williamson, O. E. (2000) The New Institutional Economics: Taking Stock, Looking Ahead. Journal of Economic Literature XXXVIII. Sep:595-613.
Zanden, J. L. van (2008) The Road to the Industrial Revolution: Hypotheses and Conjectures about the Medieval Origins of the 'European Miracle'. Journal of Global History 3:337-359.
Zanden, J. L. van (2009) The Long Road to the Industrial Revolution. Leiden, Boston: Brill.

Pál Czeglédi

Larger Prey, More Predators: Culture as a Constraint on Expropriation*

"Though admittedly, ideology is something which cannot be 'proved' (or demonstrated to be true), it may well be something whose widespread acceptance is the indispensible condition for most of the particular things we strive for."
(Hayek 1971:32)

"The ultimate guarantee for individual liberty may rest not on rules for social choice but on developing individual values that respect each other's personal choices."
(Sen 1970:155-156)

1. Does culture matter for economic growth?

Institutional economics has become a part of the mainstream research into economic growth, and the view that "institutions matter" has been widely accepted, thanks to works that explore the way different institutions lead to differences in long run economic performance (among the many see North 1990, 2005, Acemoglu 2009, or Ménard and Shirley 2005). As a result, the claim that formal institutions are important factors that shape the economic performance of countries is not very much debated. They are said to be even more fundamental causes of development than technological change or the accumulation of resources (Acemoglu et al. 2005, Acemoglu 2009:109-143, Owen et al. 2009). However, once it becomes clear that formal institutions are deeper and more important factors than capital accumulation, the question that immediately arises is why formal institutions differ to such a great extent. One answer is politics (Acemoglu et al. 2005), another is culture. This paper will try to contribute to this latter line of research.

Culture is, according to Guiso et al. (2006:23), the set of "those customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation". That is, the important features of culture are that, first, they are institutions that are not enforced by a third party (the state), and that secondly, they evolve slowly; several generations are needed for values to change. The empirical literature on economic growth deals with culture accordingly: it shows either that something which is transmitted over generations, probably by inheritance, has an impact on economic performance, or that those values that people *say* they have are significant determinants of development.

The first approach suggests using genetics and information on inheritance in the explanation of different income levels or growth rates. This is what has been done by Spolaore and Wacziarg (2009) for instance, showing that "relative genetic distance" is a significant and important determinant of income differences. These kinds of explanations give indirect evidence that culture affects economic development in that culture is the one characteristic that is transmitted through generations. In Spolaore and Wacziarg's (2009) interpretation relative genetic distance parallels differences in "habits and customs" and the difference in the latter imposes a cost on technology adoption. Using genetics to support a cultural view of economic development is not uncommon in growth econometrics. Algan and Cahuc (2010), for example, use data of second generation Americans to instrument the level of trust in the country from which these people's ancestors arrive in the United States, and show the significance of these factors.

Another branch of this literature (see the next section) uses survey results on values people hold as explanatory factors in cross-country regressions (Tabellini 2008, 2010, Licht et

* This paper was also supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

al. 2007) and come to the similar conclusion: differences in culture account for a significant part of the differences in development across countries. In sum the view that "culture matters" quickly joined the "institutions matter" argument and has become a part of the literature on growth, including econometrics.

Many researchers are still skeptical though. Their argument, in short, is, as Sala-i-Martin puts it in an informal way (Snowdon 2006:105),"you cannot explain something that changes rapidly with factors that do not change at all, or change only very slowly… every time predictions are made on the basis of culture or religion they turn out to be wrong. We keep observing countries where all of a sudden income starts to grow even though culture and religion have stayed the same". That is the same "challenge" which Acemoglu (2009:122-123) sees as the most important difficulty for a theory trying to explain economic growth with culture. As he (ibid: 123) argues, "if some Asian cultural values are responsible for the successful growth experiences of these [South Asian – the author] countries, it becomes difficult to explain why these Asian values did not lead to growth before". But this is not a real puzzle in itself. Culture may be seen as one factor of the many that are needed for economic development, so culture may not be necessary for fast growth if the other factors are in place. The puzzle becomes more serious, however, once one accepts the view that formal institutions need informal ones to work properly.

A consensus view of economic growth is that the most important fundamental factors (Acemoglu and Johnson 2005, Rodrik et al. 2004) are institutions most often understood as formal rules, and "getting the institutions right" is the most important task if you want to achieve sustainable development. Supposing, as many do (see the next section), that institutions can only be "gotten right" when these institutions are supported by the values held by the players, culture becomes a fundamental, if not necessary, factor behind economic growth. In short, while theoretically and in a historical perspective, culture seems to be a fundamental factor of economic growth, fast-growing countries are very diverse culturally. Seeing development through these lenses, the cultural variety among those that have begun to catch-up is puzzling. This puzzling contrast between formal and informal institutions is completed by the contrast between the arguments of economic historians explaining the first industrial revolution, and that of economists explaining contemporary facts of economic growth. Institutional economists examining post war growth records place a high emphasis on formal institutions while historians analyzing the industrial revolution see informal institutions (that is, culture) as the most important cause of economic growth.[61] The argument of economic historians and that of growth economists are just the opposite of each other. While economic historians say that we know that property rights security did not change, but informal institutions did[62], economists say that we know that informal institutions do not change, but formal ones do.

In addition, whether formal or informal, the role of institutions are deemed to be the "channelling" of talents towards productive activities as Baumol (1990) famously explained and as many others argue (see the next section). In this light the estimation of Comin and Hobijn (2010a:2048-2049) according to which technology specific effects are responsible for 65 percent of the variation in the diffusion of technology is also controversial, since culture is country-specific, not technology specific.

One can summarize the problem of culture versus formal institutions in the following way. On the one hand, culture is shown to be an important factor of economic growth but contemporary growth successes do not seem to result from a cultural change. On the other

[61] This contrast is very well illustrated by McCloskey's (2010:320-322) short comment on the explanation of economists for the start of economic growth in the industrial revolution. Although this explanation is generally accepted by institutional economists, McCloskey agrees with hardly any of it.

[62] See the next sections for more details.

hand, economists emphasize the role of formal institutions, first of all property rights security, which however did not improve around the time the era of economic growth began. This paper tries to contribute to the understanding of this problem by examining the role of culture. In the next section I will first describe the institutional stickiness argument (section 2) as an explanation of why property rights may need a philosophical (cultural) background. Next, in section 3 I will hypothesize that the part of culture that matters for economic development is what can be called the ideology of freedom which implies a negative attitude towards rent seeking. Interpreting culture in this way lets me explain the effect of institutions in a simple model of technology diffusion (section 4), the implications of which may shed some light on the way the puzzle above can be solved. Finally in section 5 I will test the explanation by investigating the data on the spread of technologies across countries.

2. Culture, formal institutions, and innovation

2.1 Values matter, but which ones, and why?

In the section above I reviewed some empirical results showing that something that is transmitted over generations is an important factor in growth. Clearly these results do not say anything about the sorts of culture that are important.[63] However, it is difficult to say anything about the mechanics of this effect if we cannot say anything about the content of these informal rules.

Informal rules are rules that are followed for other reasons than third party enforcement. As Stringham (2011) explains, markets cannot work if transactions are only based on contracts that are enforced by a third party, and on contracts that are self enforcing on utilitarian grounds. There must be a third mechanism, the enforcement of which comes from the internal moral constrains of the individual. These moral constraints on behavior are to be interpreted as informal institutions. More precisely, Stringham (2011) differentiates between moral constraints and social norms, arguing that the latter – together with, for example, religion – can be the source of the internal moral constraints that have a great impact on whether a market economy works properly.

Those authors who examine the relatively recent experience of economic growth come to certain conclusions concerning which values matter. Tabellini (2010) shows that culture can be made responsible to account for differences in the development of different regions within the same country. The values he shows that matter are very similar to what is emphasized by economic historians (see in section 3): individualism and respect for others. In a similar fashion, but using an alternative psychologically based measure of culture, Licht et al. (2007) came to very similar conclusions. They find that the most important trait is what they call autonomy, referring to the fact that the individual thinks of herself as autonomous of the community in which she lives.

In sum, a "good" culture from an economic point of view is individualistic in the sense that it defines the pursuing of privately defined aims as a right thing to do. This shared belief includes two implications which have to be emphasized here. First, it does not make individualism and utility maximization equal, seeing utility as a narrow concept. Secondly, it also implies that there should be values that constrain individual actions.

These extra-utilitarian values are the ones that are deemed to be important by Mokyr (2008, 2006, 2010) and McCloskey (2006, 2008, 2010). In their understanding, appropriate

[63] As Spolaore and Wacziarg (2009:471) admits, "[a]lthough we provide a general economic interpretation of genetic distance in terms of barriers to the diffusion of development from the frontier, we remain largely agnostic about specific mechanisms of technology diffusion, as well as about the specific traits and characteristics that create the barriers"

formal institutions are insufficient to enhance innovation and economic growth. Formal procedures in 18th century England were no better than before according to McCloskey (2008, 2010:317-324) or Clark (2007:145-165), not to mention the fact that people did not really turn to the state as a third a party to enforce contracts (Mokyr 2008a,b). What changed, they argue was not the (formal) constraints on the government, or on those with the capability of expropriating wealth, rather it was the norms that people followed in their everyday lives that changed around the time of the industrial revolution.

Similarly Murrell (2009) argues that it was not the change in formal institutions such as the Bill of Rights that was important. According to his evidence, the constitutional changes around 1688 has no effect on important variables associated with development, and "[n]othing in the Bill and the Act [the Bill of Rights and the Act of Settlement – the author] added anything to strengthen the rights of English citizens and several clauses diminished those rights" (ibid:28). The evidence he provides supports the view that "changes in political culture and lower-level institutions came before – and were more important than – constitutional change, a message highly complementary with that of the emerging economic literature on culture" (ibid:4). Zanden (2009:28) seems to express a somewhat consensus view saying that "from the late medieval period Western Europe already had a relatively efficient set of institutions ... property rights were well respected, and ... a relatively high level of trust was common".

That factors of this kind could not only be the engine behind industrial revolution-era Britain but also behind the economic growth experienced in more modern times is shown for example by Zak and Knack (2001). They model trust by supposing an investigation technology by which people can be more or less diligent about their agents' behavior on the market. But diligence consumes resources on the one hand, and on the other the same reason that makes more diligence pay off lowers expected returns on investment. Consequently, argue Zak and Knack (2001), there is the possibility of a "Northian poverty trap" which means that if the society is too heterogeneous, people will devote a substantial part of their labor to investigating other people's behavior and the return on investment will not be high enough to start and sustain economic growth. What in Zak and Knack's (2001) model is the "distance between the investor and the broker" is the shared "codes of behavior" for Mokyr (2010): people are "similar" if they both share a certain value system and both of them think that the others share it, too.

It is individualism as opposed to collectivism that is emphasized by Gorodnichenko and Roland (2010). They set up a growth model in which individualism has two roles. First, it makes the combination of the factors of production more difficult. Secondly, and more importantly, it increases the value individuals devote to personal achievement as such, which in their model means the product of innovation. Their results receive empirical support in Gorodnichenko and Roland (2011).

In addition to those explanations that emphasize the direct role of different cultural traits in innovation, there is a more fundamental one. It is that the formal and informal institutions must be congruent so that formal institutions can serve as efficient constraints on behavior. One of the most concise explanations of this sort comes from Boettke (2001) who summarizes the argument in three brief propositions (ibid: 259):

1 People respond rationally to incentives.
2 Incentives are a function of the rules of the game.
3 Rules are only RULES if customary practice dictates.

That is, formal rules will be stuck to only if they are in line with the informal rules in such a way that the formal rules do not provide incentives to take actions which are in contradiction with what is culturally accepted. Once one accepts this "institutional stickiness" argument,

together with the one saying that what is needed for economic development is freedom of entrepreneurial action and security of private property (Acemoglu and Johnson 2005, Baumol 1990, Parente and Prescott 2000), culture becomes a more fundamental factor of economic development than formal institutions.

The theory of institutional stickiness is even supported by some of the papers mentioned above. Tabellini (2008) and Licht et al. (2007) show that formal institutions are of better quality in certain cultural circumstances. In particular Tabellini (2008) shows that a higher level of generalized trust predicts the quality of government measured by bureaucratic quality and by the anti-diversion policies of governments. It is also shown that the effect of substitutability may be at work in some areas: when it is international trade that is to be explained, the explanatory power of formal institutions are better in that country group in which formal institutions are weaker. Licht et al. (2007) argue similarly by showing that several dimensions of good governance are predicted by the different psychological features they identify. Most importantly, they show that the cultural dimension of autonomy is the most important one[64]. Using the same psychological features in his explanation, Licht (2008) argues that before choosing whether to follow norms or not, the society should decide which rules to follow, because "an individual imposing on herself a constraint that is not socially valued would be acting foolishly, not just irrationally" (ibid:725). Thus the extent to which people(s) abide by the law will depend on values or "cultural orientation". The formalistic view of the rule of law, he shows, is very much in line with the attitude that, following Schwartz (2004), he calls personal autonomy. Facchini (2010), somewhat similarly to Licht (2008), also argues that the idea of property needs a philosophical view of the self as independent of the community. In sum, the rule of law needs a value system according to which "legal entitlements will be respected in most circumstances, *irrespective of the context*" (Licht 2008:738, italics are mine)

As for the how and why questions, the answer seems to be found in higher productivity or technological change. A general claim in this respect is that an economy with a better culture has more "dynamism" (Phelps 2006, Phelps and Zoega 2009). A dynamic economy is an entrepreneurial economy according to Phelps and Zoega (2009), which means that creativity translates into business. Although the main point in Comin, Easterly and Gong (2010) is that technological adoption is extremely path dependent, they also provide some reasons to think that culture may be a driving force behind this. They show, for example, that when it is not countries, but peoples that are investigated, their results – according to which development and technology levels are strongly correlated with the technology level in 1500 AD – become stronger. On the other hand, they also show that within-country (between-sector) variance of adoption levels are relatively large compared with cross-country variances (ibid: 91-92) .

2.2. Some challenging empirics

The view that is emerging from the review in the previous subsection can be challenged, however, with some relatively simple statistical evidence. I will show here that, first, informal institutions as they are measured are not very much explained by cultural traits as they are measured. Secondly, formal institutions do not explain a very great part of the differences in technology diffusion.

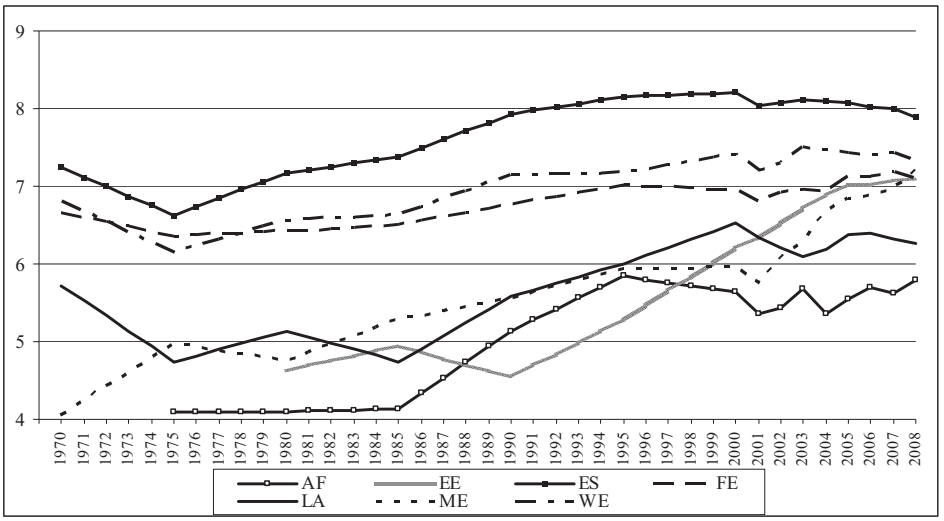
As for the first step in the causational chain of the stickiness argument, let us consider Figure 1 and 2 below. Here formal institutions are proxied by the Freedom of The World Index (Gwartney and Lawson 2010), while informal institutions are represented by cultural

[64] According to Licht et al. (2007:662) a culture described by a high level of autonomy is one "in which the person is viewed as an autonomous, bounded entity who finds meaning in his or her own uniqueness."

regions as identified in Licht et al. (2007) based on seven cultural variables[65]. In figure 1 the overall index of economic freedom is used, while in figure 2 only one of its areas is featured, namely *Legal Structure and the Security of Property Rights*, since the literature (Acemoglu and Johnson 2005) shows that property rights are the most important institutions when it comes to economic growth.

Figure 1

Economic freedom in different cultural regions of the world as described in Licht et al. (2007)



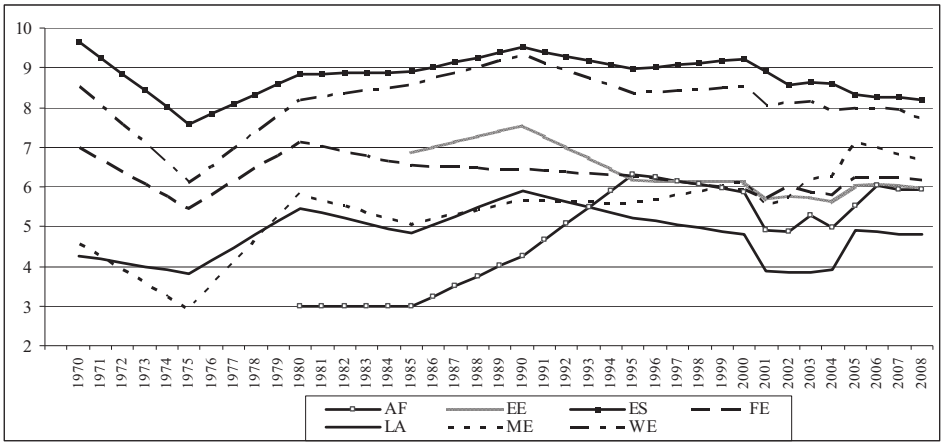
Sources: Gwartney and Lawson (2010) and Licht et al. (2007)
Notes: The values are averages of the scores of those countries that belong to a certain cultural regions.
Abbreviations: AF: African, EE: Eastern European, ES: English-speaking, FE: Far Eastern, LA: Latin American, ME: Mediterranean, WE: Western European

What can be seen here is that there are fundamental changes even within a given cultural region, and these changes do not seem to be culture specific either. Great changes seem to affect formal institutions in the same way across different cultures. To put it another way, the relative position of countries within the range of possible institutions seem to be determined by culture to a greater extent than is their absolute level on the economic freedom scale.

[65] These data are based on a unique social psychological approach, and describe people's attitudes on three dimensions. The first is (Licht et al. 2007:662) *embeddedness-autonomy*, where embeddedness describes the view that sees "the person as embedded in the group and committed to maintaining the status quo, propriety, and restraint of actions or inclinations that might disrupt group solidarity or the traditional order". The second is the *harmony-mastery* dimension which refers to the extent to which people think they should get "ahead through active self assertion". The third, the *hierarchy-egalitarianism* dimension is about the cultural emphasis on obeying role obligations within a legitimately unequal distribution of power, roles, and resources".

Figure 2

"Legal Structure and the Security of Property Rights" in different cultural regions of the world as described in Licht et al. (2007)



Sources: Gwartney and Lawson (2010) and Licht et al. (2007)

Notes: The values are averages of the scores of those countries that belong to a certain cultural regions. Abbreviations: AF: African, EE: Eastern European, ES: English-speaking, FE: Far Eastern, LA: Latin American, ME: Mediterranean, WE: Western European

A simple way to assess this impression more formally is to compare the explanatory power of culture on institutions in the short run, and in the long run. Such a comparison is shown in Table 1 and 2. Clearly the average value of the economic freedom index is much better explained than it is in 2008. However, even in the case when the average level is considered, hardly more than twenty percent of the variation between the different areas is explained by the three cultural variables. The only exception is the area of legal structure and the security of property rights, in which case almost the half of the variance is explained. The area of legal structure is determined to a much greater extent than the other four areas.

A challenge to the view that good culture creates good formal institutions and good institutions provide incentives to productive activities which will lead on to (higher) economic growth comes from an interpretation of the interesting results in Comin and Hobijn (2010a,b, 2008, 2006). They create the CHAT (Cross-country Historical Adoption of Technology) database to analyze the facts of technology diffusion. Estimating their model of technology adoption Comin and Hobijn (2010a) came to several interesting conclusions concerning the adoption lags of different technologies, the estimation of which their model makes possible. One of their conclusions is that the adoption lags are relatively large (45 years on average) and vary, not only across countries but also across technologies.

Comin and Hobijn (2010b) build a model of technology diffusion to argue that such a diffusion was a major cause of the fast growth of European countries after WWII. In addition to this they claim that the data show this diffusion was to a large extent promoted by US foreign assistance, including the Marshall Plan. Data on technology diffusion, they say, discredit other institutional explanations. In addition they seem to falsify Parente and Prescott's (2000) more recent explanations carried out in the tradition of Olson's work which also placed great emphasis on the institutional hurdles to technological diffusion. Comin and Hobijn (2010b) base this argument on the fact that there is no difference between the effect on those technologies that have "close" predecessors and on those that do not. Similarly, they argue that since the old technologies were not affected, but only the new ones, it should not be

some general institutional factor that caused the fast growth, because such a change should affect both kinds of technology.

Table 1

Regressions of different areas of the economic freedom of the world index on cultural variables from Licht et al. (2007)

	Dependent variables: areas of the Economic Freedom of the World index for 2008					
	Summary index	size of government	legal structure	sound money	freedom to trade internationally	regulation
constant	2.477 (4.38)***	-0.0042 (-0.01)	4.924 (6.14)***	0.231 (0.09)	2.498 (4.05)***	3.103 (5.83)***
harmony	0.384 (0.12)	0.374 (1.00)	-0.634 (-1.58)	1.900 (1.06)	-0.091 (-0.27)	-0.377 (-1.65)
embeddedness	-0.404 (-2.32)**	0.907 (2.06)**	-1.488 (-4.42)***	-0.644 (-1.76)***	-0.280 (-1.20)	-0.424 (-1.61)
hierarchy	-0.065 (-0.70)	0.079 (0.33)	-0.234 (-1.30)	-0.047 (-0.21)	-0.055 (-0.43)	-0.103 (-1.00)
R^2	0.105	0.157	0.325	0.111	0.034	0.094
number of observations	49	49	49	49	49	49

Notes: all variables are included in log forms. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level.

Table 2

Regressions of ranks according to different areas of the economic freedom of the world index on cultural variables from Licht et al. (2007)

	Dependent variables: average index scores according to the areas of the Economic Freedom of the World index between 1970 and 2008					
	Summary index	size of government	legal structure	sound money	freedom to trade internationally	regulation
constant	3.434 (5.35)***	0.519 (0.39)	5.600 (6.68)***	3.890 (3.98)***	2.914 (4.10)***	3.948 (4.90)***
harmony	0.388 (-1.14)	0.029 (0.05)	-0.705 (-1.79)*	-0.241 (-0.48)	-0.260 (-0.71)	-0.770 (-1.88)*
embeddedness	-0.807 (-3.52)***	0.456 (0.78)	-1.956 (-6.08)***	-1.262 (-3.75)***	-0.375 (-1.43)	-0.786 (-2.96)***
hierarchy	-0.138 (-1.10)	0.383 (1.33)	-0.335 (-2.28)***	-0.081 (-0.36)	-0.370 (-2.72)***	-0.208 (-1.34)
R^2	0.257	0.144	0.480	0.241	0.228	0.219
number of observations	49	49	49	49	49	49

Notes: all variables are included in log forms. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level.

These facts seem to be puzzling from an institutional perspective. The relatively small role of country effects as opposed to technology specific effects suggest that it is not institutions that retard technological change and thus economic growth. This puzzling fact is affirmed in Tables 3-5, where I used the CHAT data from Comin and Hobijn (2009) to give a simple assessment of their claim with a different (and simpler) method, and of what has just been concluded about institutional stickiness[66]. Table 3a, b show the results of a very simple

[66] The dataset is described in section 5.2.

conditional convergence regression with these data and with income. In addition to data on income from Heston, Sumers and Aten (2009) and the above mentioned CHAT data I use the economic freedom of the world index (Gwartney and Lawson 2010) as a proxy for "good" institutions.

Table 3a

The effect of economic freedom on the diffusion of different technologies

	Dependent variable: growth rate of							
	GDP per capita	ships	railways-pass.	railways-freight	cars	trucks	aviation-pass.	aviation-freight
constant	0.281 (6.34)***	0.476 (4.90)***	1.134 (4.15)***	0.641 (2.39)**	0.697 (3.84)**	0.125 (1.91)*	0.811 (8.12)***	0.419 (2.53)**
initial value	-0.44 (-7.96)***	-0.084 (-5.22)***	-0.132 (-4.06)***	-0.085 (-2.96)**	-0.056 (-4.57)***	-0.045 (-6.25)***	-0.097 (-12.24)***	-0.110 (-7.12)***
economic freedom	0.067 (7.83)***	0.030 (0.69)	-0.045 (-0.96)	-0.06 (-0.10)	0.057 (1.71)*	0.093 (2.80)***	0.010 (0.19)	0.084 (0.96)
no. of countries	114	43	69	64	105	88	85	82
no. of obs.	643	152	215	202	390	346	270	261
R^2 (within)	0.197	0.478	0.356	0.167	0.212	0.171	0.546	0.480
R^2 (between)	0.081	0.039	0.006	0.027	0.006	0.05	0.022	0.000

Table 3b

The effect of economic freedom on the diffusion of different technologies

	Dependent variable: growth rate of						
	telephone	cell phone	pers. computers	internet users	MRIs	Blast ox. steel	electricity
constant	0.101 (0.93)	-0.066 (-0.07)	0.927 (3.02)***	-4.187 (-2.62)***	-1.340 (-1.40)	1.636 (3.37)***	0.982 (6.97)***
initial value	-0.016 (-1.92)*	-0.115 (-3.94)***	-0.087 (-6.91)***	-0.106 (-2.48)***	-0.116 (-4.54)***	-0.238 (-3.52)***	-0.042 (-6.61)***
economic freedom	0.103 (3.16)***	1.140 (1.75)*	0.230 (0.99)	3.410 (3.42)***	0.942 (1.82)*	0.163 (1.61)	0.030 (2.16)**
no. of countries	109	95	85	84	16	48	112
no. of obs.	324	166	143	103	30	208	464
R^2 (within)	0.045	0.342	0.536	0.511	0.412	0.711	0.274
R^2 (between)	0.001	0.006	0.340	0.003	0.001	0.016	0.07

Notes: all variables are included in log forms. Standard errors are clustered. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level

The question of this simple panel regression exercise is whether the speed of the spread of different technologies is increased by "better" formal institutions. In short, the answer is not much. When it is done with the growth of income the results are what can be expected and have been much more carefully proved in the literature before: good institutions are a factor behind a fast conditional convergence. The same cannot be said for different technologies (see Tables 3a,b). Clearly, there is a convergence effect because in almost every case (with the exception of telephones where the initial value is significant but only at the ten percent level)

the coefficient of the initial values of the technology measure is negative. However, the measure of economic freedom is not significantly positive in general: there are only four cases out of fourteen when it is significant at a significance level below five percent.

To take a closer look, I split the sample into a developed and a developing group. The reason behind this is that in developed countries the technology adoption may slow down and the results may be dominated by the developed countries. The difference between developed and developing countries is established by using the results in Eichengreen et al. (2011). Examining the slowdowns in economic growth they (ibid:9) came to the conclusion that "a growth slowdown typically occurs when per capita income reaches 58 per cent of that in the lead country". Accordingly, a data point is identified as referring to a developing country if the GDP per capita in that year is lower than 58 percent of that of the United States.

Table 4a

The effect of economic freedom on the diffusion of different technologies in "developed" countries

	Dependent variable: growth rate of							
	GDP per capita	ships	railways-pass.	railways-freight	cars	trucks	aviation-pass.	aviation-freight
constant	0.323 (3.00)***	0.749 (2.33)***	1.166 (4.25)***	0.812 (3.28)***	0.771 (9.03)**	0.160 (2.42)**	0.484 (1.63)	0.432 (1.51)
initial value	-0.043 (-3.67)***	-0.104 (-2.07)***	-0.137 (-5.22)	-0.098 (-3.17)***	-0.057 (-9.11)***	-0.030 (-4.87)***	-0.042 (-3.81)***	0.027 (-1.71)*
economic freedom	0.067 (3.73)**	-0.044 (-0.73)	0.043 (0.58)	-0.120 (-1.72)*	0.062 (2.67)**	0.037 (1.23)	-0.016 (-0.13)	-0.103 (-0.75)
no. of countries	29	11	19	21	29	24	24	24
no. of obs.	171	34	70	77	118	103	5	84
R^2 (within)	0.187	0.443	0.375	0.272	0.599	0.093	0.215	0.097
R^2 (between)	0.153	0.069	0.349	0.000	0.041	0.016	0.078	0.019

Table 4b

The effect of economic freedom on the diffusion of different technologies in "developed" countries

	Dependent variable: growth rate of						
	telephone	cell phone	pers. computers	internet users	MRIs	Blast ox. steel	electricity
constant	0.720 (5.07)***	0.585 (0.37)	2.510 (5.72)***	-5.580 (-1.27)	-0.433 (-0.29)	1.314 (6.31)***	1.066 (8.17)
initial value	-0.047 (-6.44)***	-0.074 (-3.14)***	-0.077 (-7.09)***	-0.110 (-2.60)**	-0.098 (-5.02)***	-0.162 (-8.34)***	-0.044 (-8.66)***
economic freedom	0.026 (0.68)	0.422 (0.50)	-0.614 (-2.46)**	3.801 (1.69)	0.464 (0.61)	0.068 (0.94)	0.031 (1.00)
no. of countries	27	27	26	26	13	19	29
no. of obs.	96	68	52	43	25	95	137
R^2 (within)	0.532	0.321	0.850	0.532	0.752	0.587	0.369
R^2 (between)	0.074	0.057	0.530	0.012	0.000	0.007	0.034

Notes: all variables are included in log forms. Standard errors are clustered. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level

As Tables 4a,b and 5a,b show, the results are not that different in this case. Considering per capita GDP there is no large difference between the two groups (see "GDP per capita" column in Tables 4a and 5a): better institutions have a positive effect in both cases. When it comes to the technologies involved, the results say, again, that the effect is not large; however, they seem to be somewhat different between the two groups. In the developed group there is only one technology (cars) where economic freedom positively effects the catching-up process, but there is a technology (personal computers) where this effect is negative. In the developing group there is no significantly negative effect, but there are positive ones: trucks and telephones.

Table 5a

The effect of economic freedom on the diffusion of different technologies in "developing" countries

	Dependent variable: growth rate of							
	GDP per capita	ships	railways-pass.	railways-freight	cars	trucks	aviation-pass.	aviation-freight
constant	0.294 (5.31)***	0.463 (4.50)***	1.051 (3.90)***	0.670 (1.91)	0.673 (3.24)***	0.121 (1.55)	0.831 (8.32)***	0.353 (1.92)*
initial value	-0.047 (-6.51)***	-0.086 (-4.85)***	-0.129 (-3.87)***	-0.084 (-2.39)**	-0.056 (-3.73)***	-0.046 (-5.39)***	-0.110 (-13.45)***	-0.127 (-8.15)***
economic freedom	0.066 (7.09)***	0.041 (0.87)	-0.053 (-1.01)	0.019 (0.29)	0.047 (1.13)	0.094 (2.18)**	0.020 (0.37)	0.130 (1.22)
no. of countries	93	35	52	45	84	71	64	61
no. of obs.	472	118	145	125	272	243	185	177
R^2 (within)	0.204	0.505	0.335	0.159	0.184	0.175	0.619	0.565
R^2 (between)	0.144	0.000	0.001	0.046	0.003	0.105	0.043	0.002

Table 5b

The effect of economic freedom on the diffusion of different technologies in "developing" countries

	Dependent variable: growth rate of						
	telephone	cell phone	pers. computers	internet users	MRIs	Blast ox. steel	electricity
constant	-0.02 (-0.01)	-0.836 (-0.55)	0.799 (2.60)**	n. a.	n.a.	1.510 (3.33)	0.975 (5.86)***
initial value	-0.008 (-0.88)	-0.183 (-3.14)**	-0.088 (-4.27)***			-0.246 (3.41)***	-0.0426 (-5.47)***
economic freedom	0.114 (2.81)***	2.083 (1.86)*	0.301 (1.04)			0.193 (1.48)	0.028 (1.76)*
no. of countries	83	71	61			31	91
no. of obs.	228	98	91			113	327
R^2 (within)	0.035	0.439	0.440			0.727	0.255
R^2 (between)	0.008	0.009	0.231			0.003	0.049

Notes: all variables are included in log forms. Standard errors are clustered. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level

In sum, the fact that seems to offer a challenge to an institutional interpretation of economic growth is that although income convergence is accelerated by better institutions, the same cannot be said for the diffusion of different technologies. The latter are very much unaffected by better formal institutions. Secondly, there is some difference between developed and developing countries. In the latter group the effect of economic freedom on technology adoption is somewhat stronger than in developed countries. This may not be a simple convergence effect, for the reason provided by Comin and Hobijn (2006), who show that technology adoption does not follow a logistic pattern.

In what follows I will respond to this challenge by focusing on the role of culture, proceeding in two steps. First I will argue – drawing lessons from scholars of the (first) industrial revolution and others – that the culture that is important in economic development is a broad ideology, one aspect of which is the level of disgust towards rent-seeking. Second, using a very simple model of economic growth, I will try to figure out an answer to the puzzle that culturally embedded institutions are important factors in economic development but seemingly not so in technology diffusion.

3. The ideology of freedom as a constraint on expropriation

3.1. Freedom as a good

A conclusion that I draw from the results I reviewed in section 2.1 is that formal institutions play a constitutional and a sub-constitutional role. Their constitutional role is to provide the preferences when the choice between rules is made. The sub-constitutional role is the role these values play in affecting behavior within the rules. The proposition of this section is that there is a broad ideology which, following Mises (1953), I refer to as the ideology of freedom, which matters for economic development for both a constitutional and a sub-constitutional reason. As a broad ideology it makes people prefer rules that give innovators the freedom that is needed for innovation. On the sub-constitutional level this ideology is a preference for productive over rent-seeking activities.

My claim is that with hindsight Mises (1953) was right in claiming in his introduction to the *Wealth of Nations* that Smith's book "presented the essence of the ideology of freedom, individualism, and prosperity". Several economic historians support the claim that the way people thought about the economy changed around the industrial revolution, and this change was the major cause of the era of economic development that followed.

According to Mokyr (2007) the idea of a *useful* science was the most important ideological shift that induced, albeit with a considerable time lag, the industrial revolution. Mokyr's insight seems to suggest that it is the idea of usefulness that is important, not freedom. But what Mokyr describes as an ideological shift seems to be rather a change in research methods and in the views people hold of science. As scientists begin to think of their own activity as something that can and must improve people's well-being, people begin to think of science as something that could improve their well-being even if they do not understand the ways it achieves its effects. This argument is also underpinned by Goldstone (2002) who explained that it was its engineering and scientific culture that made Britain unique around the industrial revolution. In this British approach to science experimentation played a key role, together with the notion of "optimizing, perfecting, and continuously improving through engineering" (ibid: 372). Elsewhere (Goldstone 2000) he identifies Britain's unique "Newtonian culture" as one of the "chance factors" that constituted a "mechanistic world-view, a belief in fundamental discoverable laws of nature, and the ability of man to shape his world by using those laws" (ibid: 184).

As a consequence, thinking of science as a useful tool to improve people's lives does not contradict the idea of giving the value of individual freedom the highest priority on a scale of values. Once people share the belief that science is useful, they have a reason to give freedom to those who advance and apply knowledge. This argument is affirmed by the fact that the new methodology of research put great emphasis on experimentation and trial and error, and the creatively destructive nature of scientific discovery. Accepting that discovery is the result of a process of trial and error is to accept that some consequences cannot be calculated in advance. That is why Mokyr (2007:15) comes to the conclusion that "[i]ntellectual innovation could only occur in tolerant societies in which possibly outrageous ideas proposed by sometimes highly eccentric men would not incur violent responses against 'heresy' and 'apostasy'".

All things considered it comes as no surprise that it is also Mokyr (2010) who emphasizes the importance of "culture" as a crucial factor that triggered the (British) Industrial Revolution, with the term culture referring to "a set of shared beliefs, attitudes and preferences that are passed on from generation to generation through *nongenetic* (i.e. soft-wired) mechanisms" (ibid:203, endnote 3, emphasis in original). What Mokyr (2008, 2010) argues for is that the gentlemanly code of conduct was responsible for creating a social environment by 1700 in which entrepreneurship and the partnership between entrepreneurs and engineers (those possessing marketing skills and technological skills) became easy. This code of conduct meant that "people who felt constrained by the gentlemanly code of behavior behaved honorably, kept their word, and did not renege on their promises. They did not blindly maximize profit" (Mokyr 2010:190). That is, the shared belief concerning how a gentleman should behave created a high level of trust and decreased transaction costs. In order for a culture to work as a growth enhancing mechanism it must be a commonly held knowledge: culture reveals the way people think other people will behave. This mechanism has been made famous by Greif et al. (1994) who analyzed the trade of the Mediterranean area and explained why it was the individualistic culture of the Genoa traders that finally led to an acceptance of the rules of a market economy and not the collectivist culture of the Maghreb traders.

Others see the content of culture as more complicated or at least multidimensional, but these more sophisticated analyses also emphasize that to accept the ideology of freedom is to accept some constraints on individual behavior, even if the individual cannot identify a direct explanation of why it is useful to follow such rules. Wilhelm Röpke (1969[1933]) clearly contrasts materialism and "liberalism" which makes people believe that "punctual trains are too high a price for the loss of freedom" (ibid:85). He argues that liberalism means not only this high value placed on liberty, but also the use of reason which leads to promoting economic freedom, and humanity meaning "absolute respect for every individual's human dignity" (ibid: 91). Elsewhere he (Röpke 1959:233) also describes his own position as contrasting freedom with material welfare by saying that "I would stand for a free economic order even if it implied material sacrifice and if socialism gave the certain prospect of material increase. It is our undeserved luck that the exact opposite is true".

In a somewhat similar but much more sophisticated fashion McCloskey (2006) describes a complete system of virtues as a balanced system of partially ancient, partially Christian, and partially modern virtues. Her argument is also in line with the view that pure prudence would not have been enough to trigger what we now call economic growth. Rather, this was the result of creating a "bourgeois" ideology (and rhetoric), liberty and dignity for the bourgeoisie. In this she (McCloskey 2010) abbreviates a long argument, according to which a commercial society can be virtuous. Once it was discovered – thanks to the Enlightenment

thinkers – that market exchange is a positive sum game, it became accepted that profit seeking can be in line with the virtues people think of as right.[67]

Hayek (1971) also similarly claims that the precondition of a free society is a value system which is ideological or dogmatic in the sense that people who hold these values reject various actions of the government on the ground that these actions are not in line with these principles. Thus, the condition of a free society is a value system that constrains "expediency" or which gives priority to certain principles over material gains. [68]

In sum, the ideology of freedom formulated around the time of the first industrial revolution includes a valuation of freedom for its own sake. According to this order of preferences, freedom is a good for which people are willing to forgo some other goods. This is derived from the belief that freedom makes it possible for people to make discoveries that increase the wealth of all. That is, the ideology of freedom must include a belief that there are productive activities the practising of which creates wealth instead of simply redistributing it.

3.2. The ideology of freedom as an attitude towards rent-seeking

The idea that an important element of the ideology of freedom is seeing the difference between productive and rent-seeking activities is in line with the facts emphasized by the scholars cited in Section 1 and 2, or with Hayek's (1970) argument that "tradition" is the result of cultural evolution. That is, cultural rules are not chosen deliberately, they also do not result from biological evolution. Cultural evolution is the process in which those rules that support the flourishing of a community will survive. This is why, Hayek argues, the rules of a free society emerged: since these rules make a community more prosperous than others, people tend to follow them, although they do not know any reason why they should follow them. The ideology of freedom is on the one hand a belief that man should have integrity or "absolute property rights" (Facchini 2002). On the other hand it also includes, if only implicitly, the belief that the activities made possible by freedom are productive, that is, they benefit the whole community. Hayekian cultural evolution may thus explain how and why people can come to follow the rules that the ideology of freedom would prescribe for them.

As De Soto (2009) argues, if one looks on the market as the process of entrepreneurial discoveries, distributional ethics loses ground, because this type of ethics is not in line with the fundamental principles of the market. It is inherently antagonistic to the security of property rights and to the view that what is to be distributed is not given, but created by

[67] Looking at the details, however, will reveal some fundamental differences between these views, the most important of which is how western, how European, how deeply seated these view are deemed to be. Röpke (1969[1933]) clearly argues that liberal views are the results of a long intellectual evolution in Europe, while McCloskey (2006, 2010) considers them the results of intellectual changes in the 1700's and does not identify them as specifically European.

[68] It is notable that while Great Britain was the first nation to experience modern economic growth, many describe the British (or the English) as holding an ideology that is similar to what was described above. The philosopher Scruton (2006:205) for example notes, that "[w]hether collectivist or individualist, Burkean traditionalist or Benthamite utilitarian, the English political thinker would take it for granted that liberty was all-important and not to be exchanged for some other, and invariably lesser, good". Similarly, Hayek (1971:35) writes that "[t]he impression that the English in the 17th and 18th centuries, through their gift of 'muddling through' and their 'genius for compromise', succeeded in building up a viable system without talking much about principles, while the French, with all their concern about explicit assumptions and clear formulations, never did so, may thus be misleading. The truth seems to be that while they talked little about principles, the English were much more surely guided by principles, while in France the very speculation about basic principles prevented any one set of principles from taking a firm hold." (Hayek 1971:35). Moreover, a much more recent and much less philosophical expression of this idea comes from Campbell (2010) who, expressing his disappointment over England's failure to win the right to organize the 2018 FIFA World Cup writes on the English that "[w]e like the idea of freedom of the press, *even if* at times we don't like what it delivers" (italics are mine):

entrepreneurial activities. What he identifies as an appropriate ethical principle for a well-functioning market is the one that "all human beings have a natural right to the fruits of their own entrepreneurial creativity" (ibid:175). However vague this principle seems to be[69], it shows that the ethics of the market are characterized by entrepreneurship and, as a consequence, creative destruction needs an ethic which considers the private property of the entrepreneur as something that should be held in respect. That is, a person holding these ethical principles, which I previously referred to as the ideology of freedom, can be imagined as if she held the view that value is *created* on the market (that is, exchange is a positive-sum game) and expropriation of entrepreneurs' profits is a violation of property rights, which is immoral.

That amounts to saying that a useful way to model the effect of this kind of ideology is to think of it as an attitude toward rent seeking and property rights. More precisely, what I have called the ideology of freedom can be seen as an attitude toward property rights which can be seen as an attitude toward rent seeking[70]. Entrepreneurial discovery can also be a discovery of a rent seeking opportunity as Buchanan (1980) explains. "The difference lies in the unintended results. Political reallocation achieved via rent seeking, does not reduce or eliminate contrived scarcity" (ibid.11). Profit seeking, however, does, because profit comes from mutually beneficial market exchanges. Reducing rent seeking is securing the property rights of those seeking profit. In this spirit, Benson (1984) shows that we should think of rent-seeking as getting governments "to reassign the right by taking it from its current holder" as opposed to entering into "a voluntary private exchange with the current holder of the right" (ibid:390). This is why the rise of rent seeking represents a deterioration of "market orientation" and "of the constitutional constraints on the government's ability to take property rights from private citizens" (ibid:396)

Reading Crafts (1995) may also lead to the conclusion that rent seeking should be incorporated into models of economic growth. Making different growth models face the facts of the industrial revolution he puts great emphasis on rent seeking. Based on the historical data he accepts the claim that part of the explanation concerning Britain's success should be the relatively low level of rent seeking (as compared to France), but he shows that the occupations deemed to be rent-seekers' positions (positions in the church or law for example) were held in higher esteem than those that are thought to be productive (such as engineers or bankers).

4. Rent seeking as the expropriation of entrepreneurial rents

4.1. A simple model of productive and unproductive entrepreneurship

Based on the arguments in the previous section I will present a simple model to show that by introducing the idea of rent-seeking and that of the preference over productive and unproductive activities into a model of innovation the challenges mentioned in Section 2 may be answered. The model is based on the following broad assumptions:

(1) The adoption of a certain technology (and economic growth in the end) is the result of entrepreneurial discoveries. Such a discovery (or innovation) makes the entrepreneur a monopolist for a period.

[69] It is vague because it is contradictory by its own logic of the market process. If a non-consequentialist ethical rule is needed because of the unpredictability and genuine uncertainty of the market process, then how can one know what the fruits of her entrepreneurial discovery are? Using profit instead may make this notion less vague, but after all it is about respecting property rights as an end, not as means.

[70] In the light of the historical puzzle I mentioned in section 1 this is in line with the proposition of Mokyr (2009:63) saying that "[i]n the second half of the eighteenth century, most important intellectuals became increasingly hostile to what modern economists would call rent-seeking"

(2) The ideology of freedom can be described as an attitude toward rent seeking activities. Consequently its intensity can be expressed as a wage premium of rent-seeking activities over productive ones.

(3) The expropriation rate is defined as the percentage of the entrepreneur's profit that is expropriated by the rest of the society. The expropriation rate is the result of allocation decisions of possible rent-seekers, which is further determined by cultural variables.

The specification of assumption (1) is based on Aghion and Howitt (1998, 2005) who develop the Schumpeterian model of economic growth with the same assumptions. Accordingly, I am going to derive my proposition by using the simplest version of their Schumpeterian model although I am going to use the model with a slightly different interpretation than is usual. The aim is to model the behavior of just one technology and the "final good" delivered by that technology.[71] Assume that the final good is produced with a Cobb-Douglas production function in the form:

$$y = Ax^{\alpha},\ 0 < \alpha < 1, \tag{1}$$

where x is the quantity of intermediate good used to deliver the final, "technological good" (such as the extent to which cars or airplanes are used), which is characterized by an efficiency parameter A. If a new intermediate good is discovered and introduced to the market this parameter becomes γA, $\gamma > 1$. With L_a amount of labor allocated to research activity the probability of a new discovery is μL_a.

The producer of the intermediate good has a monopoly since she discovers (or applies) a new technology. The intermediate good is produced by using labor in a one-by-one technology, that is $x=L_x$, where L_x is the amount of labor allocated to the intermediate sector. Since this firm is a monopoly its profit can be written as

$$\pi = A\alpha(1-\alpha)L_x^{\alpha}. \tag{2}$$

Assumption (2) is based on the idea that rent seeking is nothing more than any other kind of redistribution that incurs costs on society (Murphy et al. 1993:409), and at the same time weakens the security of property rights. Ethical or ideological views on private property can be seen as ideological or ethical views on redistribution and thus on private property.

In this spirit I assume that the consumption of the non-entrepreneur majority comes from two sources. One is the wage paid by the entrepreneur in the intermediate sector. The other is rent-seeking which is equivalent to expropriating the profit of the entrepreneur. That is, I assume that there is a rent seeking sector which aims at expropriating the profit made in the intermediate sector. The technology of this sector is given as

$$t(L_r) = \frac{1}{\theta L_y / L_r + 1}, \tag{3}$$

where t(.) is the share of the intermediate sector's profit which is expropriated through rent seeking, L_r is the labor which is allocated to rent seeking activities, L_y is the total productive labor in the economy, and $0 < \theta$ is a parameter describing the effectiveness of the rent seeking activity[72] (the smaller it is the more effective rent seeking is), and it can be

[71] The reason is that the technology data used in section 2.2 fit this interpretation. Clearly, these data do not describe "technology" in the sense in which it is simply an idea with whose help one can produce the same output at a lower cost, or a new product etc. But "cars", "ships" or "computers" are not purely "intermediate goods" either because what we mean here is the product of using the cars, ships and computers or railways. In addition, "cars", "ships" or "computers" have been steadily improved with new ideas since they were first invented. Consequently, a technology as understood in the CHAT database used above is at best a blend of what is a new technology and what is an intermediate good in the model.

[72] This kind of technology comes from the rent seeking literature. Grossman (2002:36), for example, uses a similar technology to describe expropriation as a function of resource allocation to guard against producers. If guarding activities are seen as part of rent seeking then the technology described in equation (3) can be seen as an application of Grossman's technology to this particular problem. Here I do not suppose protection as an

thought of as being set at the constitutional level of choice. This parameter represents the constitutional role of formal institutions defined in the previous section.

The sub-constitutional role of ideology is modelled as a subjective wage premium on rent-seeking activities. This is the immediate consequence of the way the ideology of freedom was described above: as an attitude toward rent seeking. When people think that rent seeking is not held in as high esteem as productive activities, they will only be willing to devote their time to rent seeking if this activity receives a premium. To formulate this idea I assume that there is a wage premium of δ percent that people must be paid to be willing to allocate labor to the rent-seeking sector.

That is, labor market equilibrium can be characterized as:

$$t(L_r)\pi/L_r = (1+\delta)w \tag{4}$$

where the right-hand side of equation (4) is the expected payoff of rent-seeking activity, and π is the expected profit of the entrepreneur. The net payoff of innovation may or may not increase, depending on the arrival rate of innovation, which in turn depends on the amount of labor allocated to the innovative sector. The right-hand side is the wage paid in the intermediate sector plus the wage premium: the higher this wage premium is, the stronger the bias against the rent seeking sector.

Taking into consideration that

$$w = A\alpha^2 L_x^{\alpha-1}, \tag{5}$$

and considering the rent-seeking technology defined in equation (3) the equilibrium criterion breaks down to

$$\frac{1}{\theta+(1-\theta)L_r}\frac{1-\alpha}{\alpha}L_x = (1+\delta). \tag{6}$$

That is, the "objective" wage premium must be equal to the subjective one, and the objective wage premium is an increasing function of the labor allocated to the intermediate sector (other things being equal). The reason is that the profit of the intermediate sector is increased when more people work there and the wage in that sector is lower. This gives an expression for the labor allocated to the intermediate sector:

$$L_x = \frac{\alpha\theta(1+\delta)}{1-\alpha} + \frac{\alpha(1-\theta)(1+\delta)}{1-\alpha}L_r. \tag{7}$$

The other equilibrium criterion of the model is the traditional one, described by Aghion and Howitt (2005:70) as the research arbitrage equation saying that the marginal increase in net profit as a result of one unit more labor allocated to research must be equal to the wage rate:

$$[1-t(L_r)]\gamma\mu\pi = w. \tag{8}$$

The left-hand side includes the term $\gamma\mu$ because new labor allocated to the research sector will increase the probability of innovation. Taking into account what was said above, this equation will have a simple form, too:

$$\left(1-\frac{L_r}{\theta(1-L_r)+L_r}\right)\mu\gamma\frac{1-\alpha}{\alpha}L_x = 1. \tag{9}$$

Thirdly, the labor market clears. By normalizing total labor force to one this means that

$$L_x + L_a + L_r = 1. \tag{10}$$

This three-equation model (described by equations (6), (9), and (10)) can easily be solved for the equilibrium levels of L_a, L_x and L_r. Since the growth of income depends on the level of

---

alternative to which labor can be allocated. This technology of rent-seeking incorporates the most important feature of rent seeking emphasized by Murphy et al. (1993). This is that the payoff of rent-seeking is affected by the amount of rent-seeking activities in two ways: directly by showing diminishing marginal returns; and indirectly by reducing the output which is produced and can possibly be (re)distributed.

L_a, this lets me draw conclusions concerning the final output. Substituting the expression for L_x in equation (9) allows us to calculate the equilibrium value of L_r:

$$L_r = 1 - \frac{1}{\gamma\mu(1+\delta)\theta}. \tag{11}$$

Equation (11) shows that in this model the role of culture is controversial. In an institutional environment in which the informal wage premium against rent-seeking is higher the amount of labor allocated to rent seeking will also be higher. This is implied by the equilibrating process coming from the assumptions of the model. First, when δ increases, the amount of labor allocated to the intermediate sector will increase too, since its equilibrium level is:

$$L_x = \frac{\alpha}{1-\alpha}\left[(1+\delta) - \frac{\theta}{1-\theta}\frac{1}{\gamma\mu}\right]. \tag{12}$$

Second, because of this increase in L_x the potential profit of innovation will increase which, in turn, will provide higher prey for rent-seekers, and rent-seeking will increase. Since L_x and L_r increase at the same time, L_a, and as a result, the rate of innovation will *decrease*. This can be seen more easily if the expression for the arrival rate of innovation is derived explicitly:

$$\text{rate of innovation} = \mu\gamma L_a = \mu\gamma(1 - L_r - L_x), \text{ that is} \tag{13}$$

$$\text{rate of innovation} = \frac{1}{(1+\delta)\theta} + \frac{\alpha}{1-\alpha}\frac{\theta}{1-\theta} - \frac{\alpha}{1-\alpha}(1+\delta)\gamma\mu. \tag{14}$$

Clearly, the higher δ is, the lower the rate is. This is the result of the feature according to which a higher level of esteem for productive activities will increase the work allocated to the intermediate sector (see equation (12)), but it also makes the (material) payoff of rent-seeking higher. As a result, labor allocated to both activities will increase, and the labor allocated to innovation cannot but decrease. The negative effect of this dimension of culture is controversial. The effect of the other dimension represented by θ is, however, less so. A worsening of the technology of rent seeking (an increase in θ) depends on the value of δ. The effect of an increase in the disgust towards rent seeking on the constitutional level will be positive only if this attitude is negative enough already. Consider first that the innovation rate will be positive only if $L_r+L_x<1$, that is, if

$$\gamma\mu < \frac{(1-\alpha)}{\alpha\theta(1+\delta)^2} + \frac{\theta}{(1-\theta)(1+\delta)} \equiv K(\theta). \tag{15}$$

This implies that growth is only possible if $K(\theta)>0$ that is to say if

$$\delta > -\frac{1-\alpha}{\alpha}\frac{1-\theta}{\theta^2} - 1 \equiv \delta_1. \tag{16}$$

Secondly, innovation becomes more probable if $K(\theta)$ increases, and K will be increased by a higher θ ($K'(\theta)>0$), if

$$\delta > \frac{1-\alpha}{\alpha}\left(\frac{1-\theta}{\theta}\right)^2 - 1 \equiv \delta_2. \tag{17}$$

Thirdly, a worsening of the rent-seeking technology enhances the growth rate when $\partial L_a/\partial\theta > 0$ which holds under the same condition formulated in (17).

It is clear that $\delta_2 > \delta_1$ which implies that there is a range of δ ($\delta_1 < \delta < \delta_2$) within which although innovation is not zero and growth is possible, its rate will not be increased by an

increase in θ.[73] When δ is high enough ($\delta_2 < \delta$), the higher the wage premium, the more probable it is that an increase in θ will lead to a relatively high level of innovation.

4.2. The effects of a change in culture and in the expropriation rate

To see the effect of a change in the expropriation rate (t), it is useful to write the model in the following way:

$$L_x = \frac{\alpha}{1-\alpha}\frac{1}{1-t}\frac{1}{\gamma\mu}, \tag{18}$$

$$L_r = \frac{t}{1-t}\frac{1}{\gamma\mu}\frac{1}{1+\delta}, \tag{19}$$

$$t = \frac{L_r}{\theta(1-L_r)+L_r} = \frac{\gamma\mu(1+\delta)\theta-1}{(\gamma\mu(1+\delta)+1)\theta-1}. \tag{20}$$

It is then possible to make a difference within the model between an exogenous and an endogenous change of the expropriation rate. An exogenous change in the expropriation rate is meant to model the top-down change in formal institutions. In this case the expropriation rate is not at the level which would be the result of the equilibrating forces of the model. A $t = \bar{t}$ is enforced and not determined by the choice of the players as to how to allocate resources between rent-seeking and productive activities. To put it another way L_r is set below that level which would equilibrate the labor market. In this case the effect on the quantity of the final product is more straightforward: the expropriation rate is reduced (by definition), and as a result the quantity of the final good is increased, since

$$\frac{\partial \ln y}{\partial t} = -\frac{\theta}{[1-(1-\theta)t]^2} - \frac{1}{(1-t)^2}\left[\frac{1}{L_a} - \alpha(1-t)\right] < 0. \tag{21}$$

When the source of institutional change is the change in the cultural parameters, the effect is much less straightforward. Most importantly, it is not always true that when culture changes for the better, the expropriation rate will decrease, since

$$dt = \frac{\partial t}{\partial \theta}d\theta + \frac{\partial t}{\partial \delta}d\delta, \text{ where} \tag{22}$$

$\frac{\partial t}{\partial \delta}$ and $\frac{\partial t}{\partial \theta}$ are both positive if δ is large enough.

When the expropriation rate is reduced as a result of a cultural change the change in the final output can be written as

$$d\ln y = \frac{\partial \ln y}{\partial t}dt + \alpha\frac{\partial \ln y}{\partial \delta}\bigg|_{t \text{ is constant}} \times d\delta. \tag{23}$$

Culture has a direct and an indirect effect. The indirect effect works through the expropriation rate. However, this effect is not exactly as would be expected, since a change of culture for the "better" (a higher value of θ and δ) will not necessarily reduce the expropriation rate. The direct effect can be seen in equations (18) and (19): provided that the expropriation rate is constant a better culture will reduce rent-seeking, but the quantity of intermediate goods will remain intact. As a result, the partial effect of the cultural change will be positive on the final good production. That is, a cultural change may not decrease the expropriation rate even if it has a positive effect on the spread of technology.[74]

---

[73] Some further criteria that are necessary so that the model can have an interior solution are elaborated in the Appendix, including the extreme case where $\theta = 1$

[74] The mathematical details of the results in this subsection are to be found in section A3 of the Appendix.

5. Growth and technology across cultures: some empirical results

5.1. Implications and predictions

When discussing the reasons why economic growth died out in different eras before the industrial revolution, Mokyr (2004:31) points out, that "[p]rosperity and success led to the emergence of predators and parasites in various forms and guises who eventually slaughtered the geese that laid the golden eggs. Tax collectors, foreign invaders and rent-seeking coalitions as guilds and monopolies in the end extinguished much of the growth of northern Italy, southern Germany, and the Low Countries."

The simple model above focuses on a mechanism of this kind. The rule of culture in this is twofold. The first role, which can be called the sub-constitutional role of culture, is to make people dislike rent-seeking activities and to choose productive work instead. The second, constitutional, role of culture is to make rent-seeking more costly by choosing formal institutions that set constraints on it.

Distinguishing between different layers or dimensions of cultural beliefs is common. One of these levels, which is very much akin to the constitutional – sub-constitutional difference made here, is Arruñada's (2010) categorization of ethics into a social ethic and a work ethic. However, in almost every work mentioned in Section 2 and 3 culture is seen as a multidimensional phenomenon. Seen in this light this simple model is a step taken to integrate this multidimensional nature of culture into a model of economic growth.

The explanation provides an answer for the puzzle raised in section 2; namely that while it is broadly held that it is formal institutions that cause growth, many economic historians think that the industrial revolution was caused by a change in the informal institutions without the formal ones becoming freer. What is more, as McCloskey (2010:331) notes "in some ways modern economies – with their gigantic administrative states spending half of national income, and regulating still wider fields of economic activity – create less, not more, security of property than a feudal economy with diffuse centers of power, or than an early modern state such as Stuart England with a less-than-impressive ability to tax".

The model gives an explanation for why the security of property rights may not have to decrease to induce technology diffusion. This explanation is that the risk of expropriation is an endogenous variable that is determined by the allocation of labor between the three different sectors which in turn is affected by cultural factors. Culture affects rent-seeking and thus property rights security both through the constitutional and the sub-constitutional level. The expropriation risk is the function of institutions but they are not identical. Even if institutions improved because of the spread of the ideology of freedom, property rights security will not necessary grow, because the change in the informal institutions will increase the payoff of rent-seeking activities.

The overall prediction of the model is thus concerned with the effects of a change in the expropriation rate on the final good of a certain technology. First, if the institutional change comes from a shift in attitudes towards the ideology of freedom, the expropriation rate will not necessarily reduce if new formal and informal institutions disfavor rent-seeking. And even if the expropriation rate is reduced by a change in culture, the direct effect of the cultural change in the final production of the good may be the opposite of that of the expropriation rate. If the change comes from a top-down "reform" the expropriation rate is reduced and the quantity of the final good is increased. Admittedly, the model does not say anything about the sustainability of such a disequilibrium situation.

These predictions are in line with the results in Williamson and Mathers (2011a,b) who have much to say about the interactions between formal and informal institutions. They found that overall economic freedom is more important as a determinant of economic growth when

cultural variables are accounted for, but they interpret this as a consequence of a substitution between formal and informal institutions, not as two different effects.

5.2. Data and results

My crucial data, as is clear from the implications discussed above, are those on culture and those on the final goods produced by certain technologies. I will use other databases of course, but they are rather the usual ones whose use requires less reasoning.

The data on culture is the same as in Section 2.2 and comes from Licht et al. (2007).[75] One advantage of this data set, its being well-founded in theory and richness aside, is that Schwarz (2004) and Licht et al. (2007) identify cultural regions based on the specific values along the seven dimensions. This is useful in two aspects. First, one can argue that culture is too complex to be measured on a continuous scale. In this case one can use the regions as dummy variables, or can make a difference between country groups belonging to different cultures. Using the variables this way has an additional advantage. Although Licht at el. (2007) provide detailed data for only 49 countries, this list can be longer if one only needs information concerning the cultural region to which a country belongs. As these regions seem to be identifiable on relatively objective criteria, we can make the database larger by using these criteria to associate countries not included in the Licht et al. (2007) database with cultural regions. I see these cultural variables as the cultural roots that explain the differences in preferences as supposed in the model above. It must be noted, however, that these dimensions do not say anything directly about the preferences toward rent-seeking.

The second crucial question that arises when trying to falsify the predictions I have drawn from the model, is how to measure the quantity of technological goods. Here I will use the so called CHAT (Cross-country Historical Adoption of Technology) database developed by Comin and Hobijn (2009) and already used in section 2.2. They also analyzed the data and came to several interesting conclusions. Their concern is different from mine, however. They use the data to explain the factors of technology diffusion (Comin and Hobijn 2010a,b), showing several interesting conclusions (also Comin and Hobijn 2006). One of them is that technology adoption is even more diverse than income or growth rates across countries. Comin and Hobijn (2010a) show that lags in adoption of different technologies account for 25 percent of the variability in per capita income.

The reason why this database is the right one for my purpose here is that it is designed to measure technology adoption at the intensive margin. That is, these data show – using typical data on those capital goods in which the technology is embedded – on what scale a technology will spread through a country once it has been devised. Clearly, the measures such as passenger kilometers by car, or the number of computers and mobile phones per capita should be understood from the perspective of the model in Section 4. These are goods that are used to satisfy consumers' needs. They are not defined in terms of value but in physical units. The CHAT database provides data on more than a hundred technologies, but Comin and Hobijn (2010) use only 15, grouped into six categories. Since the technology of the telegraph seems to be outdated for the time span I focus on, I dropped this technology. I will focus on the fourteen technologies shown in Table A1 in the Appendix.

Measuring the expropriation rate is also not a straightforward task. The key is that in my argument the expropriation risk is a result which characterizes the equilibrium of the model. It is shaped by the parameters, most importantly by those that describe formal and informal institutions. I claim that this kind of expropriation is what is measured by the usual institutional indexes, and this is what Gleaser et al. (2004) propose; they criticize the typically used indexes on the basis that they measure the outcome of a process and not the constraints

[75] The dimensions of culture in this dataset are described in footnote 6.

that the (political) players face.[76] This shortcoming of the data is what I can make use of, since in the argument presented in this paper the expropriation rate is really an outcome measure. Secondly, the economic freedom index is really meant to measure the expropriation by governments through different channels. As Gwartney and Lawson (2003:407) explain "[t]he concept of economic freedom … is closely related to the presence of protective rights, that is, rights that provide individuals with a shield against others who would invade and/or take what does not belong to them".

Tables 6a,b, 7a,b show the results of testing the main predictions of the model above. These predictions, as we saw, concern the partial effect of the expropriation risk and of culture. The model predicts that the change in the expropriation rate without a change in culture and a change in culture without a change in the expropriation rate will have a greater effect than when the two happen together. The regressions whose results are presented below follow two strategies to capture these effects. On the one hand I will use data on culture mentioned above to be able to "hold culture constant" in the regressions, since the coefficients will show the partial effects of the variables.

However, since this strategy leads to a radical decline in the number of countries that can be included into the regressions (as compared to their number in Tables 3-5), I will apply an alternative strategy, too. Based on what was concluded in section 2.2 I will split the EFW index into two parts: one element including the security of property rights measure and the other the mean of the other four areas. The idea behind this strategy is that it was shown that it is the property rights measure that is to some extent determined by culture while the other four are only slightly affected. This helps us differentiate between that proportion of the expropriation risk that is determined by culture and that proportion that is not. To put it differently, the mean the other four areas represent is the proxy for the exogenous change in the expropriation risk.

In Tables 6a and b the initial value of the technology, of economic freedom, and of the three above-mentioned cultural dimensions are regressed on the growth rate of the technology in question with a pooled OLS method. In the light of the previous predictions the results are mixed. First, in eight cases out of fourteen some of the cultural variables are significant at a 5 per cent level, while economic freedom is also significant in six cases. Second, the signs of the effects are roughly those that can be expected, based on this paper's argument. In those cases when economic freedom is significant its effect is positive except for one case (cell phones).

Clearly, the results in Tables 6a,b do not perfectly fit the model, but it shows that the mechanism emphasized in it may be a part of the "culture and growth" story. In addition to showing that in many cases the effect of an exogenous decrease of the expropriation rate is positive, it also shows that culture directly contributes to the adoption of technology beyond its indirect contribution. The effect of formal institutions seems, however, to contradict what the model suggests. The most important cultural variable is hierarchy, which at first sight is precisely the opposite of what the ideology of freedom should mean. But this impression is wrong, since the opposite of "hierarchy" is "egalitarianism". Indeed, Schwartz (2006:173) arrives at the conclusion empirically that "[p]eople are more likely to view competition as good if they live in countries with cultures that emphasize hierarchy and mastery". In

[76] Although Gleaser at al. (2004) criticize three institutional measures, none of which is identical with the index I use above, their criticism applies. They note, for example that "[w]hatever expropriation risk measures, it is obviously not permanent rules, procedures or norms supplying checks and balances on the sovereign" (ibid:276) referring to data from the International Country Risk Guide. Criticizing the usually used "Good Governance" data they claim that "[t]hese are clear *ex post* outcomes, highly correlated with the level of economic development, rather than political constraints *per se*" (ibid:276, emphases in original). Finally, on the Polity IV's "constraints on the executive" data they comment that "it is an outcome measure, which reflects not the constraints, but what happened in the last elections" (ibid:277).

addition, his results (ibid:172, Table 3) also show that hierarchy is the most important value that predicts an attitude that emphasizes hard work as a "desirable quality for children to learn". Consequently, the significance of this value dimension for the diffusion of technology seems to support the theoretical argument.

Table 6a

The effect of the expropriation rate and culture in technology diffusion (pooled OLS)

	Dependent variable: growth rate of						
	ships	railways-pass.	railways-freight	cars	trucks	aviation-pass.	aviation-freight
constant	-0.790 (-4.11)***	-0.075 (-0.40)	-0.063 (-0.27)	-0.196 (-0.96)	-0.430 (-2.95)***	-0.268 (-1.71)*	-0.243 (-0.75)
initial value	-0.018 (-2.32)**	-0.002 (0.42)	0.001 (0.61)	-0.007 (-1.91)*	-0.002 (-0.89)	-0.013 (-4.01)***	-0.021 (-3.50)***
economic freedom	0.060 (2.71)**	0.040 (1.74)*	0.007 (0.26)	-0.003 (-0.09)	0.013 (0.65)	0.105 (3.68)***	0.145 (2.24)**
embedded-ness	0.313 (4.73)***	0.004 (0.05)	0.076 (0.82)	0.037 (0.56)	0.164 (3.05)***	0.078 (1.16)	0.077 (0.64)
harmony	0.242 (2.58)**	0.004 (0.06)	-0.023 (-0.32)	0.152 (1.80)*	0.154 (2.79)***	0.069 (1.23)	-0.022 (-0.17)
hierarchy	0.074 (2.30)**	0.031 (0.95)	-0.027 (-0.73)	0.108 (3.58)***	0.055 (2.71)***	0.099 (3.05)***	0.171 (3.17)***
no. of obs.	86	128	133	206	176	143	142
adj. R^2	0.303	-0.024	-0.012	0.094	0.121	0.206	0.203

Table 6b

The effect of the expropriation rate and culture in technology diffusion (pooled OLS)

	Dependent variable: growth rate of						
	telephone	cell phone	pers. computers	internet users	MRIs	Blast ox. steel	electricity
constant	-0.282 (-1.35)	2.255 (1.87)*	1.182 (2.69)***	1.478 (0.66)	0.458 (0.34)	0.404 (0.46)	-0.182 (-1.50)
initial value	-0.005 (-1.20)	-0.092 (-3.59)***	-0.040 (-4.84)***	-0.120 (-6.50)***	-0.071 (-6.25)***	-0.046 (-1.25)	-0.006 (-3.74)***
economic freedom	0.023 (0.81)	-0.882 (-4.17)***	-0.048 (-0.66)	0.092 (0.27)	0.543 (2.70)**	-0.039 (-0.61)	0.030 (2.16)**
embedded-ness	0.070 (0.63)	-0.371 (-0.72)	-0.151 (-1.21)	0.181 (0.24)	-0.660 (-1.88)*	-0.280 (-0.78)	0.106 (2.25)**
harmony	0.129 (1.96)*	0.824 (1.50)	-0.167 (-0.90)	-0.115 (-0.15)	-0.337 (-0.74)	0.211 (0.70)	0.094 (2.00)*
hierarchy	0.144 (2.68)***	0.741 (3.09)***	0.157 (1.98)*	0.655 (1.77)*	0.383 (1.47)	0.194 (1.49)	0.060 (3.66)***
no. of obs.	143	93	87	64	25	151	225
adj. R^2	0.290	0.472	0.552	0.516	0.261	0.134	0.252

Notes: all variables are included in log forms. Standard errors are clustered. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level

Tables 7a,b show the results when the second strategy is applied: the second area (legal structure and the security of property rights) of the EFW index and the mean of the remaining four are included separately as independent variables. The results derived from running fixed effect panel regressions show that it is this remaining element of the economic freedom index which is significant, if any. It must be added, however, that this happens in only six instances of the possible fourteen. There is no one case when the property rights variable is significant at the 5 percent level.

Table 7a

The effect of exogenous and endogenous change in the expropriation rate (fixed effects panel regressions)

	Dependent variable: growth rate of						
	ships	railways-pass.	railways-freight	cars	trucks	aviation-pass.	aviation-freight
constant	0.523 (3.64)***	1.640 (4.55)***	0.694 (2.63)**	0.736 (3.35)***	0.116 (1.67)	0.758 (6.05)***	0.389 (2.81)***
initial value	-0.092 (-4.15)***	-0.189 (-4.66)***	-0.100 (-3.87)***	-0.061 (-3.98)***	-0.044 (-9.71)***	-0.084 (-9.18)***	-0.092 (-6.76)***
property rights	-0.007 (-0.51)	-0.003 (-0.26)	0.003 (0.24)	0.027 (-1.19)	-0.005 (-0.29)	0.017 (0.75)	-0.008 (-0.25)
EFW excl. p. r.	0.039 (0.88)	-0.036 (-0.59)	0.021 (0.32)	0.110 (5.28)***	0.110 (3.77)***	-0.015 (-0.25)	0.087 (1.10)
no. of obs.	118	165	166	313	273	204	198
no. of countries	39	60	56	91	78	72	71
R^2 (within)	0.452	0.557	0.266	0.218	0.170	0.487	0.449
R^2 (between)	0.001	0.007	0.007	0.000	0.021	0.018	0.003

Table 7b

The effect of exogenous and endogenous change in the expropriation rate (fixed effects panel regressions)

	Dependent variable: growth rate of						
	telephone	cell phone	pers. computers	internet users	MRIs	Blast ox. steel	electricity
constant	0.269 (2.27)**	0.037 (0.04)	0.966 (3.48)***	-5.429 (-3.08)***	-1.572 (-1.14)	0.725 (7.76)***	1.261 (3.42)***
initial value	-0.024 (-2.64)***	-0.112 (-4.13)***	-0.086 (-7.57)***	-0.150 (-2.81)***	-0.101 (-6.52)***	-0.106 (-7.50)	-0.057 (-3.34)***
property rights	0.007 (0.56)	0.183 (0.75)	0.016 (1.07)	-0.367 (-0.64)	0.551 (1.36)	0.017 (1.08)	0.003 (0.25)
EFW excl. p. r.	0.066 (1.77)*	0.923 (1.93)*	0.193 (1.07)	4.700 (3.02)***	0.445 (1.46)	0.069 (0.092)	0.077 (4.01)***
no. of obs.	240	159	138	101	30	194	370
no. of countries	91	90	80	82	16	47	98
R^2 (within)	0.053	0.352	0.534	0.543	0.441	0.511	0.352
R^2 (between)	0.001	0.014	0.382	0.038	0.010	0.009	0.071

Notes: all variables are included in log forms. Standard errors are clustered. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level

Tables 8a,b show the results when the two strategies are combined: both parts of the EFW index and the cultural variables are included. The economic freedom variable performs much worse: there are only five cases when any of its two parts has a significant (at the 5 percent level) coefficient and, with one exception (MRI's), it is the non-property rights part that does. However, in one case (cell phones) the coefficient has a negative sign. The effect of culture shows about the same pattern as in Tables 6a, b: again, the "hierarchy" variables prove to be most important, being positively significant in ten cases.

All in all, these results lend some support to the predictions derived from the theoretical argument presented in the paper, inasmuch they suggest that a low rate of expropriation shaped by informal institutions will not contribute to the spread of one particular technology.

Rather, it is culture directly, or a low expropriation rate that is enforced in a top-down fashion, that will.

Table 8a

Endogenous change, exogenous change and culture (pooled OLS)

	Dependent variable: growth rate of						
	ships	railways-pass.	railways-freight	cars	trucks	aviation-pass.	aviation-freight
constant	-0.717 (-5.41)***	-0.136 (-0.78)	-0.040 (-0.19)	-0.209 (-0.98)	-0.333 (-1.93)*	-0.385 (-2.38)**	-0.326 (-0.95)
initial value	-0.018 (-2.28)**	-0.003 (-0.61)	0.001 (0.62)	-0.006 (-1.56)	-0.003 (-1.07)	-0.014 (-3.51)***	-0.026 (-3.76)***
property rights	-0.030 (-2.39)**	0.004 (0.22)	0.004 (0.17)	0.000 (0.00)	-0.021 (-1.36)	0.020 (1.08)	0.024 (0.71)
EFW excl. p. r.	0.085 (6.20)***	0.031 (0.99)	-0.004 (-0.10)	0.004 (0.15)	0.032 (1.92)*	0.093 (4.12)***	0.088 (1.91)*
embedded-ness	0.260 (4.40)***	0.021 (0.28)	0.084 (0.99)	0.055 (0.79)	0.118 (1.92)*	0.129 (1.98)*	0.106 (0.78)
harmony	0.241 (3.26)***	0.047 (0.70)	-0.032 (-0.44)	0.130 (1.50)	0.137 (2.15)**	0.104 (1.75)*	0.058 (0.44)
hierarchy	0.079 (2.96)***	0.030 (0.79)	-0.037 (-0.92)	0.100 (3.32)***	0.053 (2.25)**	0.092 (3.65)***	0.202 (3.05)***
no. of obs.	78	112	120	193	163	128	126
R^2	0.408	0.015	0.028	0.101	0.160	0.267	0.305

Table 8b

Endogenous change, exogenous change and culture (pooled OLS)

	Dependent variable: growth rate of						
	telephone	cell phone	pers. computers	internet users	MRIs	Blast ox. steel	electricity
constant	-0.222 (-0.98)	1.119 (0.84)	0.999 (2.10)**	2.26 (0.90)	-0.667 (-0.43)	-0.375 (-1.21)	-0.143 (-1.05)
initial value	-0.007 (-1.52)	-0.088 (-3.65)***	-0.041 (-4.79)***	-0.121 (-6.20)***	-0.055 (-4.70)***	-0.007 (-2.02)*	-0.006 (-3.51)***
property rights	0.011 (0.84)	0.134 (0.47)	0.050 (1.00)	-0.218 (-0.85)	0.544 (2.53)**	-0.008 (-0.35)	-0.012 (-1.17)
EFW excl. p. r.	-0.001 (-0.02)	-0.911 (-2.89)***	-0.071 (-1.24)	0.262 (0.82)	0.120 (0.85)	0.012 (0.40)	0.039 (2.18)**
embedded-ness	0.054 (0.41)	0.073 (0.12)	-0.074 (-0.53)	-0.130 (-0.16)	-0.453 (-1.18)	0.058 (0.85)	0.092 (1.86)*
harmony	0.129 (1.94)*	0.945 (1.73)*	-0.150 (-0.81)	-0.232 (-0.28)	-0.054 (-0.12)	0.214 (1.71)*	0.090 (1.73)*
hierarchy	0.167 (2.41)***	0.858 (3.05)***	0.177 (2.15)***	0.556 (1.32)	0.473 (2.46)***	0.098 (2.61)**	0.055 (3.23)***
no. of obs.	126	93	87	64	25	146	206
R^2	0.350	0.513	0.585	0.562	0.473	0.121	0.271

Notes: all variables are included in log forms. Standard errors are clustered. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level

It would be a mistake to conclude then that a low expropriation risk that is culturally embedded through an acceptance of the ideology of freedom is not important in economic development. One should not forget that the results concern the spread of a certain technology, which does not say much about how these measures are translated into human value. And this might be a crucial insight when interpreting the effect of those "good" institutions that need to be culturally embedded. These results suggest that their effect does not lie in enhancing the spread of one certain technology. Instead, their effect may be to

inspire the discovery of new technologies and new ways of making value out of old technologies[77].

5.3. Some robustness checks

Table 9 and 10a,b show the results of two kinds of robustness checks. First, I included time dummies and two education variables in the panel regressions presented in Tables 7a,b. Time dummies seem to be useful to account for worldwide trends in technology-adoption, such as the spread of information and communication technologies. The importance of educational variables for technology diffusion is shown in Comin and Hobijn (2004). Here I use the average years of primary and secondary education from Barro and Lee (2010).

Table 9

Robustness: time dummies and schooling (fixed effects panel regressions)

	Dependent variable: growth rate of					
	cars	trucks	aviation-pass.	aviation-freight	MRIs	electricity
constant	0.650 (3.08)***	0.400 (6.12)	1.33 (7.94)	0.625 (3.91)***	-4.986 (1.30)	2.372 (2.90)***
initial value	-0.049 (-4.22)***	-0.090 (-10.81)***	-0.166 (-9.84)***	-0.157 (-6.39)***	-0.410 (-4.19)***	-0.105 (-2.93)***
property rights	-0.061 (-1.72)*	-0.017 (-0.92)	0.012 (0.60)	-0.019 (-0.55)	1.721 (2.51)**	0.002 (0.15)
EFW excl. p. r.	0.107 (3.74)***	0.094 (3.86)***	0.077 (2.27)**	0.186 (2.83)***	0.940 (1.90)*	0.050 (2.88)***
average years of primary educatio)	-0.004 (-0.39)	-0.021 (-1.17)	0.007 (0.60)	-0.012 (-0.70)	0.630 (0.39)	-0.004 (-0.46)
average years of secondary education	0.007 (0.89)	0.024 (1.92)*	-0.002 (-0.25)	0.016 (1.40)	-0.011 (-0.05)	0.010 (1.00)
no. of obs.	290	253	184	178	30	344
no. of countries	88	77	70	69	16	95
R^2 (within)	0.213	0.318	0.695	0.567	0.911	0.479
R^2 (between)	0.002	0.022	0.018	0.002	0.007	0.060

Notes: all variables are included in log forms. The coefficients of the time dummies are not shown. Standard errors are clustered. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level

In Table 9 only those cases are included in which any of the institutional variables in questions were statistically significant at the ten percent level. As can be seen, there are six such cases, in only one of which is the coefficient of the property rights variable statistically significant. The main conclusion is not changed by adding these new variables: when better institutions matter for technology diffusions, it is the non-property rights institutions which, as I argued above, are the ones that are not greatly determined by informal, cultural, or ideological factors.

[77] An example of the explicit test of a specific form of this hypothetical conclusion is Leeuwen and Földvári (2011) who show that the fall of the communist system in Hungary brought an increase in the value of human capital, the inclusion of which into a growth accounting framework makes TFP growth close to zero or even negative.

Table 10a:

Robustness: pooled regressions with cultural region dummies and time dummies

	Dependent variable: growth rate of				
	railways-pass.	railways-freight	cars	trucks	aviation-pass.
constant	0.134 (1.82)*	0.053 (0.76)	0.189 (1.71)*	0.262 (3.19)***	0.121 (1.17)
initial value	-0.005 (-1.06)	0.001 (0.31)	-0.006 (-1.61)	-0.002 (-0.68)	-0.014 (-1.49)
economic freedom	-0.049 (-1.52)	-0.032 (-0.87)	-0.044 (-1.07)	-0.098 (-2.70)***	0.065 (1.44)
EFW×ES	0.083 (0.56)	0.122 (1.08)	0.187 (3.07)***	0.209 (3.38)***	0.072 (0.47)
EFW×EE	0.074 (1.07)	0.116 (1.08)	-0.166 (-0.73)	0.192 (2.24)**	0.055 (3.64)***
EFW×AF	0.054 (0.88)	0.077 (1.16)	0.041 (0.83)	0.086 (1.85)*	-0.037 (-0.38)
EFW×LA	0.120 (1.34)	0.191 (3.65)***	0.090 (1.93)*	0.133 (2.78)***	-0.012 (-0.18)
EFW×CO	-0.131 (-4.01)***	-0.189 (-2.26)**	-0.186 (-2.67)***		
EFW×SA	0.096 (0.094)	0.132 (1.19)	0.000 (0.00)	0.020 (0.34)	0.014 (0.21)
no. of obs.	196	182	352	313	238
no. of countries	66	61	100	87	82
R^2	0.088	0.142	0.224	0.178	0.298

Notes: see Table 10b.

As a second robustness check of the results above, I will use the idea of cultural regions developed in Schwartz (2004, 2006) and Licht et al (2007). In accordance with what was said in Section 5.2 I adjusted their database and included countries that were not originally included.[78] In addition to the reason provided above, making use of regions can also be considered a credible option because several theoretical approaches come to identify roughly the same cultural regions (Schwartz 2006:157). In what follows I will use dummy variables for seven cultural regions developed in Schwartz (2004, 2006) and include these dummies into the panel regressions run in the similar fashion as they were before. These regions are West Europe, English Speaking, Confucian, Africa and Middle East, South Asia, East Europe, Latin America

Again I selected only those cases in which any one of the institutional variables in question has a statistically significant coefficient (at the 5 percent level). The regressions are run with the cultural region dummies (not shown), the two education variables mentioned above (not shown), and interaction variables between these dummies and the economic freedom variables. The idea is to see whether the effect of economic freedom is different across different cultural regions. The answer is that in nine cases at least one region differs. When they do, non Western European regions tend to have a higher coefficient on the interaction term with the notable exception of the Confucian region[79]. This means that a change in economic freedom usually has a different effect on the spread of these technologies than they do in Western Europe and usually a higher economic freedom will lead to a faster diffusion than in Western Europe.

[78] The cultural regions that Licht et al. (2007) and Schwartz (2004, 2006) use are not precisely the same. Particularly, Licht et al. (2007) merge Schwartz's (2004, 2006) "Confucian" and "South Asia" regions into one called Far East, while they use an "Africa" region instead of Schwartz's "Africa and the Middle East"

[79] Running the regressions without China does not change this result.

Table 10b

Robustness: pooled regressions with cultural region dummies and time dummies

	Dependent variable: growth rate of				
	aviation-freight	telephone	computer	steel	electricity
constant	0.070 (0.44)	0.146 (2.23)**	0.529 (3.73)***	0.222 (0.76)	0.361 (2.65)**
initial value	-0.019 (-2.18)**	0.002 (0.65)	-0.042 (-6.34)***	-0.048 (-1.53)	-0.006 (-1.24)
economic freedom	0.052 (0.67)	-0.068 (-2.85)**	0.122 (1.79)*	0.210 (1.42)	-0.086 (-1.72)*
EFW×ES	0.046 (0.31)	-0.069 (-1.22)	0.213 (0.51)	0.226 (1.08)	0.186 (2.90)***
EFW×EE	0.270 (8.45)***	0.253 (2.87)***	-0.031 (-0.16)	-0.420 (-1.26)	0.147 (2.06)**
EFW×AF	-0.016 (-0.14)	0.164 (2.66)**	-0.042 (-0.37)	-0.120 (-0.68)	0.057 (1.01)
EFW×LA	0.085 (0.58)	0.159 (3.46)***	-0.251 (-1.42)	-0.100 (-0.55)	0.112 (2.17)**
EFW×CO				-0.371 (-2.67)**	-0.023 (-0.38)
EFW×SA	0.245 (2.89)***	0.140 (1.12)	-0.181 (-1.75)*	-0.804 (-1.99)	0.116 (1.99)**
no. of obs.	230	291	139	202	421
no. of countries	79	103	83	48	108
R^2	0.299	0.295	0.580	0.319	0.256

Notes: all variables are included in log forms except for region dummies. The coefficients of the time and cultural region dummies and of the schooling variables are not shown. Israel is dropped because Licht et al. (2007) do not associate it with any of the cultural regions. Standard errors are clustered. Heteroskedasticity robust t-statistics are in parentheses. Letters in the upper index refer to significance: ***: significance at 1 percent, **: 5 percent, *: 10 percent. T-values without an index mean that the coefficient is not significant even at the 10 percent level

6. Conclusion

The story of culture and economic growth in this paper is based on the strong proposition that there is an ideology of freedom which can be identified as the informal basis for economic growth. This ideology, I claim, can be formulated from the works of economic historians dealing with the era of the industrial revolution and from the thoughts of classical liberal thinkers. Another strong proposition of the paper is that a useful simplification of such an ideology is to see it as an attitude towards rent seeking which includes of course the ability to differentiate between productive and unproductive activities.

Applying this view of informal institutions in a very simple model of technological adoption leads to some results which can be seen as answers to the puzzle raised in the introduction: although property rights security is seen as the most fundamental factor of economic growth, leading economic historians do not identify its increase at the time when the era of economic growth began. According to the model, we have to differentiate between culture, formal institutions and property rights security (risk of expropriation): culture has an effect on technology adoption through its sub-constitutional role by providing an incentive for productive activities, and through its constitutional role by choosing the degree to which formal institutions should prevent rent seeking. Property rights security is the end result of the process framed within these incentives. With the help of this model I argued that a cultural turn-away from rent seeking may not abate the risk of expropriation. In addition, since cultural change has a direct effect, it will affect the spread of technology without a change in the risk of expropriation, even if this risk is supposed to depend on the cultural variables. The regression results gave some support to this claim.

What is important to bear in mind is that this explanation is concerned with the spread of one particular technology that has already been invented. The lack of significance of a culturally-backed low expropriation risk may add something to what we know about the role of institutions in economic development. It seems that a low expropriation risk based on people's beliefs (and not on top-down enforcement) is more important in discovering new technologies than in the diffusion of old ones.

As a result, the paper may have provided some (unintended) contribution to the "unbundling institutions" issue by pointing out how different parts of the institutional mix that constitute economic freedom differ in their contributions to economic development. I argued that while the security of property rights is an important determinant of innovation, the rest of economic freedom has more to do with the diffusion of already existing technologies. The paper thus suggests a strategy to unbundle institutions that is different from what is common in the literature. This strategy is to distinguish between that aspect of institutions that is determined by culture and that which is not.

Appendix

A1. Conditions for an internal solution of the model in section 4.1.

By an internal solution I mean a solution of the model described by equations (6), (9) and (10) which satisfies the conditions

$$0 < L_x < 1,\ 0 < L_r < 1,\ 0 < L_a = 1 - L_r + L_x < 1. \tag{A1}$$

Using the solutions that are given in equations (11) and (12) these conditions are satisfied if

$$\gamma\mu > \max\left\{\frac{1}{(1+\delta)\theta}, \frac{\theta}{1-\theta}\frac{1}{1+\delta} + \frac{1}{(1+\delta)^2}\frac{1}{\theta}\frac{1-\alpha}{\alpha}\right\} \text{ and } \delta < \frac{1-\alpha}{\alpha} - 1 \text{ or} \tag{A2}$$

$$\frac{\theta}{1-\theta}\frac{\alpha}{\alpha(1+\delta)-(1-\alpha)} > \gamma\mu > \max\left\{\frac{1}{(1+\delta)\theta}, \frac{\theta}{1-\theta}\frac{1}{1+\delta} + \frac{1}{(1+\delta)^2}\frac{1}{\theta}\frac{1-\alpha}{\alpha}\right\}$$

$$\text{and } \delta \geq \frac{1-\alpha}{\alpha} - 1. \tag{A3}$$

A2. The case of $\theta = 1$

In the case of $\theta = 1$, the rent-seeking technology (equation (3)) becomes

$$t(L_r) = L_r. \tag{A4}$$

This leads to a modification of three equilibrium conditions of the model (equations (6), (9), (10)):

$$\frac{1-\alpha}{\alpha} L_x = 1 + \delta \tag{A5}$$

$$(1 - L_r)\gamma\mu\frac{1-\alpha}{\alpha} L_x = 1 \tag{A6}$$

$$L_x + L_a + L_r = 1. \tag{A7}$$

Solving for L_x, and L_r gives

$$L_x = (1+\delta)\frac{\alpha}{1-\alpha} \tag{A8}$$

$$L_r = 1 - \frac{1}{\gamma\mu(1+\delta)}. \tag{A9}$$

These interior solutions apply if

$$\gamma\mu > \max\left\{\frac{1}{(1+\delta)}, \frac{1}{(1+\delta)^2\frac{\alpha}{1-\alpha} + 1 + \delta}\right\} \text{ and } -1 < \delta < -1 + \frac{1-\alpha}{\alpha}. \tag{A10}$$

A3. The effect of culture on the expropriation rate

Using equation (20) it can be shown that

$$\frac{\partial t}{\partial \theta} = \frac{(1-t)\gamma\mu(1+\delta) - t}{[\gamma\mu(1+\delta)+1]\theta - 1}$$

which is positive if

$$\delta > \frac{1}{\gamma\mu} \times \max\left(\frac{t}{1-t}, \frac{1-\theta}{\theta}\right) - 1 \text{ or } \delta < \frac{1}{\gamma\mu} \times \min\left(\frac{t}{1-t}, \frac{1-\theta}{\theta}\right) - 1.$$

Similarly,

$$\frac{\partial t}{\partial \delta} = \frac{\gamma\mu\theta(1-t)}{[\gamma\mu(1+\delta)+1]\theta - 1} \text{ which is positive if } \delta > \frac{1}{\gamma\mu}\frac{1-\theta}{\theta} - 1.$$

Table A1

Categories of technology and the measures describing them

Technologies	Technology measures
transportation	**Steam and motor ships:** Gross tonnage (*above a minimum weight*) of steam and motor ships in use at midyear. *Invention year:* 1788; the year the first (US) patent was issued for a steam boat design.
	Railways–Passengers: Passenger journeys by railway in passenger-kilometres. *Invention year:* 1825; the year of the first regularly scheduled railroad service to carry both goods and passengers.
	Railways–Freight: Metric tons of freight carried on railways (*excluding livestock and passenger baggage*). *Invention year:* 1825; same as passenger railways.
	Cars: Number of passenger cars (*excluding tractors and similar vehicles*) in use. *Invention year:* 1885; the year Gottlieb Daimler built the first vehicle powered by an internal combustion engine.
	Trucks: Number of commercial vehicles, typically including buses and taxis (*excluding tractors and similar vehicles*), in use. *Invention year:* 1885; same as cars.
	Aviation–Passengers: Civil aviation passenger-kilometres travelled on scheduled services by companies registered in the country concerned. *Invention year:* 1903; the year the Wright brothers managed the first successful flight.
	Aviation–Freight: Civil aviation ton-kilometers of cargo carried on scheduled services by companies registered in the country concerned. *Invention year:* 1903; same as aviation–passengers.
telecommunications	**Telephone:** Number of telegrams sent. *Invention year:* 1876; year of invention of telephone by Alexander Graham Bell.
	Cell phone: Number of users of portable cell phones. *Invention year:* 1973; first call from a portable cell phone.
IT	**Personal computers:** Number of self-contained computers designed for use by one person. *Invention year:* 1973; first computer based on a microprocessor.
	Internet users: Number of people with access to the worldwide network. *Invention year:* 1983; introduction of TCP/IP protocol.
medical	**MRIs:** Number of magnetic resonance imaging (*MRI*) units in place. *Invention year:* 1977; first MRI-scanner built.
steel	**Blast Oxygen Steel:** Crude steel production (*in metric tons*) in blast oxygen furnaces (a process that replaced Bessemer and OHF processes). *Invention year:* 1950; invention of blast oxygen furnace.
electricity	**Electricity:** Gross output of electric energy (*inclusive of electricity consumed in power stations*) in KwHr. *Invention year:* 1882; first commercial power station on Pearl Street in New York City.

Source: Comin and Hobijn (2010a:2043, 2055-256)

References

Acemoglu, D. (2009) Introduction to Modern Economic Growth. Princeton University Press, Princeton.

Acemoglu, D. – Johnson, S. – Robinson, J. A. (2005) Institutions as the Fundamental Cause of Long-Run Growth. In: Aghion, Ph. – Durlauf, S. N. (eds.) Handbook of Economic Growth, Volume 1A. Amsterdam: Elsevier :385-472.

Acemoglu, D. – Johnson, S. (2005) Unbundling Institutions. Journal of Political Economy 113. 5: 949-995.

Aghion, Ph. – Howitt, P. (1998) Endogenous Growth Theory. The MIT Press, Cambridge, MA.

Aghion, Ph. – Howitt, P. (2005) Growth with Quality-improving Innovations: An Integrated Framework. In: Aghion, Ph. – Durlauf, S. N. (2005) Handbook of Economic Growth, Volume 1A. Amsterdam: Elsevier, North Holland. pp. 69-110.

Algan, Y. – Cahuc, P. (2010) Inherited Trust and Growth. American Economic Review 100.5:2060-92.

Arruñada, B. (2010) Protestants and Catholics: Similar Work Ethic, Different Social Ethic. Economic Journal 120. 547:890-918

Barro, R. J. – Lee, J.-W. (2010) A New Dataset On Educational Attainment in the World, 1950-2010. NBER Working Paper 15902.

Baumol, W. J. (1990) Entrepreneurship: Productive, Unproductive, Destructive. Journal of Political Economy 98. 5:893-921.

Benson, B. L. (1984) Rent Seeking from a Property Rights Perspective. Southern Economic Journal 51. 2:388-400.

Boettke, P. J. (2001) Why Culture Matters? In: Boettke, P. J. (2001) Calculation and Coordination: Essays on Socialism and Transitional Political Economy. Routledge, New York. pp. 248-265.

Buchanan, J. M. (1980) Rent Seeking and Profit Seeking. In: Buchanan J. M. – Tollison, R. D. – Tullock, G. (eds.) Toward a Theory of a Rent-seeking Society. Texas A&M University Press, College Station. pp. 3-15.

Campbell, A. (2010) If FIFA were a country, it would be Russia. The Daily Telegraph, Online edition, 3 December. http://www.telegraph.co.uk/news/newstopics/politics/8180427/If-Fifa-were-a-country-it-would-be-Russia.html

Clark, G. (2007) A Farewell To Alms. A Brief Economic History of the World. Princeton, N.J.: Princeton University Press.

Comin, D. – Easterly, W. – Gong, E. (2010) Was the Wealth of Nations Determined in 1000 BC? American Economic Journal: Macroeconomics. 2.3:65-97.

Comin, D. – Hobijn, B. – Rovito, E. (2006) Five Facts You Need to Know About Technology Diffusion. NBER Working Paper 11928.

Comin, D. – Hobijn, B. – Rovito, E. (2008) Technology Usage Lags. Journal of Economic Growth 13. 4:237-256.

Comin, D. – Hobijn, B. (2004) Cross-country technology adoption: making the theories face the facts. Journal of Monetary Economics 51:39-83.

Comin, D. – Hobijn, B. (2010a) An Exploration of Technology Diffusion. American Economic Review 100. 5:2031-2059.

Comin, D. – Hobijn, B. (2010b) Technology Diffusion and Postwar Growth. NBER Working Paper 16378.

Crafts, N. F. R. (1995) Exogenous or Endogenous Growth? The Industrial Revolution Reconsidered. Journal of Economic History 55. 4:745-772.

De Soto, J. H. (2009) The Ethics of Capitalism. In: De Soto, J. H. (2009) The Theory of Dynamic Efficiency. London: Routledge. pp. 169-181.

Eichengreen, B. – Park, D. – Shin, K. (2011) When Fast Growing Economies Slow Down: International Evidence and Implications for China. NBER Working Paper 16919.
Facchini, F. (2002) Complex Individualism and Legitimacy of Absolute Property Rights. European Journal of Law and Economics 13. 1:35-46.
Facchini, F. (2010) Religion, Law, and Development: Islam and Christianity – Why Is It in the Occident and Not the Orient that Man Invented the Institutions of Freedom? European Journal of Law and Economics 29. 1:103-129.
Glaeser, E. L. – La Porta, R. – Lopez-De-Silanes, F. – Shleifer, A. (2004) Do Institutions Cause Growth? Journal of Economic Growth 9. 3:271-303.
Goldstone, J. A. (2000) The Rise of the West – or Not? A Revision of Socio-economic History. Sociological Theory 18. 2:175-194.
Goldstone, J. A. (2002) Efflorescenes and Economic Growth in World History: Rethinking the "Rise of the West" and the Industrial Revolution. Journal of World History 13. 4:323-389.
Gorodnichenko, Y. – Roland, G. (2010) Culture, Institutions, and the Wealth of Nations. Working Paper, University of California, Berkeley, September. http://www.econ.berkeley.edu/~ygorodni/gorrol_culture.pdf accessed: January 6, 2010.
Gorodnichenko, Y. – Roland, G. (2011) Which Dimensions of Culture Matter for Long-Run Growth? American Economic Review 101.3:492-498.
Grossman, H. I. (2002) "Make Us a King": Anarchy, Predation and the State. European Journal of Political Economy 18.1: 31–46.
Greif, A. (1994) Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies. Journal of Political Economy 102. 5:912-950.
Guiso, L. – Sapienza, P. – Zingales, L. (2006) Does Culture Affect Economic Outcomes? Journal of Economic Perspectives 20.2:23-48.
Gwartney, J. – Lawson, R. (2003) The Concept and Measurement of Economic Freedom. European Journal of Political Economy 19. 3:405-430.
Gwartney, J. D. – Lawson, R. (2010) Economic Freedom of the World. Annual Report 2010. Vancouver: The Fraser Institute.
Hayek, F. A. (1970) Epilogue. The Three Sources of Human Values. In: Law, Legislation and Liberty. Volume III: pp.153-176.
Hayek, F. A. (1971) Principles or Expediency? In: Hayek, F. A. – Hazlitt, H. – Read, L. E. – Velasco, G. R. – Harper, F. A. (eds.) Toward Liberty. Essays in Honor of Ludwig von Mises on the Occasion of his 90th Birthday. Volume I. Menlo Park: Institute for Humane Studies. pp. 29-45.
Heston, A. – Summers, R. – Aten, B. (2009) Penn World Table Version 6.3, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.
Leeuwen, B. van – Földvári, P. (2011) Capital Accumulation and Growth in Hungary, 1924-2006. Acta Oeconomica 61. 2:143-164.
Licht, A. (2008) Social Norms and the Law: Why Peoples Obey the Law. Review of Law and Economics 4.3:715-750.
Licht, A. N. – Goldschmidt, Ch. – Schwartz, Sh. H. (2007) Culture Rules: The Foundations of the Rule of Law and Other Norms of Governance. Journal of Comparative Economics 35.4:659-688.
McCloskey, D. N. (2006) Bourgeois Virtues: Ethics for an Age of Commerce. Chicago: The University of Chicago Press.
McCloskey, D. N. (2008) Bourgeois Towns: How Capitalism Became Virtuous, 1600-1776. Max Weber Lecture No. 2008/3. European University Institute, Max Weber Programme, Badia Fiesolanan, Italy.

McCloskey, D. N. (2010) Bourgeois Dignity. Why Economics Can't Explain the Modern World. Chicago: The University of Chicago Press.
Mises (1980[1953]) Why Read Adam Smith Today? In: Mises, L. von (1980) Economic Freedom and Interventionism. Online edition, The Ludwig von Mises Institute http://mises.org/efandi/ch24.asp
Ménard, C. – Shirley, M. (2005) (ed.) Handbook of New Institutional Economics. Dordrecht:Springer.
Mokyr, J. (2004) The Gifts of Athena: Historical Origins of the Knowledge Economy. Princeton, N.J: The Princeton University Press.
Mokyr, J. (2007) The European Enlightenment, the Industrial Revolution, and Modern Economic Growth. Max Weber Lecture, European University Institute, Florence, Italy.
Mokyr, J. (2008) The Institutional Origins of the Industrial Revolution. In: Helpman. E. (ed.) Institutions and Economic Performance. Cambridge, Massachusetts: Harvard University Press. pp. 64-119.
Mokyr, J. (2009) The Enlightened Economy. An Economic History of Britain 1700-1850. New Haeven: Yale University Press.
Mokyr, J. (2010) Entrepreneurship and the Industrial Revolution in Britain. In: Landes, D. S. – Mokyr, J. – Baumol, W.J. (eds.) The Invention of Enterprise. Entrepreneurship form Ancient Mesopotamia to Modern Times. Princeton: Princeton University Press. pp. 183-210.
Murphy, K. M. – Shleifer, A. – Vishny, R. W. (1993) Why Is Rent-Seeking so Costly to Growth. American Economic Review 83. 2:409-414.
Murrell, P. (2009) Design and Evolution in Institutional Development: The Insignificance of the English Bill of Rights. Working Paper, Department of Economics, University of Maryland, College Park, USA.
North, D. C. (1990) Institutions, Institutional Change, and Economic Performance. Cambridge: Cambridge University Press.
North, D. C. (2005) Understanding the Process of Economic Change. Princeton University Press, Princeton and London.
Owen, A. L. – Videra, J. – Davis, L. (2009) Do All Countries Follow the Same Growth Process? Journal of Economic Growth, 14. 4:265-286.
Parente, S. L. – Prescott, E. C. (2000) Barriers to Riches. The MIT Press, Cambridge, Massachusetts.
Phelps, E. – Zoega, G. (2009) Entrepreneurship, culture and openness. In: Audretsch, D. B. – Litan, R. E. – Strom, R. J. (eds.) Entrepreneurship and Openness. Theory and Evidence. Edward Elgar, Cheltenham, UK. pp. 101-130.
Phelps, E. (2006) Economic Culture and Economic Performance: What Light Is Shed on the Continent's Problem? "Perspectives on the Performance of the Continent's Economies" Conference of CESifo and Center on Capitalism and Society, Venice, 21-22 July 2006. http://www.columbia.edu/~esp2/Center3rdAnnualConf2006Sept27.pdf
Rodrik, D. – Subramanian, A. – Trebbi, F. (2004) Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. Journal of Economic Growth, 9.2:131-165.
Röpke, W. (1959) The Economic Necessity of Freedom. Modern Age 3.3:227-236.
Röpke, W. (1969[1933]) End of an Era. In: Röpke, W. (1969) Against the Tide. Chicago: Henry Regnery Company. pp. 79-97.
Schwartz, Sh. H. (2004) Mapping and Interpreting Cultural Differences around the World. In: Vinken, H. – Soeters, J. – Ester, P. (eds.) Comparing Cultures, Dimensions of Culture in a Comparative Perspective. Leiden, The Netherlands: Brill. pp. 43-73.
Schwartz, Sh. H. (2006) A Theory of Cultural Value Orientations: Explication and Applications. Comparative Sociology 5. 2-3:137-182.

Scruton, R. (2006) England, an Elegy. London: Continuum.
Sen, A. (1970) The Impossibility of a Paretian Liberal. Journal of Political Economy 28. 1:152-157.
Snowdon, B. (2006) The Enduring Elixir of Economic Growth. Xavier Sala-i-Martin on the wealth and poverty of nations. World Economics 7. 1:73-130.
Spolaore, E. – Wacziarg, R. (2009) The Diffusion of Development. Quarterly Journal of Economics, 124.2:469-529.
Stringham, E. P. (2011) Embracing Morals in Economics: the Role of Internal Moral Constraints in a Market Economy. Journal of Economic Behavior and Organization 78. 1-2:98-109.
Tabellini, G. (2010) Culture and Institutions: Economic Development in the Regions of Europe. Journal of the European Economic Association 8.4:677-716.
Tabellini, G. (2008) Institutions and Culture (Presidential Address). Journal of the European Economic Association 6.2-3:255-294.
Williamson, C. – Mathers, R. (2011a) Economic Freedom, Culture, and Growth. Public Choice 148.3:313-335
Williamson, C. – Mathers, R. (2011b) Cultural Context: Explaining the Productivity of Capitalism. Kyklos, 64. 2:231-252.
Zak, P. J. – Knack, S. (2001) Trust and Growth. The Economic Journal 111. 470:295-321.
Zanden, J. L. van (2009) The Long Road to the Industrial Revolution. The European Economy in a Global Perspective, 1000-1800. Leiden, The Netherlands: Brill.

Judit Kapás

The Factory: An *Historical* Theory of the Firm View

1. Introduction

One stylized fact about the factory is that it was the result of the British Industrial Revolution (BIR); another is that it was the first form of the capitalist firm. As a result of the rise of the factory, the *firm* as such became the dominant form of production.

Economic historians have always investigated the rise and the spread of the factory system, analyzing, in this way, the *origin* of the factory system in depth. However, by its nature this inevitably lacks an analysis of the essence and *nature* of the factory, which is a core question in the theory of the firm. The theory of the firm, however, due to the dominance of formal models explains *the* firm as such, and in this way is ahistorical.[80] There is no doubt that the first papers in the theory of the firm (e.g., Coase 1937) explained the nature of the firm irrespective of time and space, a perspective that later became dominant. The rare exceptions who put the firm in a historical context were Langlois (1999), Pitelis (1998) and Leijonhufvud (1986).

So, the two major questions, namely what caused the factory system to emerge (the *origin* question) and what is the nature or essence of the factory (the *nature* question) are asked and answered separately by economic history and the theory of the firm, irrespective of their organic relationship. The major argument of this paper is that the *origin* and the *nature* questions have to be answered simultaneously, which leads to a *historical* theory of the firm view for an understanding of the factory. The starting point of this perspective is the view that the capitalist firm is historically specific (Hodgson 2001); that is, it emerged in a concrete historical context – accordingly, an explanation for the factory needs a historical perspective. On the other hand, the *nature* of the capitalist firm, i.e., its distinctive characteristic, has to be emphasized in the explanation, too. Having said that, an *historical* theory of the firm perspective helps us not only to explain why *the* capitalist firm emerged, but also highlights its distinctive attribute at the same time.

A basis for this framework is a critical analysis, from the viewpoint of the theory of the firm, of the existing economic history views on the factory which attribute to the factory three major characteristics, namely centralization of production in one place, the use of machinery and factory discipline. Economic historians put one or more of the above characteristics at the centre of their argumentation concerning the rise of the factory. From the viewpoint of my approach, the main shortcoming of these views is that they do not identify the distinctive attribute of the factory, i.e., the characteristic that marks its basic difference compared to previous production forms, because the presence of the above three attributes do not necessarily imply firm-ness. This analysis relies basically on a comparison with the previously existing production form, namely the putting-out system (cottage industry).

To sum up, in this paper I approach the factory from a historical perspective, that is, through the process of its emergence in the BIR. This historical approach will lead me not only to the Coasean argument that the employment relationship (authority) is the essence of the factory (firm) but here the *why* will also be highlighted through historical events.

The paper is organized as follows. In section 2 I will show the historical background of the factory, that is, the system that prevailed before it: the putting-out system. The economic history views on the factory will be summarized and criticized in section 3. Section 4 will be devoted to a clarification of what the factory system is. In section 5 I will bring the *origin* and

[80] See Pitelis (1998) for a critique of transaction costs economics for its lack of a historical basis.

the *nature* questions together, developing the framework for a *historical* theory of the firm view on the factory. In section 6 I will provide a summary.

2. Historical background of the factory: the putting-out system

The dominant production system before the BIR was the cottage industry which operated mainly in the framework of the putting-out system.[81] The putting-out system was already known by the late Middle Ages, especially within the textile industry and it was the major organizational form for commercial production until the early 19th century.

In fact the putting-out system was a "big production organization" in the sense that the putter-out subcontracted with many artisans producing in their own homes, managing in this way a complex network of contracts of manufacture. The putter-out supplied raw material and capital, especially working capital, because fixed capital was insignificant at that time (Marglin 1974). The head of the domestic craftshop was the artisan who worked with family members and a couple of apprentices. The household was the unit of production and the work was divided between members of the family; so household and workplace were not separated (Mokyr 1993, 2002). The raw materials, the tools and the products were owned by the merchant – the putter-out, who outsourced the production to the artisans working at home and paid not a wage, but a price, for the products.[82] Accordingly, the putting-out system was a decentralized production system, supervised by the putter-out, in which the artisans were subcontractors. The putter-out performed both managerial and entrepreneurial roles since he or she supplied capital, supervised the output, and organized distribution and sales.

This system was of course adapted to different economic, social and cultural settings: it had prevailed for a long period in England, France, Germany and Italy. In particular settings the putter-out had only a subordinated role, while in others, where investments were important, he or she played the principal role.

From a theory of the firm perspective, the putting-out system should not be considered a firm, rather it consisted of market contracts, although these were long-term contracts.[83] The artisan was formally independent, he or she determined the production, and exchanged the product against raw material with the putter-out (Cohen 1981). The putter-out did not have authority over the artisans; accordingly, he or she could motivate the artisans only through prices, i.e., market coordinating mechanisms. The shortcomings of this system, in this way, derive from the absence of authority-based coordinating devices such as command. All this led to relatively high costs for the putter-out in monitoring the quality, and what is more, the monitoring of the production process itself was simply impossible. Embezzlement and fraud were common features of the putting-out system, which served to increase artisans' income. It covered different activities: pilfering of raw material for direct sales, embezzling of materials, working secretly with stolen materials and pilfering of finished goods. This was a kind of principal-agent problem. Later on, when the division of labor became more extended and the product more complicated, the putter-out had to face even higher agency costs, which loosened the system.

The factory system, however, did not arise overnight and the transition to the factory system was a slow process. In this process the putting-out and the factory system co-existed for a long period, and it took almost 100 years for the factory to become the dominant form of production (Mokyr 1993, 2001, 2002, Jones 1987). This means that the cottage industry and

[81] For details on putting-out see Landes (1966) and Magnusson (1991).
[82] Of course, there were slight differences as regards the practice of different merchants, but the common feature was that the producers worked up raw material put out by the putter-out and the product was taken care of by the putter-out. In some cases, however, artisans utilized their own tools (Magnusson 1991).
[83] Note also that the putting-out system had no organizational characteristics either (Kieser 1994).

the factory system were alternative trajectories; either of them could become more advantageous under particular circumstances.[84]

However, it must be noted that the BIR did not invent the factory system, the industrial revolution contributed only to its transformation and spread (Mokyr 2002). Here factory means a production unit (organization) involving many workers under one roof. But, as Geraghty (2003) argues, such plants existed in Britain largely before the BIR.[85] One category of such workplaces involved production processes too large or energy-intensive to be performed in a small shop or at home. These were fulling mills, glassworks, breweries, paper mills, and hammer forges. The other category of large-scale production units was the protofactory, an agglomeration of workers using more-or-less traditional hand technologies. The primary rationale for these centralized workplaces was organizational: direct supervision allowed improved quality control and a more intensive work pace. According to Geraghty (2003), the emergence of these early factories was not the result of the technological changes brought about by the BIR; the first factories appeared in textile, more specifically in carding and bleaching.

Factories satisfying the criteria of the capitalist firm[86] appeared first in textiles, in the silk industry. Thomas Lombe's silk mill, built in Derby in 1718 marks a radical departure from the typical pre-industrial factory by using a centralized power source. Jones (1987), analyzing in detail the development of the silk industry, shows convincingly that the spread of the factory system in silk followed technological breakthroughs. The first wave of factories was in silk throwing following the exploration of Thomas Lombe's patent in 1732. Weaving was slower to enter the factory: the Jacquard-loom invented in France was first used in England in about 1822, which resulted in the movement of silk weaving into the factory. To sum up, the transition to the factory system in the silk industry was relatively significant, but this process was never completed in the sense of including all production units (Jones 1987, 1999).

The spread of factories in other branches of the textile sector was less spectacular. First, a kind of mixed system developed, as in the cotton industry. Here some production activities were outsourced to small craftsmen working in their homes, while the rest of the production was organized within mills.[87] While the transition to the factory from domestic work was the most dramatic in textiles, it still took a century or more to complete (Mokyr 2002).[88]

Besides textiles, factories also penetrated other industries in England (Geraghty 2003). In the British iron industry two major innovations transformed both scale and organization: coke smelting and the puddling and rolling process. Coke smelting increased the capital-intensity and the minimal efficient scale in iron production. The puddling and rolling process extended the division of labor and permitted an almost continuous process. In metalworking the large-scale production and the use of steam power were commonplace. In the pottery industry Josiah Wedgwood pioneered large-scale factory production, and used innovative production techniques and steam power. In Britain by 1860 the progress of mechanization had led to a kind of industrial dualism in many industries.

3. Economic history views on the factory

[84] In fact, large-scale production and small craftshops represent two extreme poles as regards the organization of production. In many regions of Europe (Lyon, Sheffield, Northern Italy) networks in which small- and large-scale production were combined developed (Piore and Sabel 1984).

[85] See the Pollard's examples (1965).

[86] The capitalist firm is historically specific and has certain different characteristics compared to the firms of the previous era. These are: private ownership of assets and the employment relationship (see Hodgson 2001).

[87] Richard Arkwright's cotton spinning mills of the 1770s were the archetype of the early modern factory, and utilized sophisticated water power systems and a nearly continuous flow of materials (Geraghty 2003).

[88] In the Continent the emergence of factories lagged behind that in Britain (Geraghty 2003). In addition, the system itself, due to the different social-economic-cultural environment, was slightly different. As Kieser (1994) argues, in Germany for instance, as a result of poor education, there was a lack of a skilled labor force, which led to Germany's well-developed apprenticeship system, assisted by mill owners.

Since the rise of the factory was the result of the BIR, economic history has long devoted special attention to an explanation of the emergence of the factory, dealing, in this way, with the *origin* question: Why did the factory system arise when it did? Traditional economic history sees new technology brought about by the BIR as the core factor culminating in the rise of the factory (Mantoux 1961, Landes 1969, 1986, Mokyr 1990, 2001, 2002, Jones 1982, 1987): "The factory system was the necessary outcome of the rise of machinery" (Mantoux 1961:252).

Three major strands of explanation can be distinguished within the large literature at our disposal. One focuses on those factors that led to the centralization of production "under one roof" (Mokyr 1990). The second sees the cause of the rise of the factory in the use of machinery (Mantoux 1961, Landes 1986, Marx 1867). The third strand equates the rise of the factory with the appearance of factory discipline. In what follows I will present these theories, while also providing a criticism on the grounds of the neglect of the *nature* question.

3.1. (Centralization of) production "under one roof"

Mokyr (2001, 2002) is the leading scholar in arguing that technology, together with knowledge drove the emergence of the factory. According to him, one of the major novelties of the BIR was a huge expansion in the knowledge base of the techniques in use. This means that efficient production required more knowledge than a single household could possess. Due to the macroinventions of the BIR many of the industries increasingly required a level of knowledge and a set of operating procedures that were beyond the capacity of the individual household. Factories became the repository units for technical knowledge and reduced access costs to this knowledge for individual workers. In addition, factories could employ experts (engineers, mechanics, chemists) to assure the critical knowledge for production.

What is emphasized by Mokyr (2001) is not simply the new technique, but the changed character of the technique: the inventions of the BIR required new knowledge, basically mechanical-technical knowledge, which led to the separation of households and production: the location of production became a centralized place, namely the factory (mill). In this way, Mokyr believes that the rise of the factory is the result of the increased technical knowledge required by the inventions, and such knowledge was simply not available within households. Inside the plant agents knew and could trust each other, and this turned out to be an efficient way of sharing knowledge. As long as the minimum knowledge requirement was small, plants could be small and coincided with households (Mokyr 2001, 2002). When production required a wide knowledge base, many specialists had to be employed within one production unit, which extended the division of labor within the plant: everyone specialized in one task, increasing in this way the distributiveness of knowledge (see also Hayek 1945). The plant not only made the workers specialize, but it also coordinated the exchange of knowledge between them.

Accordingly, the whole issue of the emergence of the factory is reduced to the question of the physical location of production. In the theoretical framework of Mokyr, production units needed increasing internal specialization and a higher level of competence – which led to a better division of labor – because more knowledge was necessary to operate the best-practice technique.

So, in fact, Mokyr (2001) argues that technology affected not only the output, the income and well-being, but also the location where production took place: a major attribute of the BIR was the concentration of former artisans and domestic workers under one roof. As previously noted, according to him, this phenomenon was largely driven by technology. Mokyr also recognizes that large firms were quite widespread before the Industrial Revolution, but their employment was domestic labor (cottage industry), on a putting-out

basis. In this system the technology did not require that physical production be located in a central place; the workers were independent farmers or craftsmen.

As a leading factor behind the concentration of workers under one roof Mokyr (2001) highlights the change in the ratios of costs and benefits of moving information relative to that of moving people: due to cheaper transportation (railroads, roads, urbanization), it became relatively cheaper to move people (Sosztak 1989). So, the benefits of the concentration of production were related to the size and the complexity of the knowledge needed for production to take place (Mokyr 2001). It was simply more efficient to move specialized workers to the job than try to communicate all the necessary information through a decentralized production network.

The fact that the factory system centralized the production under one roof is, of course, an important aspect of the factory. However, the "under one roof" view is related only to the *origin* question; more precisely, it provides an answer – from a given perspective – to the question of why and how the factory emerged. But this perspective does not deal with the *nature* of the factory, and accordingly, cannot establish the nature of the factory. Clearly, the essence of the factory was not centralization under one roof. Even historical facts undermine this. As Landes (1969:14, 24) shows, production in many industries, including iron, chemicals and ship-building took place in a single location, but was not organized along "factory lines"; that is, although production was centralized, it was not organized as a firm (factory). As I mentioned above, Geraghty (2003) highlights that some production processes, such as fulling mills, glassworks, breweries, paper mills and hammer forges were too large or energy-intensive to be performed in small shop or in the home. So while location distinguished the putting-out system from the factory, it does not follow that centralization automatically implied factory organization (Cohen 1981): in many cases workers at these places were single artisans working with their own tools, being is this way independent (sub)contractors.

To conclude, centralization of production in one place is, of course, an important attribute of the factory, but cannot distinguish the factory from previous organizational modes of production.

3.2. The use of machinery: large-scale production

Another strand of the economic history literature also sees the rise of the factory as a wholly technical event by emphasizing the role of the use of machinery. In fact, mechanization increased fixed capital, and accordingly, by leading to increasing returns to scale, it increased the optimal scale of production (Mantoux 1961, Landes 1986).

So, the increase in the optimal scale of production was largely due to the use of the new machines brought about by the BIR.[89] As argued by Landes (1986), machines and new techniques meant gains in productivity and a shift in the relative importance of factors of production. He also emphasized that the logic of technology was towards even wider mechanization (ibid p. 615).

Of course, some equipment could not be made equally efficiently in small craftshops and in large plants (e.g., chemicals, iron making). Heating, lighting, power supply and security were all activities in which scale economics were the result of technical considerations. In addition, there were non-technical economies in scale, too, such as marketing or finance.

[89] "… what made the factory successful in Britain was not the wish, but the muscle: the machines and engines. We do not have factories until these were available, because nothing less would have overcome the cost advantages of dispersed manufacture (Landes 1986:607).

Besides these factors, according to Landes (1986), more extended specialization also contributed to the cost advantage of the factory as compared to the putting-out system.[90]

The cotton industry was the first to use machinery and became the example of modern large-scale industry (Mantoux 1961). Thomas Lombe's silk factory was the real beginning of the factory system in England, he was a precursor. Just to give an idea of the scale of his plant, let me recall here Mantoux's description of the factory: five hundred feet long, five or six storeys high and pierced by 460 windows, it employed 300 workers, used automatic tools and maintained continuous and unlimited production.

Economies in scale were important for the factory, of course, but they were not omnipotent and cannot fully explain the rise of the factory (Leijonhufvud 1986). As argued by Cohen (1981) the use of machinery in itself cannot explain the rise of the factory system; rather, the use of machinery contributed primarily to a decrease in production costs, and, accordingly, to the spread of the factory. Moreover, according to Pollard (1964), the role attributed to fixed capital is exaggerated: while in some industries the proportion of fixed capital as compared to working capital increased during the BIR, in a typical mill it was about 50% between 1780 and 1830, which is not so high (Pollard 1964:302). In fact, capital finance problems were much more related to the working capital, consequently it was not a factory-specific problem, but rather that of the putting-out system.[91]

The fact that large-scale production was not synonymous with the factory can be historically proven. As mentioned before, large-scale production units were present largely before the BIR: in even in 16th century England large cotton mills or mines operated whose size was not dictated by technology (machinery). Moreover, as argued by Mantoux (1961), large-scale production was almost exclusively artificial in France and it was supported by the French Crown. The best example of this was the Gobelins works. But the creation of these royal manufacturers in the 17th century must not be confused with the spontaneous growth of the factory in the following century in England. Mokyr (2002), when characterizing the putting-out system, clearly argues that, in terms of the size of production, it also was a large-scale production unit, since the merchant-entrepreneurs worked with a large number of artisans; and on the other hand, as argued by Landes (1969), there were many small-scale factories, as well.

Besides the above-mentioned historical facts which cast doubt on the equivalence between the factory and large-scale production, there is an even more fundamental problem, namely that it is very difficult to decide at what point machinery begins and tools end, since the factory did not arise overnight and the introduction of machinery was not accomplished at once; rather, as argued before, the transition to the factory system was a slow, gradual process, characterized by a mixed system; there was no clearly marked division between manufacture and the factory system.[92] But how can we distinguish manufacture from the modern factory? According to Marx (1867) the distinguishing feature of the factory is the use of machinery. However, machinery was used even in domestic production, such as the jennies and mules that were present almost exclusively in domestic production (Berg 1991).

To conclude, the use of machinery is a non-exclusive attribute of the factory; accordingly, it cannot establish the *nature* of the factory. However, it clearly refers to macroinventions of the BIR, establishing in this way an important relationship with historical facts.

3.3. Factory discipline

[90] The extent of specialization is well documented by the data of the Wedgwood porcelain plant: out of the 278 people that Wedgwood employed in June 1790, only five had no specified post, the rest were specialists (McKendrick 1961).

[91] Pollard (1964) also emphasizes that the finance of fixed capital was much more difficult, something which caused serious problems in utilities (roads, canals).

[92] According to Mantoux (1961), the use of machinery was only one of the principal factors, but probably the most fundamental one, in the modern factory system.

The third line of the economic history literature on the factory devotes attention to an aspect that brings us closer to the theory of the firm perspective. Rather than emphasizing technological aspects like the other two, it puts an organizational issue at the centre of the explanation, namely the nature of the work involved. The nature of work altered greatly during the BIR, as employees were subject to supervision, coordination and discipline (Geraghty 2003). As soon as fixed cost became important, the employer had an interest in supervising workers because shirking reduces the utilization rate of fixed capital.

Those scholars (e.g., Geraghty 2003, Pollard 1963, McKendrick 1961, Clark 1994) who share this perspective seem to admit that technology alone does not explain the rise of the factory system. This kind of explanation is related to organizational considerations: the factory arose to solve asymmetric information problems.[93]

In factories there was expensive capital equipment, interdependent production processes, and the need for improved quality. For these reasons, as explained by Geraghty (2003), owners introduced fixed working hours, punctuality and consistent attendance, high levels of work effort, an emphasis on the uniformity of finished product and proper care of the equipment. In smaller factories direct process supervision and face-to-face contact between supervisors and workers was sufficient to establish rudimentary levels of discipline. In larger factories where direct communication and control was not feasible factory discipline regimes were codified into work rules. Enforcement of discipline relied on deterrent mechanisms such as corporal punishment or the threat of dismissal (Pollard 1963). Later owners turned to positive incentives, including piece rate pay, or bonuses tied to productivity. Owners tried to engender loyalty by offering their workers various paternalistic fringe benefits (housing, sickness and accident insurance, pensions, medical care, educational facilities, etc.).

The factory discipline contributed to a large extent to a standardization in and improvement of product quality through the introduction of quality standards (Magnusson 1991), quality control (Cohen 1981) and the training of workers within the plant[94] (McKendrick 1961). This quality control contrasts sharply with the experience of earlier modes of production. Under the putting-out system quality was largely unobservable as direct process supervision was not possible.

To state it explicitly, supervision took two basic forms. Where a large number of skilled workers were used, owners relied on a subcontracting system. Master craftsmen were responsible for hiring, supervising, disciplining and paying their own workers. The masters were also often responsible for setting up and maintaining their own machinery and had the power to determine the pace of work. Where many workers were unskilled, shop-floor management was typically carried out by a group of foremen who formed a lower level of management.

This strand of the literature also suggests that the factory system was important not only as a way to centralize physical plant, but also as a means to centralize organizational decision-making concerning the aspects of production: the factory was a new type of organization in which the factory owner had more control over both technological and organizational issues.[95]

From a theory of the firm perspective, factory discipline, in fact, was monitoring. Quality standardization, punctuality, supervision, etc. served to overcome the information asymmetric problem present in the production process (see Alchian and Demsetz 1972). From the perspective of my concern here, an important advantage of this strand of economic history is that it takes the factory as a both technical and organizational unit, a view that comes closer to

[93] Note that some scholars from the theory of the firm school (e.g., Langlois 1999) offer a hybrid theory that incorporates elements of both the technology and organizational views.

[94] Mokyr (2002) points to the fact that factories changed the formation of human capital as well: the factory assumed a role in training its workers both within the factory and via a subsidy of schools.

[95] This view is given evidence by Gerarghty (2007) in his empirical analysis testing his complementary thesis.

the theory of the firm. However, it is still not clear what caused factory discipline and whether factory discipline marks the distinctive attribute of the factory.

4. What is a factory?

What exactly constituted a factory is difficult to establish based on the economic history literature. This is partly because the *nature* question is not at the forefront of this literature. However, when dealing with the rise of the factory even this literature sheds some light on what the distinguishing feature of the factory was. So, the term "factory" is not as unambiguous as it would seem to be at first glance. Basically two different meanings are present in the economic history literature, and these two are not always clearly separated from one other. One meaning is rather technical, the other is organizational. The views adopting the "under one roof" and the "use of machinery" arguments rely on a technological perspective, while completely neglecting the organizational aspect of the factory. The "factory discipline" view emphasizes organizational aspects, but does not refute its technological roots. How is the factory understood by these two views?

In the first view the essence of the factory is almost exclusively embodied in how production was organized within it: a central power source, machinery, continuous production, scale and efficiency (Jones 1999) etc. But even scholars within this branch of the literature admit that the factory was more than just a large production unit; it was rather a system of production in which the worker and the capitalist were bound by supervision and discipline (Landes 1986). Mantoux (1961) also argues that by the factory one means a particular organization, a particular system of production. Sombart (1902:26) tries to define the factory both by technical and economic characteristics. From the technical point of view he emphasizes the same points as Mokyr, namely that its main feature is the concentration of production in one establishment, with machinery moved by central force. From an economic point of view he points to the special relationship that existed between the capitalist and the worker: a kind of commanding power.

The above technical view of the factory must be augmented by the Smithian one, emphasizing the introduction of line production replacing the craft production of the putting-out system. He stressed that line production (Smith 1776) may allow an increase in the degree of the division of labor. In the Smithian story, however, the shift to the factory is not initially characterized by a new technology, but by the extension of market (demand), which led to the division of labor (Morroni 1992). Adam Smith, in his example of the pin factory, showed how an increase in demand may allow a reorganization of production, which brings about an increase in productivity and a shift from craft production to line production, typical of the factory system.

In contrast to the above views, Georgescu-Roegen (1970) clearly argues that the factory system is independent of technology. He explains that in a factory the economy of time reaches its maximum because line production allows a shift in workers and tools without interruption to the following process. He stresses that in every elementary process every agent is idle over some definite periods, and there is only one way to eliminate this idleness: the factory system. So, according to him, the root of the factory can be found in domestic workshops that introduced line production due to an increased demand. In his sense not every production activity can be turned into a factory. In Georgescu-Roegen's (1970) view the factory is a new type of organization of work, based on line production, which is different from the production characteristic of previous times.

The above views are in sharp contrast with the modern theory of the firm which identifies the nature of the firm in elements different from a simple association with production issues. Even the "factory discipline" view of economic history is much closer to the theory of the firm because it clearly recognizes that the factory was an organization. What is missing from

this view is a clear recognition of the fact that the factory was the first form of the capitalist firm. In this spirit Mantoux (1961) was right when saying that "[t]here was more difference between a spinning mill and a domestic workshop as they existed side by side between 1780 and 1800, than between a factory of that day and the modern one" (Mantoux 1961:251).

To sum up, what is missing from economic history is the recognition that the factory was the first form of the capitalist firm. Following Hodgson (2001, 2002) who argues that the firm is a historically specific institution, I propose to rely on his definition: "A firm is defined as an integrated and durable organization involving two or more people, acting openly or tacitly as a legal person, capable of owning assets, set up for the purpose of producing goods or services, with the capacity to sell or hire these goods or services to consumers" (Hodgson 2002:56). This definition points to two aspects of a firm: technological and organizational, and both are present within the factory

In what follows, when proposing a *historical* theory of the firm view on the factory, I take the factory as a *firm*. Accordingly, I augment the economic history views discussed and criticized above, by relying on what the theory of the firm says on the distinctive feature (nature) of the firm.

5. Taking the "origin" and the "nature" questions together

The stylized fact about the BIR, emphasized by economic history, is that it brought about the factory system. The one proposed by the theory of the firm is that it is the capitalist firm that came into existence with the emergence of the factory. These two stylized facts are stated by two different disciplines whose analyses are centered on different key questions. Economic history focuses on analyzing why the factory system arose (the *origin* question), while the theory of the firm is concerned with analyzing the essence of the factory (the *nature* question). The two disciplines basically developed in separation from one other. My argument is that the *nature* and the *origin* questions should be answered simultaneously. To arrive at such a perspective, the two disciplines must admit, at least partially, each other's characteristics.

As for the theory of the firm, except for – among others – Langlois (1999) and Pitelis (1998), it explains *the* firm without paying attention to historical specificities (Hodgson 2001). However, there is no doubt that the theory of the firm needs a historical background when it comes to the factory.[96]

Economic history, on the other hand, as shown by the above discussion, provides us with a detailed analysis of three aspects of the factory, namely the centralization of production under one roof, large-scale production and factory discipline. The major criticism vis-à-vis these views from a theory of the firm perspective is that they cannot reveal the distinctive attribute of the firm, the one that differentiates it fundamentally from other forms of production of previous times. Thus the factory had numerous common attributes of previous modes of production; accordingly, these cannot be considered distinctive ones. Table 1 summarizes the attributes of the factory as contrasted with those of the previous forms of production as is emphasized in economic history. So, the characterization of the factory, as featured in economic history, is an important analysis, but it cannot be equated with an explanation for the *nature* of the factory.

	Scale (workers)	**Technology**	**Work organization**
Domestic system	1-5	Simple hand tools	Subcontractor for merchants Limited division of labor Mostly family labor

[96] Kieser (1994) provides general arguments in favor of why the theory of the firm necessities historical analysis.

Protofactory	6-30	Simple hand tools, limited use of powered machinery	Supervision and discipline Female and child labor
Factory	30+	Powered machinery	Supervision and discipline Division of labor

Table 1: Historical organizational forms of production

That the above-mentioned three attributes of the factory are not *sine qua non* attributes of the factory can be shown by historical evidence. As for the centralized "production under one roof", as already argued above, it is not necessarily a firm-like organization. In many cases the workers used their own tools and worked as subcontractors in ship-building, iron and chemicals (Landes 1969:14, 24). Neither is large-scale production (due to the use of machinery) the distinctive attribute of the factory. There is plentiful evidence in the literature to prove that large-scale production was present in various non-firm organizations such as Gobelins in France or even in the putting-out system itself. On the other hand, some factories produced only on a small-scale.

The third strand of economic history literature sees factory discipline as an important aspect of the factory (Geraghty 2003, Pollard 1963, McKendrick 1961). Here it is argued that technology brought about the need for coordination between workers, which led to the introduction of factory discipline: supervision and regulations. However, factory discipline, as opposed to what is suggested in economic history, was not in itself the essence of the factory, rather the introduction of factory discipline was a consequence of the appearance of authority and the coordinating mechanisms related to that authority. In this sense, factory discipline in itself cannot be considered the distinctive attribute of the factory. At this point the question arises: How do we approach the essence of the factory in a *historical* theory of the firm?

To integrate the *nature* of the factory into economic history, one should take into account what the theory of the firm literature (e.g., Foss 2002, Kapás 2004) says on the distinctive characteristic of the firm. The issue of the distinctive attribute of the firm has received more attention during the past 10 years, due to an increasing analysis of the boundaries of firms in the knowledge economy. This literature (Foss 2002, Kapás 2004) takes the view that the distinctive attribute of the firm is the preponderance of firm-like coordinating mechanisms (authority) among the coordinating mechanisms used within the firm. Applying this to the factory, the distinctive attribute of the factory, i.e., the one that implies firm-ness, is firm-like authority, as opposed to the putting-out system which was characterized by a market-type network of contracts.

Thus, the major question is why and how authority appeared during the Industrial Revolution. Since the essence of the BIR was macroinventions (see Mokyr 1990), the question, in fact, is how radical technological changes led to an authority-based organization of production. That is, a *historical* theory of the firm has to show how and why authority became an inevitable coordinating mechanism due to macroinventions.

Thus, a *historical* theory of the firm view on the factory must be built on what the essence of the BIR was. Mokyr (1990) clearly argues that a clustering of macroinventions was the essence of the industrial revolution which, from the viewpoint of my approach, led to two fundamental and closely interwoven processes, namely the extension of markets and an improved division of labor to an extent never seen before, which in turn increased the demand for new and better quality goods (Mantoux 1961). New products, that is, products due to macroinventions, required the new technology, and the improvement of the already existing products also needed the new technology. The production of the new goods required a monitoring different from that of the putting-out system. As noted before, this new kind of monitoring was the most important element of factory discipline, and it relied on firm-like authority.

How can we characterize this new type of monitoring? The monitoring under the putting-out system differed from that in the factory in two respects. First, the subject to be monitored changed. While in the putting-out system, the merchant-entrepreneur supervised the product itself, in the factory it was possible to monitor the production process. Second, according to Cohen (1981) the essential difference between the factory and the putting-out system lay in who controlled the production process: in the household it was the family head, in the factory it was the factory owner.

So, the factory did not invent monitoring and the essential difference between the two systems was not to be found in the fact that there was monitoring in one but not in the other. Both systems had monitoring, but the factory fundamentally changed its character: the contract (market)-based monitoring of the putting-out system (Langlois 1999) turned into firm-like monitoring.[97]

Accordingly, monitoring – which, at that time, was the most important building block of authority – as such was not a distinctive feature of the factory; rather its distinctive attribute was firm-like monitoring. Put differently, the distinctive attribute of the factory was that authority-based monitoring became the most important coordinating mechanism among the coordinating mechanisms used within the factory.

To sum up, based on historical facts, I associate firm-like monitoring with two developments: (1) a monitoring of the production process instead of a monitoring of the output, (2) a change in the individual charged with monitoring; instead of the family head, it became the factory owner. The above two characteristics of monitoring brought about by the factory are precisely those the theory of the firm understands by firm-like monitoring (see Foss 2002), which constitutes the essence of the firm.

Moreover, firm-like monitoring, in its turn, brought about the employment relationship, which, according to Coase (1937) is an essential element of the firm. That is, the factory is a Coasean firm. The employment relationship emerged as a result of a fundamental change in the nature of labor exchange (Gintis 1976): the worker sold his/her labor power for a specific period of time and in return agreed to accept the authority of the factory owner (in the sense of Simon 1951) in matters of discipline, supervision and organization of the work process.[98] So, basically the authority that became the major coordinating device within the factory embodies the relationship between the employee and employer (entrepreneur): the entrepreneur exercises direction, and employees agree to obey him/her within certain limits. This concept of authority was formalized by Simon (1951) and has become common in the literature: authority refers to the manager's right to direct the workers within their "zone of acceptance". Here authority is based upon the control and monitoring of individual efforts.[99]

To summarize, as opposed to the modern theory of the firm view (e.g., Alchian and Demsetz 1972, Williamson 1980, 1985) which sees an economizing in transaction costs as the major cause, the historical account of the transition from the putting-out system to the factory can be claimed to have been, at least in part, the result of the extension of markets and the division of labor leading to new technology-intensive product (see above). In this sense, the roots of the factory are to be found in the putting-out system itself that prevailed before the

[97] Note that firm-like monitoring was also present in the putting-out system in the sense that there was a hierarchy in monitoring itself. When the number of producers tended to become unmanageable, the putter-out simply engaged a sub-putter-out who dealt on their behalf with a number of producers. Normally, this strategy could not increase costs, since the sub-putter-out made a profit by increasing supervision on producers (Kieser 1994).

[98] As shown by Pitelis (1998), another explanation of why there is a capitalist employment relation to start with is the Marxist one as developed by, among others, Marglin (1974) and Hymer (1979). According to this, the employment relation was imposed on workers by capitalists through coercion. The Marxist view focuses on the power-control related distributional benefits for capitalists of the employment relation.

[99] In line with Pitelis (1998), note that authority existed in a family craftshop as well, but the latter would not be a Coasean firm.

BIR. That is, the rise of the factory can be best understood as an evolutionary process (for more details see Kapás 2008).

The above perspective, integrating economic history and the theory of the firm, shows why firm-like monitoring implied at the same time an employment relationship (authority), and led, accordingly, to the emergence of a *firm (factory)*. Of course, once the Coasean firm had been established, the productive benefits related to teamwork and knowledge enhancement in the framework of an administrative organization could lead to additional changes as regards the organization of the (Coasean) firm.

Summary

The factory was one of the most significant institutions of the BIR. Economic historians tended to point to the technological origins of the factory, while the theory of the firm has not paid special attention to this, its general concerns being related to the nature of the firm. Seemingly, the two disciplines focus on different aspects of the factory and are separated from one other. In this paper I argued that a better understanding of the factory needs a framework in which the two disciplines are taken together.

In this endeavor, I built upon the idea that the factory was the first form of the capitalist firm and I augmented economic history views on the factory with a theory of the firm perspective. The added value of this *historical* theory of the firm view on the factory was to recognize the distinctive feature of the factory through the historical process of its emergence.

References

Alchian, A. A. – Demsetz, H. (1972) Production, Information Cost, and Economic Organization. American Economic Review 62. 5:777-795.

Berg, M. (1991) On the Origins of Capitalist Hierarchy. In: Gustafsson, B. (ed), Power and Economic Institutions: Reinterpretations in Economic History Aldershot: Edward Elgar. pp. 173-194.

Clark, G. (1994) Factory Discipline. Journal of Economic History 54. 1:128-163.

Coase, R. H. (1937) The Nature of the Firm. Economica 4. 16:386-405.

Cohen, J. S. (1981) Managers and Machinery: An Analysis of the Rise of Factory Production. Australian Economic Papers June:24-41.

Foss, N. J. (2002) 'Coase vs Hayek': Authority and Firm Boundaries in the Knowledge Economy. International Journal of the Economics of Business 9. 1:9-36.

Georgescu-Roegen, N. (1970) The Economics of Production. American Economic Review Papers and Proceedings 60. 2:1-9.

Geraghty, T. M. (2003) Factory System. In: Mokyr, J. (ed), Oxford Encyclopedia of Economic History. Oxford: Oxford University Press.

Geraghty, T. M. (2007) The Factory System in the British Industrial Revolution: A Complementarity Thesis. European Economic Review 51. 6:1329-1350.

Gintis, H. (1976) The Nature of Labor Exchange and the Theory of Capitalist Production. Review of Radical Political Economics 8. 2:36-54.

Hayek, F. A. (1945) The Use of Knowledge in Society. American Economic Review 35. 4:519-530.

Hodgson, G. M. (2001) How Economics Forgot History: The Problem of Historical Specificity in Social Sciences. London, New York: Routledge.

Hodgson, G. M. (2002) The Legal Nature of the Firm and the Myth of the Firm-Market Hybrid. International Journal of the Economics of Business 9. 1:37-60.

Hymer, S. H. (1979) The Multinational Corporation: A Radical Approach. In: Cohen, R. B. – Felton, N. – Liere, J. van – Nkosi, M. (eds), Papers by Stephen Herbert Hymer. 1971. Reprint. Cambridge: Cambridge University Press.

Jones, S. R. H. (1982) The Organization of Work. A Historical Dimension. Journal of Economic Behavior and Organization 3. 2-3:117-137.

Jones, S. R. H. (1987) Technology, Transaction Costs, and the Transition to Factory Production in the British Silk Industry, 1700-1870. Journal of Economic History XLVII. 1:71-95.

Jones, S. R. H. (1999) The Rise of the Factory System in Britain. Efficiency or Exploitation? In: Robertson, P. L. (ed), Authority and Control in Modern Industry, London: Routledge. pp. 17-44.

Kapás, J. (2004) Mutant-Firms in the New Economy. Economie et institutions 5. 2:77-96.

Kapás, J. (2008) Industrial Revolutions and the Evolution of Firm Organization: An Historical Perspective. Journal of Innovation Economics 1. 2:15-33.

Kieser, A. (1994) Why Organization Theory Needs Historical Analyses and How This Should Be Performed. Organization Science 5. 4:608-620.

Landes, D. (1969) The Unbound Prometheus: Technological and Industrial Development in Western Europe from 1750 to Present. Cambridge: Cambridge Universirty Press. 2003.

Landes, D. S. (1966) (eds) The Rise of Capitalism, New York: Macmillan.

Landes, D. S. (1986) What Do Bosses Really Do? Journal of Economic History 46. 3:585-623.

Langlois, R. N. (1999) The Co-evolution of Technology and Organization in the Transition to Factory. In: Robertson, P. L. (ed), Authority and Control in Modern Industry, London: Routledge. pp. 45-72.

Leijonhufvud, A. (1986) Capitalism and the Factory System. In: Langlois, R. N. (ed), Economics as a Process. New York: Cambridge University Press. pp. 203-223.
Magnusson, L. (1991) From Verlag to Factory: the Contest for Efficient Property Rights. In: Gustafsson, B. (ed), Power and Economic Institutions: Reinterpretations in Economic History Aldershot: Edward Elgar. pp. 195-221.
Mantoux, P. (1961) The Industrial Revolution in the Eighteenth Century: An Outline of the Beginnings of the Modern Factory System in England. Chicago: University of Chicago Press. Revised Edition.
Marglin, S. A. (1974) What Do Bosses Do? The Origin and Functions of Hierarchy in Capitalist Production. Review of Radical Political Economics 6. 2:20-112.
Marx, K. (1867) Capital: A Critique of Political Economy, Vol. III. The Process of Capitalist Production as a Whole.
Available: http://www.econlib.org/library/YPDBooks/Marx/mrxCpC.html
McKendrick, N. (1961) Josian Wedgwood and Factory Discipline. In: Landes, D. S. (ed), The Rise of Capitalism. New York: Macmillan. 1966. pp. 65-80.
Mokyr, J. (1990) The Lever of Riches: Technological Creativity and Economic Progress. New York, London: Oxford University Press.
Mokyr, J. (1993) Editor's Introduction: The New Economic History and the Industrial Revolution. In: Mokyr, J. (ed), The British Industrial Revolution: An Economic Perspective. Boulder and Oxford: Westview Press. pp. 1-127.
Mokyr, J. (2001) The Rise and Fall of the Factory System: Technology, Firms, and Households since the Industrial Revolution. Carnegie-Rochester Conference Series on Public Policy 55:1-45.
Mokyr, J. (2002) The Gifts of Athena. Historical Origins of the Knowledge Economy. Princeton University Press.
Morroni, M. (1992) Production Process and Technical Change. Cambridge: Cambridge University Press. 2009.
Piore, M. J. – Sabel, C. F. (1984) The Second Industrial Divide: Possibilities for Prosperity. New York: Basic Books.
Pitelis, C. N. (1998) Transaction Costs and Historical Evolution of the Capitalist Firm. Journal of Economic Issues 32. 4:999-1017.
Pollard, S. (1963) Factory Discipline and the Industrial Revolution. Economic History Review (New Series) 16. 2:254-271.
Pollard, S. (1964) Fixed Capital and Industrial Revolution. Journal of Economic History 24. 3:299-314.
Pollard, S. (1965) The Genesis of Modern Management. London: E. Arnold.
Simon, H. A. (1951) A Formal Theory of the Employment Relationship. Ecometrica 19. 3:293-305.
Smith, A. (1776) An Inquiry into the Nature And Causes of the Wealth of Nations. London: The Adam Smith Institute. Electronic Edition, 2001.
Available: http://www.adamsmith.org/smith/won-index.htm.
Sombart, W. (1902) Der moderne Kapitalismus. Leipzig: von Duncker und Humblot. Available:http://www.archive.org/stream/dermodernekapit01sombgoog#page/n6/mode/2up
Szostak, R. (1989) The Organization of Work. The Emergence of the Factory Revisited. Journal of Economic Behavior and Organization 11: 343-58.
Williamson, O. E. (1980) The Organization of Work: A Comparative Institutional Assessment. Journal of Economic Behavior and Organization 1. 1:5-38.
Williamson, O. E. (1985) The Economic Institutions of Capitalism. New York: Free Press.