Caste Systems and Technology in Premodern Societies

Caste Systems and Technology in Pre-Modern Societies*

Elena Esposito‡

January 28, 2015

Abstract

We explore the link between the presence of a particular type of social institution, the caste system, and technology in pre-modern societies. We find that caste-societies are associated with a higher level of technological sophistication and specialization in pre-modern times. The correlation is robust to all main established determinants of pre-modern technological sophistication. In order to assess causality, based on anthropological theories suggesting that caste systems represented a social device aiming at reducing disease contagion, we instrument castes with an index of disease exposure. The results indicate the existence of a positive causal link between caste institutions and technological sophistication at an early phase of development.

Keywords: Technology Adoption, Division of Labor, Caste Systems, Disease Exposure.
JEL Classification: O10, O33, N10, P50, J62.

*I am extremely grateful to Matteo Cervellati, for the kind supervision and the encouraging attitude. A special thank to Giovanni Prarolo for many useful advices. The author wishes to thank Luz Marina Arias, Jeorg Baten, Maria Carillo, Carl-Johan Dalgaard, Oded Galor, Andrea Ichino, Anastasia Litina, Holger Strulik and Emanuele Tarantino for helpful comments and suggestions, as well as participants at the conference "Intergenerational Transmission of Entrepreneurship, Occupations and Cultural Traits in the process of long-run economic growth" at the University of Naples Parthenope, and at the Economic History Panel of the EEA-ESEM 2012 Congress in Malaga. The Department of Economics at the University of Bologna is being a very supportive and comfortable learning environment.

†University of Bologna

‡*Contact details:* University of Bologna, Department of Economics, Piazza Scaravilli 2, Bologna, ITALY. E-mail: elena.esposito6@unibo.it. Tel: 0039 051 20 98 887.

1 Introduction

Caste systems have long been considered an extremely inefficient type of economic and social institution, insofar as they constrain social mobility, fractionalize societies and foster discrimination. The long term negative detrimental effects of castes on modern economic growth can be hardly questioned. Nonetheless, the depth and persistence of this very ancient form of social fractionalization[1] is suggestive of the possibility that castes, under different historical constraints, may had represented an asset, instead of a cost. In fact, caste systems have not simply been a special form of social stratification, but have constituted a complex institutional set up with important effects over the economic organization of societies.

In this article, we explore the possibility that caste systems had been comparatively advantageous at an early phase of economic development. In particular, we test the hypothesis that castes had positively affected technological sophistication and labor specialization in the institutional, economical and technological context of pre-modern societies. The hypothesis builds on the idea that caste systems, by promoting strong ties of solidarity and cooperation within groups, might had facilitated and accelerated the process of labor specialization and technological sophistication. In fact, the strong reciprocal feelings of trust might have contributed to solve common problems of production and exchange in pre-modern societies, such as the danger of information disclosure, the enforcement of contracts through collective action and the cost of transferring skills. Moreover, anytime that economies of scale emerged, the endogamous group might had represented the natural basin where to recruit labor for extending production beyond the family unit. On top of this, workers in the same endogamous group might had benefited of technological spillovers from other kins. In this sense, castes might had facilitated the development of artisan clusters in pre-modern societies.

We use data from Murdock's Standard Cross Cultural Sample and *Ethnographic Atlas*, two ethnicity-level datasets which permit to compare premodern societies before the European colonization. To our knowledge, these are the only data sources allowing to map the worldwide distribution of caste systems. For comparing technological sophistication in pre-modern societies, we need to pull far back the frontier of technology, back to a time where

[1]The origin of Indian caste systems has been dated variously between 3000 B.C. to approximately 1000 B.C. (Thapar, 2004). It is more complicated to date the origin of castes in Western and Eastern Africa, because of the lack of written records. Tamari (1991) summarized evidence regarding the emergence of caste systems in western Africa, which appeared no later than 1300 AD according to some sources, and no later than 1500 according to other sources.

the mere presence of specialized occupational practices represented, comparatively, a leap forward. The first measure of technological sophistication we use is based on the presence of three key types of artisanal skills, i.e. pottery, weaving and metalworking. Indeed, less than one fifth of societies in our sample disposed of all these three basic types of technologies. The second index of technological sophistication we employ, inspired by the work of Comin, Easterly, and Gong (2006), attempt to measure the level of technological sophistication along several dimensions, and namely: agriculture, artisan crafts, writing and land transportation. Finally, we also investigate the degree of labor specialization. For this reason, we explore to what degree occupational practices are performed as specialized arts or occupations.

We find a strong positive relationship between caste stratification and the level of technological sophistication and labor specialization in pre-industrial societies. The relationship remains economically meaningful and statistically significant even after controlling for various potentially relevant geographical and environmental characteristics of the land occupied by the society, as well as location controls such as distance from rivers and seas. Nevertheless, the correlation may be driven by omitted factor, and a reverse effect of technological sophistication on caste stratification cannot be ruled out.

In order to understand whether the effect of castes on technology is causal, we follow three strategies. First, we try to control for all major established determinants of pre-modern development. The correlation between caste stratification and pre-modern technological sophistication is robust to all relevant controls: to the timing of transition to agriculture, the migratory distance from Adis Abeba and the distance from the regional technological frontier (Ashraf and Galor (2011a) and Ashraf and Galor (2013)). Moreover, controlling for observable society characteristics, such as population density and the level of political centralization, reduces the magnitude of the correlation only slightly. This evidence suggests that for controlling for observable characteristics affect our results only mildly.

The second strategy we employ rely on an exogenous source of variation in the likelihood of observing caste stratification. Following historian William McNeill (1976), who argued that castes represented an early social attempt to reduce the risk of contagion in highly diseased environments, we test whether castes prevailed in areas that, for bio-geographical reasons, had to deal with a higher burden of pathogens. We construct an index of geographical disease exposure exogenous to population size and development level and we find that it significantly predicts a higher probability of observing societies with castes, suggesting thus a tentatively causal relation. Instrumenting the variable castes with our index of disease exposure, we find a strong and causal effect running from caste stratification to pre-modern societies technological sophistication and labor specialization.

This paper contributes to multiple strands of the literature. First, it complements the evidence of recent studies documenting the crucial role of social institutions and cultural norms for economic outcomes (Greif, 1993, 1994; Guiso, Sapienza, and Zingales, 2008; Tabellini, 2007; Nunn and Wantchekon, 2009) and, along the same line, the impact of social fractionalization on growth and institutional quality (Easterly and Levine, 1997; Alesina and Ferrara, 1999, 2004; Fearon and Laitin, 2003, among many others). In particular, our study enriches this literature substantiating the idea that the economic effects of institutions and norms, as the effects of social fractionalization, are not fixed but vary with the historical context. Second, this research add new elements to the literature exploring determinants of historical development (Ashraf and Galor (2011a), Ashraf and Galor (2011b), Aiyar, Dalgaard, and Moav (2008) Comin, Easterly, and Gong (2006)).

Moreover, the article tangentially contributes to the debate over the role of institutions versus geography in economic development (Gallup, Sachs, and Mellinger, 1998; Acemoglu, Johnson, and Robinson, 2001; Sachs, 2001; Acemoglu, Johnson, and Robinson, 2001; Sokoloff and Engerman, 2000, among the various contributions). In fact, from one hand, we document that institutions mattered for economic development even in pre-modern societies, insofar as we find that castes spurred technology and labor specialization. On the other hand, we slightly depart from the institutional view since the effect of geography on institutions that we find pre-dates European colonization. In the same way, our findings slightly diverge from the standard geography view as we show that, in pre-modern times, diseases did not negatively affect technology but, through caste stratification, indirectly favored labor diversification and early technological sophistication. Along the same line, by highlighting an effect of disease exposure on the presence of caste systems, our analysis complements recent literature on the geographical origin of institutions, norms and social diversity, which traces the roots of their historical emergence in the bio-geographic environments that societies had to confront (Alesina, Giuliano, and Nunn, 2011b,a; Michalopoulos, 2008). Finally, this paper, despite focusing on pre-modern times, indirectly adds new elements to the literature exploring the impact of a diseased environment over various economic, social and political outcomes (Gallup, Sachs, and Mellinger, 1998; Cashdan, 2001; Bloom, Canning, and Sevilla, 2003; Sachs, 2003; Lorentzen, McMillan, and Wacziarg, 2005; Acemoglu and Johnson, 2006; Weil, 2007; Fincher, Thornhill, Murray, and Schaller, 2008; Cervellati, Sunde, and Valmori, 2011a,b).

In Section II, we begin our analysis by laying out the historical and conceptual framework. In Section III, we turn to a description of the data and present OLS estimates. In Section IV we present the identification strategy for pinning down the causal relationship and report our IV estimates. Section V concludes.

2 Historical Background and Conceptual Framework

2.1 Historical Background

Caste-like stratification existed - and often still persist - in various geographical regions and within ethnic communities at different levels of economic development. The most studied and known example of caste system is the Indian one, by far the most structured caste system we know, being at the time an economic, religious, social and political institution. But castes - or better, system of social stratification with significant similarities to the Indian one - are wide-spread in western and eastern Africa and in other parts of Asia[2]. Sure enough, caste-like stratifications vary widely across regions, complicating the task of cross-cultural analysis. For instance, in the Indian system virtually any Hindu was recognized as part of an endogamous caste, or *jati*, while in Western Africa caste people formed a small minority of the population (Tamari, 1991)[3]. In order to be able to compare castes cross-culturally, we proceed by providing a parsimonious operational definition. Following Berreman (1968), we define caste systems as "a hierarchy of endogamous divisions in which membership is hereditary and permanent". This definition stresses two main features of caste stratification, and namely: the hereditary status - one belongs to a caste only if he/she is born in that caste - and the rigid endogamic nature of the group - one belongs to a caste only by marrying within the caste, otherwise risking expulsion.

The definition we employ is a sort of lowest common denominator of caste-like institutions found in all the societies that we will compare. However, in no way this definition exhausts the religious meanings, the functions

[2] Anthropologists and sociologists have been long debating over whether it is appropriate to name "castes" the caste-like institutions found outside the Indian peninsula. In this context, we take a step back from the debate and circumscribe our analysis by providing a clear operational definition of what we consider a caste.

[3] "They have been estimated to form ten to twenty per cent of the Wolof population, and about five per cent of the Soninke population. They also seem to form about five per cent of the Bambara population. Among the Dogon, and a fortiori among the Dan, Minianka and other peoples who recognize only one or two castes, they form a yet smaller proportion of the population" (Tamari, 1991).

and roles that castes have been playing cross-culturally. Besides these common features, there are several characteristics that have been found across the majority of societies examined, but not in all of them. First of all, in the majority of cases castes are associated with a traditional hereditary occupation, which is transmitted from father to son (or mother to daughter) and is often forbidden to other members of the community. Moreover, caste stratification is very frequently supported by ideological-religious beliefs of purity-pollution. Caste people - or lower caste people - are considered impure and thus to merit segregation and disdain. The impure nature of caste people is sometimes attributed to the profession practiced (i.e. washers, sweepers...), or to the dietary habits of the community (i.e. eating porks, dead animals etc...). Along the same lines, several communities prescribe limits to the interactions among members of different castes, the strict prohibition of commensality being the most frequent one (Hutton, 1946; Dumont, 1980; Tamari, 1991).

Our notion of caste might, in some cases, not permit a clear-cut distinction with other kinds of social institutions or other type of social fractionalization, like kin groups, ethnic groups and social classes. Indeed, these various types of social divisions often overlap. Still, caste institutions maintain their specificities. For instance, we might say that a caste is a sort of kin group, since kin groups sometimes have very precisely set endogamic rules which may recall the endogamicity of castes. However, very often kin groups marry outside the group, in many cases the relationships among groups have no hierarchical nature and they are rarely associated to a traditional occupation. For what concerns ethnic stratification, different castes sometimes belong to different ethnic groups. Theories over the origin of caste systems argued that castes originated from different ethnic groups which happened to form economically-integrated but socially segregated communities (Hutton, 1946)[4]. However, in the majority of cases, ethnic differences are not apparent and castes of a same community tend to speak the same language. Finally, caste stratification might coincide with class stratification. For in-

[4]Recent evidence from genetics supported the idea that Indian castes are genetically distinct and might had originated from different tribal groups, however, the debate is still ongoing. From one hand Mountain, Hebert, Bhattacharyya, Underhill, Ottolenghi, Gadgil, and Luca Cavalli-Sforza (1995) found no clear genetic separation along caste lines, on the other hand, Bamshad, Kivisild, Watkins, Dixon, Ricker, Rao, Naidu, Prasad, Reddy, Rasanayagam, et al. (2001) documented that higher caste Indians are closer to Europeans and lower caste Indians to Asians. Similar studies are not available for other caste systems located in Asia and Africa. Nonetheless, ethnographic evidence acquired by anthropologists raised similar hypothesis regarding the origins of caste systems in other areas. See, for instance Shack (1964) on castes from South-West Ethiopia, and Barth (1956) on a caste system in Northern Pakistan.

stance, in India the possession of land was often prohibited to outcasts and lower castes. At the same time, being granted an exclusive right to perform a specific occupation often permitted caste people to reach a considerable wealth:

> "Sanitary workers veiled themselves in Lagos to avoid recognition, but everywhere they showed much corporate solidarity and some achieved considerable economic success ... The Dorze weavers of Addis Ababa suffered extreme insult and discrimination but earned perhaps three or four times the national average in 1970[5]"

2.2 Conceptual Framework

Castes have long been considered an obstacle to economic growth. Weber (1958) was among the first social scientists pointing to an alleged incompatibility between modern economic growth and the caste system. First of all, a division of labor based on hereditary status instead of contracts may discourage entrepreneurship no less than forestalls workers' ambitions (Hutton, 1946; Berreman, 1968). Beside the distortionary effects on workers' incentives, Bose (1916) considered the hereditary passing of crafts and skills as an intrinsic limit to the introduction of innovation, since it made an innovation in method appear as a sin against the craftsman's ancestors. In a recent analysis, Santacreu-Vasut (2009) argued that social diversity generates information asymmetries which cause the adoption of less growth-enhancing managerial institutions. She explained the diverging path of productivity between Indian and Japanese cotton industries as the result of a higher social diversity, and of the less efficient institutions which it engendered. Even within the same Indian state, factories with more diverse labor force tended to recur to less profitable managerial set ups, with more workers supervision when the labor force is diverse.

Moreover, castes might prevent an efficient allocation of talents across professions by prescribing hereditary occupational status or by discriminating workers of lower castes. The effects of Indian caste-based discrimination are dramatically persistent, in a recent experiment Hoff, Pandey, and Team (2004) showed how low-caste students tend to perform significantly worse anytime salience is placed on their caste status.

Beside the direct effects on productivity and innovation, caste-based fractionalization might affect economic growth even through indirect channels, such as the quality of institutions a society decide to adopt and the level of

[5] Iliffe (1987).

public goods to provide. Banerjee, Iyer, and Somanathan (2005) find that social heterogeneity in Indian villages weakens the political relevance of the community and reduces the amount of public goods it is able to secure. More broadly, caste stratification might damage the quality of institutions just as ethnic fractionalization has been found to do (Alesina and Ferrara, 1999, 2004).

All this evidence suggests to consider castes as extremely costly institutions, among other things, in terms of efficiency. Still, caste systems are present in historically relatively prosperous societies, and in the majority of cases have lasted till our days. As pointed by Greif (2006), institutions tend to be persistent, because of the significant adjustment costs to be faced for transforming them. On top of this, institutions are for the most part embedded in a complex matrix of cultural and social norms, which are intrinsically slowly changing (North, 1990). In this light, it becomes relevant to inquire whether what represents a truly inefficient institution in the context-specificity of a modern society, might have provided some advantages under different technological, social, economic and institutional circumstances. In other words, can castes had been a booster of growth at a certain point in history? In this article, we test whether this had been the case. More precisely, our hypothesis build on the idea that the previously documented detrimental effects of castes on growth might have prevailed only at a relatively late stage of technological sophistication. And, on the contrary, caste stratification had boosted technological sophistication and specialization at an earlier phase. Indeed, caste organization might have had the crucial function of "artificially" spur the division of labor, which since Adam Smith has been recognized as the first driver of growth. In this light, Bougle's remark on the economic consequences of the Indian castes is particularly suggestive:

> "..the caste which mends shoes refrains from making a pair of them... it takes three distinct craftsmen to make a bow and arrows..."[6].

The role of castes at an early stage of technological sophistication has not been explored in the economic literature. On the contrary, anthropologists and historians have long been arguing in this sense. Diop and Salemson (1987) claimed that labor specialization strictly followed clan and caste organization. Along the same line, Nesfield (1901)[7], one of the first indianologist studying the economic features of the caste system, was reading the Indian system as an institutional device for achieving labor specialization.

[6]From Hutton (1946).

[7]From Hutton (1946).

A considerable size of the literature on castes had been emphasizing the exploitative traits of the economic relations in caste societies. And indeed, the power relationship between social groups in caste societies cannot be underplayed. Nevertheless, limiting the analysis to the exploitative side of caste stratification does not allow to fully account for some relevant features of the economic relationships in caste societies. In this sense, it is worth quoting again Diop and Salemson (1987):

> "For each caste: inconveniences and advantages, alienations and compensations balance each other out... the stability of the caste system is secured by the hereditary transmission of social functions that corresponds, to a certain extent, to a monopoly, disguised in religious interdiction, in order to eliminate competition".

It is necessary to qualify that Diop is specifically referring to caste institutions in Wester Africa and, still, is probably underplaying some of the dramatic social costs of being part of a despised group. Moreover, the same analysis would misrepresent the truly dismal conditions of many Indian out-castes. Still, the economic analysis of castes merits further scrutiny and cannot be simply dismissed as a particular form of slavery. In this light, the question we need to explore is why we expect castes to favor technological sophistication and labor specialization in pre-modern societies.

One possible explanation is that the caste - the kin-group or the clan - was the natural boundary where to recruit whenever arose the possibility of exploiting economies of scale in production, and whenever the production process required an amount of labor larger than what could be found within the family unit. The fact that each caste tended to be geographically localized in a specific area of the village could represent a favoring pre-condition for artisan clustering, which is a common feature of pre-modern crafts because of the positive organizational and technological externalities it engendered[8].

Another possible explanation is that the trust and cooperative nature of the relationships which were likely to prevail within the endogamous group could solve well-known problems of production in pre-modern societies. First of all, confining the artisan knowledge to the group could work as an early form of intellectual property protection. It is suggestive to recall how, in

[8]Epstein (1998) pointed to the tendency of pre-modern craft guilds to cluster in the same neighborhood as a common feature of European medieval cities. It is sufficient to remember the district of the silk workers in Bologna, and the names of the roads in the ancient parts of the town which recall the craft professionals clustering along the road (goldsmiths, carpenters...).

medieval Europe, craft guilds emphasized the "brotherly" nature of the association through elaborate symbolic liturgies and ceremonies, with the primary scope of promoting feelings of trust and cooperation among members. In a caste, such trustworthiness was granted by birth status within the endogamous group and this common background might had reduced fears of information disclosure. Epstein (1998) stresses the importance of intellectual property protection for European guilds, arguing that "deliberate inventions will not be forthcoming if the inventor cannot claim more than his proportional share of the gains". Of the possible solutions to the problem (patent rights, state support for research, etc.), only secret transmission of crafts to trusted members was likely to be an available option in pre-modern societies.

Another possible explanation why endogamous groups might had found it convenient to restrict production within the caste is the potentially high access cost to artisanal skills in pre-literate societies. Stressing again the comparison with medieval Europe might provide some further insights. Epstein (1998) argued that craft guilds emerged in order to provide transferable skills through apprenticeship and to share out the unattributed costs and benefits of training among its members. Again, the transfers of skills within a group of kin might be safer and less costly.

Finally, in a context where enforcement institutions may had been missing or lacking, specialization within the caste may had represented a tool for contract enforcement. Freitas (2006) argues that caste, as an information-sharing institution, may had facilitated collective action as a way to enforce contracts, with beneficial effects on trade.

3 Data and Empirical Estimations

3.1 Data Sources

Our analysis relies on information from two main data sources: the *Ethnographic Atlas* the Standard Cross Cultural Sample (SCCS), two ethnicity-level datasets constructed by George Peter Murdock providing information on societies as early in time as information is available (Murdock, 1967). The *Ethnographic Atlas* includes 1,267 ethnic groups, primarily from observations recorded in the late 19th and early 20th century. The SCCS, a sub-sample of the Ethnographic Atlas, is composed by 186 societies chosen to maximize independence in cultural and historical origin. The scope of these datasets is to "photograph" societies in their pristine state, before that contact with Europeans radically modified their technologies, economies and institutions. With this aim, the earliest documenting material available from missionaries, voyagers, merchants and anthropologists was analyzed and coded.

One limitation of these data is that societies are observed at different points in time[9]. However, it is reassuring to note that the pin-pointing time is much more homogeneous within the same region. Moreover, since the dataset provides precise information regarding the year each society was observed, we try to guarantee robustness of results in two additional ways: i) we control for the pin-pointing date; ii) we provide estimates of the baseline specifications for homogeneous sub-periods. To our knowledge, these are the two only datasets available providing information about the presence of castes in pre-modern societies.

Castes Systems SCCS variable $v272$ and Ethnographic Atlas Variable $v98$ track the presence of caste stratification in each society, specifying which type of endogamous stratification is observed, and namely whether stratification takes the form of despised occupational groups, ethnic stratification or a complex type of stratification[10]. In the SCCS, out of 181 societies for which we have information, 18 societies had despised occupational groups, 3 societies presented ethnic endogamous stratification and 7 societies had a complex type of endogamous stratification. Endogamous stratification in the Ethnographic Atlas follows a similar pattern: of the 1079 societies of the Atlas for which we have information, about 10% had endogamy of despised occupational groups, 3% a complex type of caste stratification and a little more than 2% an ethnic endogamous stratification. Based on these variables, we created a dummy variable indicating whether a type of endogamous caste stratification is present ($CASTE_i$=1) or not ($CASTE_i = 0$) within the society[11]. As Figure 1 shows, castes have been mainly African and Euroasiatic institutions.

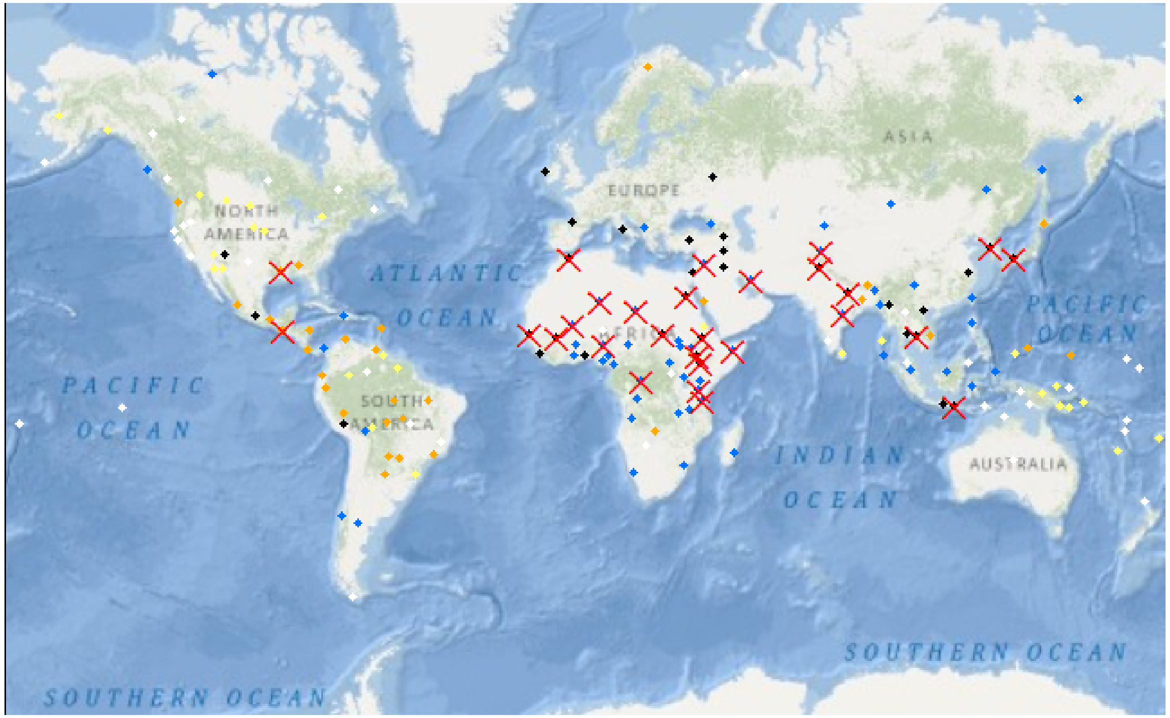
Technological Sophistication In order to compare the degree of technological sophistication, we exploit two indexes aiming at capturing the level

[9]In the Ethnographic Atlas, the pinpointing date for 3 societies dates before 1000 AD, for 42 societies between 1001 and 1800 AD, 32 after 1950 and all remaining between 1801 and 1950. For the SCCS, the pinpointing date for 3 of the 186 societies is before 1500 AD (Khmer, Hebrews and Romans), for 23 societies between 1500 and 1850, and the remaining societies' information is from late 19th and 20th century.

[10]Complex caste stratification include all multidimensional types of stratification, which can be at the time occupational, traditional, ethnic etc...

[11]For robustness, we created a second dummy variable $CASTEnoethnic_i$ where we exclude ethnic stratification and replicated the analysis. Results remain qualitatively unchanged.

Figure 1: Caste and Technology in the SCCS[a]



[a]The figure maps the distribution of societies in the SCCS. Each dot corresponds to a society. The color of each dot represents the level of technological sophistication of the society: white stands for no artisan, yellow stands for pottery only, green stands for weaving only, blue stands for metal-work and one other craft (either potters or weavers), black stands for societies with potters, weavers and metal-workers. The red cross around some of the societies in the sample indicates the presence of endogamous groups within the society.

of sophistication in artisanal skills and the level of general technological sophistication.

The first index is intended to measure the degree of complexity and specialization in artisanal crafts and corresponds to SCCS variable $v153$. The index takes value one in societies where metalworking, loom-weaving and pottery are all absent while 5 is assigned to societies reported to have a variety of craft specialists, including at least smiths weavers and potters. Between these boundaries, the index is equal to 2 for societies where only pottery is practiced, 3 for societies which have specialized weavers (weavers only or weavers and potters) and 4 to societies which have metalworking but lack loom-weaving and/or pottery[12]. In the Ethnographic Atlas, a corresponding variable is missing but, by recombining information from other variables, it is possible to reconstruct a similar index of artisanal sophistication[13]. The measure of craft sophistication we are using might strike as a very raw and reductive indicator of technology. However, in order to make possible a comparison of technologies in pre-modern societies we need to pull back the frontier of technological sophistication. Indeed, it is worth emphasizing that less than 20% of all societies in the sample disposed of the three types of artisan practice.

The second index is meant to rank societies according to their overall level of technological sophistication in multiple fields, and namely: agriculture, artisanal crafts, writing and land transportation. The index is intended to mimic the technological index devised by Comin, Easterly, and Gong (2006), and was constructed aggregating information from several SCCS variables[14]. The All-Technologies index ranges from 4, the lowest value of technological sophistication, to 12, the highest rank, which is attributed to societies scoring the highest in all dimensions. Table 1 shows that the Circum-Mediterrean region and the EuroAsiatic region presented a higher level of both artisanal and general technological sophistication. Note, however, that in our analysis we will mainly focus on within-region variation of technological sophistication[15].

[12]The specific definition of crafts employed for devising SCCS variable $v153$ were defined as follows: i)Metal working: only such art as smelting, casting and forging, which involve the application of fire ; ii)Weaving: only the manufacture of true cloth on a lume or frame is indicated, not the manufacture of nets, baskets, mats or non-woven fabrics like barkloth or felt; iii)pottery: only the manufacture of earthenware utensils is indicated.

[13]See the Appendix for a detailed description of variables' construction and sources.

[14]The only technological area which is not covered by SCCS is military technology. See the Appendix for details regarding index construction.

[15]Results would not change if we were to lead the analysis by exploiting overall variation, and not just within region variation.

Table 1: Technological Sophistication

	Afr	Med	Asia	Pac	N.Am	S.Am	Total
				Artisan Crafts			
None	3	1	4	14	13	4	39
Pottery	0	1	2	8	11	5	27
Loom weaving	1	2	4	2	5	17	31
Metalwork	21	8	15	5	2	5	56
Smiths, weavers, potters	3	16	9	2	2	1	33
Total	28	28	34	31	33	32	186
				All Technologies			
Index: 4-6	2	0	7	15	17	16	57
Index: 7-9	24	4	11	12	15	11	77
Index: 10-12	1	22	14	2	1	3	43
Tolal	27	26	32	29	33	30	177

Notes: The table reports the number of societies for each technological sophistication category. The upper panel summarizes the artisanal craft index, while the lower panel groups society into three level based on the All-Technology Index. Societies are grouped into six geographical region: "Afr." stands for Sub-Saharan Africa, "Med." for Circum-Mediterranean, "Asia" for East-Eurasia, "Pac." for Insular Pacific, "N.Am" for North America and "S.Am" for South America.

Thanks to the larger size of the Ethnographic Atlas, we are also able to explore the society level of labor specialization for various artisanal crafts. While previous indexes focus on the mere existence of a specific technological know-how, we are interested in inquiring whether, given the presence of a specific technique, the occupation is exclusively practiced by specialized artisans or not. For several crafts, the dataset informs us on whether the occupation is performed as a craft specialization or not. We thus constructed a dichotomous variable $CRAFT_i$ which, given the existence of the technique, takes value 1 if the occupation is performed as a specialized craft and 0 if not. Table 2 has to be read as such: among the 435 societies which were known to work metal, only 17 did not have specialized metalworkers. On the other hand, among the 549 societies known to perform some kind of pottery, for the vast majority (475) pottery was not performed as a specialized craft[16].

[16]It is important to point that the Ethnographic Atlas is an unbalanced sample, i.e. some variables - regarding both the existence and the degree of specialization - are available for some societies and not for others. Therefore, descriptive statistics for each craft cannot be directly compared - we cannot say, for instance, that there are more societies which practice metalwork than weaving, because such a conclusion would heavily depend on the amount of missing data for each variable.

Table 2: Technological Sophistication

	Afr	Med	Asia	Pac	N.Am	S.Am	Total
				Potters			
Not Specialized	178	32	40	38	113	74	475
Specialized	21	33	15	1	2	1	73
Total	199	65	55	39	115	75	548
				Leather Workers			
Not Specialized	65	34	32	4	227	16	378
Specialized	16	41	9	0	1	1	68
Total	81	75	41	4	228	17	446
				Weavers			
Not Specialized	62	36	52	30	88	54	322
Specialized	7	29	20	2	1	1	60
Total	69	65	72	32	89	55	382
				Metal-Workers			
Not Specialized	8	1	0	0	2	6	17
Specialized	210	98	79	27	3	1	418
Total	218	99	79	27	5	7	435

Notes: The table reports the number of societies with specialized artisans. Societies are grouped into six geographical region: "Afr." stands for Sub-Saharan Africa, "Med." for Circum-Mediterranean, "Asia" for East-Eurasia, "Pac." for Insular Pacific, "N.Am" for North America and "S.Am" for South America.

3.2 Preliminary Evidence

We begin by estimating the relationship between the level of technological sophistication of the society and the presence of endogamous groups within the community. Given that the dependent variable $TECHN_i$ is ordinal, we proceed with two estimation strategies. First, we ignore the nature of the data and estimate an OLS model. The second strategy is to maintain the categorical feature of the variable and estimate an ordered logit model. As results remain qualitatively identical, we report results for the second strategy in the Appendix. Our baseline estimating equation is:

$$TECHN_i = \beta_0 + \beta_1 CASTE_i + \beta_2 \mathbf{X_i} + \epsilon_\mathbf{i}$$

where i indexes the society. The variable $TECHN_i$ denotes the level of technological sophistication. From Column (1) to Column (3), the dependent variable is the index of artisanal sophistication, while from Column (4) to Column (6) the dependent variable is the index of general technological sophistication. The dichotomous variable $CASTE_i$ takes value 1 if in the community is present at least one endogamous group, and 0 otherwise. From the second specification onwards, we add a vector of strictly exogenous control variables, **X**, which include geographical, climatic and location controls. Finally, we add region fixed effects in all specifications[17].

[17]The geographical areas are: 1)Sub-Saharan Africa, 2) Circum-Mediterrenean (which includes Northern Africa and Western Europe), 3) East Eurasia (which include Eastern Europe and Asia), 4) Insular Pacific, 5) North America and 6) South America.

Table 3: OLS for Castes and Technological Sophistication

	Technological Sophistication					
	Artisanal Craft			All Technologies		
	(1)	(2)	(3)	(4)	(5)	(6)
CASTES	0.788***	0.836***	0.873***	1.124**	1.102**	1.094**
	(0.241)	(0.263)	(0.270)	(0.481)	(0.469)	(0.486)
Latitude		0.013*	0.012*		0.019*	0.018*
		(0.007)	(0.007)		(0.010)	(0.010)
Mean Temperature		0.002	0.002		-0.005	-0.006
		(0.010)	(0.010)		(0.017)	(0.017)
Mean Rainfall		0.032	0.016		-0.098	-0.096
		(0.044)	(0.047)		(0.060)	(0.063)
Mean Elevation		0.000	0.000		0.001**	0.001**
		(0.000)	(0.000)		(0.000)	(0.000)
Agricultural Potential		0.010	0.015		0.118**	0.122***
		(0.029)	(0.029)		(0.045)	(0.046)
Sea Distance			-0.406*			-0.122
			(0.220)			(0.345)
River Distance			-0.215*			-0.139
			(0.120)			(0.213)
Island Dummy			-0.343			-0.488
			(0.335)			(0.710)
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	160	160	160	153	153	153
R-squared	0.419	0.440	0.458	0.408	0.464	0.469

Notes: The table reports OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Regression results, reported in Table 3, highlight a positive and highly significant correlation between castes and the level of technological sophistication of the society, which changes modestly when including geographical controls, climatic controls and location controls. Note, from Column (1) and (4), that the variable $CASTES$, together with unobservables captured by the regional fixed effects, explains as much as around 40% of the variation in both technological sophistication in artisanal craft and general technological sophisticationWithout region fixed effects, the estimated coefficient of caste stratification on artisan craft is 1.57 (with a standard error of 0.170). The R-squared of the regression without regional fixed effects 0.17, implying that caste stratification alone explains up to 17% of the overall variation in craft sophistication. For the all-technology index, the coefficient is 2.55 (standard error 0.36) and part explained by caste stratification is up to 20%. The presence of castes is associated with a jump of almost one ladder in both indexes of sophistication, more precisely, a 0.8 increase in the craft index and an about 1 point increase in the technology index.

Table 4: Castes and Technological Specialization

	Technological Sophistication	Craft Specialization			
		Pottery	Leather	Weaving	Metal
	(1)	(2)	(3)	(4)	(5)
CASTES	0.420***	0.220***	0.449***	0.186***	0.020
	(0.146)	(0.034)	(0.058)	(0.068)	(0.018)
Geography Controls	Yes	Yes	Yes	Yes	Yes
Location Controls	Yes	Yes	Yes	Yes	Yes
Continent FE	Yes	Yes	Yes	Yes	Yes
Observations	647	497	416	335	383
R-squared	0.511	0.301	0.504	0.254	0.390

Notes: The table reports OLS estimates. The unit of observation is the Ethnographic Atlas society. The dependent variable in column (1) is an index of artisanal sophistication which captures the types of artisanal crafts performed in the society and ranges from 1 to 5. In column (2), (3), (4) and (5), the dependent variable is a dichotomous variable indicating whether the existing craft is performed by specialized artisans or not. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Geography control variables include: latitude, average agricultural suitability, average annual cumulated temperature above 0 degree, average precipitation and mean elevation, all measured within a 100 km radius of society centroid. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for societies located on islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors, clustered at the linguistic family level (the total number of linguistic family cluster is 68), are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Column (1) of Table 4 shows how the estimated relation between castes and artisanal sophistication is almost half of the one estimated in the SCCS,

however, reassuringly, it is highly statistically significant.
The larger size of the sample allows us to go a step further and explore patterns of labor specialization. In Column (2) to (6) we estimate a linear probability model for each main artisan craft. Having castes increases by around 20% the probability of having pottery and weaving performed by specialized artisans, while up to 40% the same probability for leather-working. We do not observe the same pattern for metal-working specialization and this is likely to be because metal-working tended to be practiced by specialists in the great majority of cases. However, if we were to run the same regressions on the probability of having metal-working practiced within the society, i.e. if we were interested in mere practice of metal work within the society, we would observe a strong and positive correlation.

Robustness Checks We performed a number of robustness checks. First, as previously mentioned, we replicated the baseline estimate with an ordered logit model. Since results are qualitatively unchanged we report estimations in Table 13 of the Appendix. Second, we pursue two strategies to control for the possible bias generated by comparing societies pin-pointed at different points in time: we add the pin-pointing date as a control variable; alternatively, we reduce the sample keeping only societies observed respectively after 1850 and 1900. Again, results, reported in the Appendix, Table 14, remain qualitatively unchanged, with estimated sizes of coefficients even larger in the smaller sub-samples than in the baseline.

4 Identifying a Causal Relationship

The regressions presented above document the existence of a previously unexplored strong correlation between castes and technological sophistication and specialization. This correlation cannot be interpreted causally, though. First, there might be some omitted factors which explain both technology and the presence of castes. For example, we might expect castes to prevail in historically more developed areas which are more technologically advanced to begin with. Second, the existence of a causal effect running from technological sophistication to castes cannot be excluded *a priori*. In fact, it might well be that castes emerged as a by-product of technological sophistication. Castes, just as class stratification, may be consequential to the degree of economic complexity which systematically follows technological innovations. Finally, we cannot exclude that early missionaries and voyagers, from whose diaries anthropologists draw information for coding the variables in the datasets, might had misunderstood some of the institutional features of societies they were visiting, not recognizing or misinterpreting endogamous stratifications[18]. If this is a true concern, the explanatory variable caste would be measured with error, exposing the OLS estimator to attenuation bias.

For untying this thigh knot we rely on three strategies. We first attempt to control for relevant observable society characteristics, mainly in order to exclude that the correlation is driven by well-known patterns of historical development. The second strategy is based on the introduction of an external source of variation in the explanatory variable - caste - which we claim exogenous to the main equation. With this aim, we exploit an early hypothesis put forward by several social scientists that has remained empirically untested, i.e. caste systems might have been an early evolutionary social device adopted by societies for reducing the transmission of diseases in areas facing a high burden of pathogens. Finally, with our last strategy, we try to exclude the existence of a direct effect running from technological sophistication to the presence of castes.

4.1 Controlling for Observables

The literature points at three main determinants of pre-modern development. Ashraf and Galor (2011a), inspired by the insightful perspective of Diamond and Ford (2000), document the effect of the Neolithic revolution - more precisely, of the timing of transition to agriculture - on comparative historical development, measured in terms of population density. Given the well-documented long lasting effects of early development (see Bockstette, Chanda, and Putterman, 2002; Olsson, Hibbs, et al., 2005; Putterman, 2008; Comin, Easterly, and Gong, 2006, among many others), the timing of transition to agriculture is very likely to affect technological sophistication and institutional features of societies in our sample and to act as a possible confounder. Therefore, we imputed the variable year since agricultural transition (from Putterman (2008)[19]) to our data. Moreover, Ashraf and Galor (2011a) further show how the migratory distance from Adis Abeba is a strong predic-

[18]For instance, James Vaughan (1970) (in Essay included in the volume by Tuden, Plotnicov, and of Congress. Hebraic Section) (1970)) documented the presence of endogamous groups within the Margi population in Western Sudan, which was not coded as a society with endogamous stratification by the *Ethnographic Atlas*.

[19]Note that the Putternam index is currently available only at country level and not disaggregated at the society level. This means that societies once localized in what is today the same country are imputed the same year for transition to agriculture. This is likely to attenuate the magnitude of the variable estimated coefficient.

tor of historical development, measured again in terms of population density. The relationship they document is non-monotonic, for this reason we add a measure of Migratory Distance together with its quadratic form. Finally, as technologies can spread, following Ashraf and Galor (2011a) we add a measure of aerial distance to the regional technological frontier. Results, reported in Tabel 5 and 6, imply that none of these variables was driving the results, as coefficients even marginally increase after the introduction of the variables.

Another possible option is to control for other society characteristics, in order to compare societies at similar level of development. The SCCS offers plenty of potentially interesting proxy for development, i.e. a measure of population density, the level of political centralization of the society and an index meafsuring the intensity of agriculture. As regression results reported in 15 of the Appendix show, including all these controls on top of Regressions in Tabel 5 and 6, reduces the magnitude of the caste coefficient by around 20% for crafts and by about 30% for all technologies. Again, the estimated coefficients remain significant at standard statistical levels. Another way to look at these results, it is to exploit the measurable bias from omitting observables characteristics for assessing the bias from unobservable ones, following Altonji, Elder, and Taber (2005)Detailed results and computations are reported in Table 15, 16 and 17 of the Appendix.. By taking as baseline a regression of castes on technological sophistication with latitude as the only control estimated with region fixed effects, we find that, for unobsevables to be driving the results, the overall effects of unobservables on artisanal crafts should 29 times higher than the combined effects from all included controls, and namely all geographical and location controls, population density, agricultural intensity and political centralization, the Ln(Timing of Neolithic Transition), the distance to technological frontier and Migratory Distance and Migratory Distance squared. The same unobserved effect for general technological sophistication should be 3.3 times higher.

In any case, it is important to point that contemporaneous controls are likely to be endogenous to the level of technological sophistication, thus coefficient magnitudes should be assessed with extreme care, as adding endogenous variable would bias the size of our coefficient of interest.

Table 5: Castes and Technological Sophistication

	Technological Sophistication					
	Artisanal Craft			All Technologies		
	(1)	(2)	(3)	(4)	(5)	(6)
CASTES	0.936***	0.930***	0.994***	1.046**	1.092**	1.179**
	(0.280)	(0.282)	(0.294)	(0.515)	(0.498)	(0.493)
Ln(Timing of Neolithic Trans.)	0.497**	0.177	0.267	0.950***	0.343	0.728*
	(0.196)	(0.208)	(0.226)	(0.326)	(0.424)	(0.428)
Distance to Techn. Frontier		-0.339***	-0.347***		-0.545***	-0.619***
		(0.078)	(0.078)		(0.152)	(0.159)
Migratory Distance			0.119			0.332**
			(0.111)			(0.130)
Migratory Distance square			-0.004			-0.017***
			(0.005)			(0.005)
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes
Location Controls	Yes	Yes	Yes	Yes	Yes	Yes
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	150	150	150	144	144	144
R-squared	0.447	0.498	0.503	0.475	0.528	0.551

Notes: The table reports OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. The variable Ln(Timing of Neolithic Transition) is the natural log of the number of years passed since the area where the society is located went through the transition to agriculture. Distance to technological frontier is the distance, in thousands of km, from the closest regional technological frontier. Migratory Distance and Migratory Distance squared are, respectively, the migratory distance (in thousand of km on a land path) and square migratory distance from Adis Abeba. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table 6: Castes and Technological Specialization

	Technological Sophistication	Craft Specialization			
		Pottery	Leather	Weaving	Metal
	(1)	(2)	(3)	(4)	(5)
CASTES	0.629**	0.200***	0.404***	0.168**	0.023
	(0.258)	(0.047)	(0.086)	(0.074)	(0.014)
Ln(Timing of Neolithic Trans.)	Yes	Yes	Yes	Yes	Yes
Distance to Techn. Frontier	Yes	Yes	Yes	Yes	Yes
Migratory Distance	Yes	Yes	Yes	Yes	Yes
Migratory Distance square	Yes	Yes	Yes	Yes	Yes
Geography Controls	Yes	Yes	Yes	Yes	Yes
Location Controls	Yes	Yes	Yes	Yes	Yes
Continent FE	Yes	Yes	Yes	Yes	Yes
Observations	643	493	413	332	379
R-squared	0.594	0.313	0.524	0.311	0.393

Notes: The table reports OLS estimates. The unit of observation is the Ethnographic Atlas society. The dependent variable in column (1) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (2), (3), (4) and (5), the dependent variable is a dichotomous variable indicating whether the existing craft is performed by specialized artisans or not. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. The variable Ln(Timing of Neolithic Transition) is the natural log of the number of years passed since the area where the society is located went through the transition to agriculture. Distance to technological frontier is the distance, in thousands of km, from the closest regional technological frontier. Migratory Distance and Migratory Distance squared are, respectively, the migratory distance (in thousand of km on a land path) and square migratory distance from Adis Abeba. Geography control variables include: latitude, average agricultural suitability, average annual cumulated temperature above 0 degree, average precipitation and mean elevation, all measured within a 100 km radius of society centroid. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for society located on islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors, clustered at the linguistic family level (the total number of linguistic family cluster is 68), are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

bose

4.2 Caste Systems and Disease Exposure

The impossibility to fully account for all possible confounders that may drive the relation between castes and technology, together with the necessity to exclude a reverse effect of technology on caste stratification, motivate us for pursuing a second alternative strategy. Relying on anthropological theories proposing castes as an evolutionary social device for reducing disease transmission, we exploit an index of disease exposure as an instrument for the presence of castes.

Anthropological Background The paternity of the idea that links castes to diseases should be accredited, in our knowledge, to historian William McNeill (1976). In the well-known book "Plagues and Peoples", he argued:

> "the caste organization of Indian civilization may have partly been a response to the kind of epidemiological standoff that arose when intrusive Aryans, who had probably learned to live with some acute "civilized" diseases - e.g. perhaps smallpox - encountered various "forest folks" who had acquired tolerance for formidable local infections ... therefore " ... instead of digesting the various primitive communities that had occupied southern and eastern India... Indian civilization expanded incorporating ex-forest folks as castes".

More broadly speaking, the idea of the existence of a relationship between the disease environment and social norms, habits and practices that prevail in a given society is not new. Evolutionary anthropologist Alland (1970) claimed that "Social isolation of various subgroups is also important to disease ecology. A rigid caste system with minimal social and physical contact between groups may affect epidemic routes in community". Along the same line, medical anthropologist Cockburn (1971) was pointing that societies evolve "pragmatic means of sanitation that, whatever the intended religious or magical purpose, may in fact function to protect health and ward of diseases". More recently, evolutionary psychologists Curtis, Aunger, and Rabie (2004); Curtis, De Barra, and Aunger (2011) argued, very much in line with McNeill's ideas, that "The Law of Manu (one of the major Hindu holy books) prescribes... avoidance of polluted caste. These social distinctions may have their origins in biological avoidance strategies... through the emotional resources of disgust and contagion". Not referring precisely to castes but to ethnic fractionalization in general, Cashdan (2001) added that "parasites are like a wedge driving groups apart through their effective creation of anti-contagion behaviors". Finally, interesting complementary evidence is emerging from recent studies in genetics. Brahmajothi, Pitchappan, Kakkanaiah, Sashidhar, Rajaram, Ramu, Palanimurugan, Paramasivan, and Prabhakar (1991) and Siddiqui, Meisner, Tosh, Balakrishnan, Ghei, Fisher, Golding, Narayan, Sitaraman, Sengupta, et al. (2001) found that isolated caste and sub-caste populations of southern India differ in their HLA (Human leukocyte antigen) and other immune repertoire for diseases such as tuberculosis and leprosy. With this evidence in mind, some early historical descriptions of the treatment reserved to Indian low castes can be read under a new insightful perspective:

> "there were castes that had to carry bells just as medieval lepers... contact with them by sight or by the passage of a breath of air necessitated ceremonial purification[20]".

Given the previously mentioned similarities of Indian caste system and caste systems found in other geographical areas along dimensions such as the purity-pollution concept or the taboo regarding commensality, we test whether castes prevailed in environments burdened by higher disease exposure. Our hypothesis is that a highly diseased environment might had induced societies to create social separations in order to limit the spread of contagious diseases, which could otherwise had spread and thrived taking advantage of the large size of communities. The high burden of diseases might also had prevented different neighboring ethnic groups, adapted to different types of diseases, to merge by promoting norms of social distance. Even when the migrating groups had adapted to the new disease environment, the social norms imposing an endogamous stratification of society might had persisted[21].

An Exogenous Measure of Disease Exposure For testing whether caste systems prevailed in regions facing a heavier burden of diseases, we need a proxy for measuring disease exposure. There are several methodological difficulties which complicate our search for a suitable instrument. First of all, the risk of endogeneity prevent us from using classical epidemiological measure of disease exposure, i.e. people at risk, number of new cases per year etc... The problem is that pathogens' effects are highly dependent

[20] Hutton (1946).

[21] The idea that diseases might affect long-term economic outcomes by shaping social institutions have recently received much attention by economists. Birchenall (2010), following anthropological evidence, find a significant and positive effect of disease exposure on the level of ethnic fractionalization in Sub-Saharan Africa. Birchenal claims that ethnic fractionalization slowed down the process of state centralization in Sub-Saharan Africa. Fogli and Veldkamp (2011), relying on recent findings in evolutionary psychology, argue that geographical areas dealing with higher disease hazards developed more collectivist cultures, with the implicit aim of promoting "safe" relations within the groups and limiting contacts with external potential source of contagion. Fogli and Veldkamp (2011) claim that the lack of individualism reduced the spread of innovations and was at the root of long run growth stagnation of countries with more collectivist values.

on several social and economic variables related to the level of development, such as the quality of the diet, the sanitary habits and the type of household just to name a few. Second, several diseases, and namely those carried and transmitted by humans, tend to spread, it is therefore intrinsically difficult to obtain a measure of exogenous and time-invariant geographical disease exposure. Third, we would ideally need an index that captures disease exposure at least as early in time as societies were pin-pointed. Of course, we do not have information regarding the historical presence of diseases except for few diseases in a few number of countries.

In order to circumvent the methodological problems and in line with previous works (Murray and Schaller, 2010; Cervellati, Sunde, and Valmori, 2011b), we implement the following strategy: i) we rely on information retrieved from historical epidemiological Atlas from early 20th century (before the epidemiological revolution of the 40's); ii) among the diseases that we combine in the index, we rely mainly on vector diseases, i.e. diseases that are transmitted through a vector whose existence in a area strictly depends on exogenous climatic-environmental conditions; iii) we recode the index of intensity on a two point scale, in order to have it equal to 1 if the disease is present in the area where the society live, and 0 if not, for reducing the potential endogeneity of the variable disease intensity[22]. The SCCS provides an index of disease exposure for seven diseases, namely Leishmaniasis, Trypanosomes, Malaria, Schistosomes, Filariae, Spirochetes and Leprosy - note that 5 of the seven diseases are vector-transmitted[23].

It can be argued that the choice of employing vector diseases, despite guaranteeing exogeneity, might not serve the purpose of explaining the presence of castes, as vector diseases do not spread by human contact but are spread by an intermediate agent (mosquito, fly, etc...) which could even infect people living relatively far apart. However, there are two sets of reason why focusing mainly on vector-transmitted diseases can be an appropriate choice. First of all, epidemiological studies repeatedly shown that different diseases

[22]In fact, in a pre-modern context, human vector diseases such as tuberculosis or leprosy may prevail in highly populated societies, more than among hunters living spread in the forest. Therefore, the difference in disease intensity may be explained by the population density and not by exogenous climatic and environmental conditions.

[23]The index was made available by Low (1988), which collected the data for her study linking diseases to polygyny. In the SCCS, there is another index of disease exposure, made available by Ludvico Ludvico and Kurland (1995), who exploited the index for studies on human scarification practices.). This index is composed by 12 diseases. We exploited the second index for robustness checks: first, we constructed an alternative index for the seven main diseases of the baseline index (Leishmaniasis, Trypanosomes, Malaria, Schistosomes, Filariae, Spirochetes and Leprosy); second, we constructed an alternative index composed by all 12 diseases. Results remain qualitatively unchanged.

are "complementary shocks", in the sense that, for instance, having heavy malaria weakens the immunitary system and raise the risk of contracting other contagious diseases. Moreover, misconceptions regarding the way vector diseases are transmitted is vague, even today. For instance, interestingly for our analysis, Ramaiah, Kumar, and Ramu (1996) points that a share of the people interviewed in Tamil Nadu, India, (about 7%) still believed that filariasis can be contracted by the action of weaving.

Instrumental variable estimates are unbiased if the instrument, disease exposure, affects technological sophistication only through the presence of castes. In other words, disease exposure should not affect technological sophistication and artisanal specialization in pre-modern societies beyond the channel of caste stratification. In the economic literature the effects of diseases, through life expectancy, on economic performance of countries have been largely investigated. A large amount of contributions finds that increases in life expectancy cause increases in country income per capita (see Gallup, Sachs, and Mellinger, 1998; Bloom, Canning, and Sevilla, 2003; Lorentzen, McMillan, and Wacziarg, 2005; Shastry and Weil, 2003; Weil, 2007; Ashraf, Lester, and Weil, 2008, among others). Acemoglu and Johnson (2006) reached an opposite conclusion, arguing that positive changes in mortality raise population growth but decrease income per capita. Cervellati and Sunde (2009) shows that the effects of higher life expectancy on income per capita tend to be negative before the demographic transition, but highly positive after the country transitions. Regarding the individual effects of health, diseases and high mortality rates have been found to negatively affect workers' productivity and days of work and to reduce investments in schooling (Weil, 2007; Bleakley, 2006; Ashraf, Fink, and Weil, 2010, just to mention a few).

On the contrary, the historical effects of diseases on development and, in particular, on technological sophistication have not been subject to empirical investigations. Acemoglu, Johnson, and Robinson (2001) argued that the impact of a diseased environment on the economic development of nations is not due to a direct effect of health conditions on income but passes through the channel of institutional quality settled at the time of colonization. On the other hand, historian McNeill (1976) argued that higher disease exposure historically represented a serious burden for societies because of the energy drain on individuals it engendered.

For our purposes, it is not obvious to see how lower productivity (days of work) and scarcer energies could had prevented societies from developing artisanal crafts like pottery, weaving and metalworking. A possible channel linking a diseased environment with lack of labor specialization runs through the potential population shocks which follows sudden epidemics and which could reduce demands and decrease specialization. However, the high level of technological sophistication and labor specialization historically reached in areas widely exposed to diseases, like the ancient civilizations of Southern India, the Egyptians or the Ghana Empire, seems to contrast this possibility. Moreover, if a direct relation between geographical disease exposure and technology exists, according to the literature this should affect technological sophistication negatively. For this reason, since we expect disease to increase caste stratification and caste to foster technology, our estimate of the effect of castes on technological sophistication would be biased downward.

IV estimates In Table 7, we present instrumental variables estimates of the effect of castes on technological sophistication[24]. Since results do not change, for brevity, we report only the first sets of estimation results.

[24]As we have a dichotomous endogenous variable, we proceed in two different ways. First, we estimate a linear first and second stage, which generates consistent second-stage estimates even with a dichotomous endogenous variable (Angrist and Pischke, 2008). Another possible way is to estimate a non-linear first stage, and to instrument the endogenous variable with the fitted values of the non-linear first stage. The second procedure would enhance efficiency in the estimation as far as the non-linear model better approximates the CEF in the first stage (Angrist and Pischke, 2008).

Table 7: IV for Castes and Technological Sophistication

	Technological Sophistication					
	Artisanal Craft			All Technologies		
	(1)	(2)	(3)	(4)	(5)	(6)
CASTES	4.036**	4.162**	2.976*	6.880**	7.068**	4.850*
	(1.752)	(1.836)	(1.614)	(3.035)	(3.251)	(2.644)
	[1.47 , 15.88]	[1.50, 19.05]	[0.41, 24.13]	[2.65, 31.06]	[2.57, 40.56]	[0.91, 51.99]
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes
Location Controls		Yes	Yes		Yes	Yes
Population Density			Yes			Yes
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	159	159	159	152	152	152
	First Stage Statistics					
Disease Index	0.057**	0.055**	0.050**	0.056**	0.053**	0.050**
	(0.022)	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)
Partial R-squared	0.043	0.040	0.032	0.042	0.038	0.032
F Stat	6.64	6.02	4.86	6.17	5.51	4.61

Notes: IV estimates are reported. The unit of observation is the Standard Cross Cultural Sample society. In the upper panel we report the second stage estimated coefficients. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if in the society are present endogamous groups and 0 otherwise. The excluded instrument is an index of geographical disease exposure ranging from 0 to 7. Geography controls include: latitude, average temperature of the coldest month and average rainfall. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Second stage coefficients are reported with robust standard errors in brackets, 95% confidence regions based on Moreira's (2003) conditional likelihood ratio (CLR) approach are reported in squared brackets. In the lower panel, the first stage coefficient estimates are reported. First stage statistics F statistics and Partial R-squared are also reported. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

As shown in Column (1) and (2) of 7, the instrumental variable estimated effect of caste on artisan technological sophistication is about 5 time higher than the OLS one in the first and second specification. For what concern general technological sophistication, the IV coefficients is about 7 times larger. In specification (3) and (6), we propose a third specification where we include population density, this is in order to reassure that violations of the exclusion restrictions are not driving the results. However, since population density might be endogenous to the level of technological sophistication, the estimated coefficient cannot be directly compared with the baseline ones.

The wide difference in the size of OLS and IV estimated coefficients may result from the combination of several effects. The higher magnitude of the coefficients suggests that omitted variables might have weakened the estimated effects of castes on technological sophistication. Moreover, there might be attenuation biases caused by imperfections in the data source. On top of this, it is worth pointing that the OLS and IV estimators measure the effect of castes on technologies for two different sub-groups. OLS measures the effect on technologies for all societies which are recorded to have endogamous stratification. On the other hand, IV estimates the effect of castes on technological sophistication for those societies whose caste stratification, in our data, is related to the adaptive relationship with a highly diseased environment. We have no way to say whether all societies with endogamous stratification generated such institutions as a consequence of disease exposure.

It is crucial to note, as F Statistics Reported in Table 7 suggests, that we might be dealing with a weak instrument. Weak instruments generate problems both for what concern the coefficient estimation, which risk to be biased in the same direction of the OLS' one, and the size of the confidence intervals for inference, that tend to be too narrow and lead to an underejection of the null[25]. In order to find support for our estimates, we undertake two strategies: i) we follow Moreira (2003) and compute confidence intervals which take into account the presence of a weak instrument, these are reported in square brackets in Table 7; iii) following suggestion by (Angrist and Pischke, 2008), we report reduced form estimates of the effect of diseases on technology, as the OLS estimator of the reduced form is always unbiased. Results are reported in the Appendix.

Reassuringly, the 95% confidence regions based on Moreira's (2003) conditional likelihood ratio (CLR) approach are confined in the positive axis. As you can see from Column (1), (2), (4) and (5), the lower bound for the IV estimates is always higher than the OLS estimating coefficients, suggesting that OLS indeed bias the effect downwards. Moreover, the reduced form relation (reported in Table 18 of the Appendix) between our disease index and technological sophistication confirm the existence a positive and significant effect of diseases on technology which, in our view, can possibly be attributed

[25]See Murray (2006a,b) for insightful reviews of best practices when dealing with weak instrument.

to no other channel other than the one of caste stratification.

To conclude, despite the fact that, given possible weak instrument issues, our IV coefficients may not be precisely estimated, weak-inference robust confidence intervals confirm a positive effect of caste stratification on technological sophistication.

4.3 Excluding Reverse Causality

Our main argument for excluding a reverse effect of technological sophistication over the likelihood of observing castes, it is to claim that castes are a pre-existing form of stratification. Therefore, our final attempt is to show that caste stratification is not the results of social complexity, which might derive from technological advancements. In order to do so, we exploit a well-established exogenous determinants of technological sophistication, i.e. the timing of transition to agriculture, and show how societies in areas that transitioned earlier to agriculture tend to have higher levels of technological sophistication, present a more structured division in social classes, but are not more likely to have caste stratification.

The Ethnographic Atlas and the SCCS report the presence and on the type of social classes reported within the society. The Atlas codes several type of social classes[26] but the coding offer no straightforward ordering. Our working hypothesis is that technological advances had generated new professional figures, and related new occupational positions, as well as major wealth distinctions between different professions. We create a variable $CLASS_i$ which takes value 1 for societies that present no sorts of class stratification (among freeman, as slavery is another matter); 2 for societies having a dual stratification into a hereditary aristocracy and a lower class of ordinary commoners or freeman, a landlord versus landless class; we assign value 3 to the index for societies with complex stratification into social classes and to societies which present wealth distinctions based on possession and distribution[27].

[26]Variables codes are: Absence of class stratification among freeman; Complex stratification into social classes correlated in large measure with extensive differentiation of occupational statuses; Dual stratification into a hereditary aristocracy and a lower class of ordinary commoners or freeman; Landlord versus landless class and Wealth distinction, based on possession and distribution

[27]Note that results do not rest on this coding. By considering wealth distinction as an intermediate type of stratification, we would obtain similar results

Table 8: Neolithic Transition, Technology, Classes and Castes

	Dependent Variables					
	Artisanal Craft		Classes		Castes	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Time of Neolithic Trans.)	0.241***	0.221**	0.179***	0.148***	-0.011	-0.009
	(0.086)	(0.088)	(0.058)	(0.055)	(0.023)	(0.022)
Absolute Latitude	Yes	Yes	Yes	Yes	Yes	Yes
Geography Controls		Yes		Yes		Yes
Location Controls		Yes		Yes		Yes
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	684	684	1,003	1,003	997	997
R-squared	0.666	0.669	0.160	0.214	0.302	0.337

Notes: The table reports reduced form OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (3) and (4) the dependent variable is an index capturing the presence of social classes within the society. In column (5) and (6) the dependent variable is a dummy variable that takes value 1 if in the society are present endogamous groups and 0 otherwise. Geography controls include: mean elevation and average rainfall. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Absolute latitude and continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Results reported in Table 8 show, consistently with our hypothesis suggesting that castes are pre-existing forms of stratification, that the timing of Neolithic transition is positively correlated with both the technological sophistication index and the social classes index, however, there is no correlation with caste stratification.

5 Conclusion

We tested the hypothesis that caste systems positively affected technological sophistication and specialization at an early stage of development by "artificially" spurring division of labor in pre-modern societies. We used data from Murdock's *Ethnographic Atlas* and Standard Cross Cultural Sample and found a robust and strong relationship between the presence of castes and two indexes of basic technological sophistication, the first index measuring artisan craft sophistication and the second general technological sophistication. Moreover, we found that castes increased the probability of having occupations performed by specialists. The correlations are robust to all major determinants of pre-modern development. In order to give a causal interpretation to the observed results, we exploit an exogenous variation in the presence of caste systems, i.e. the geographical disease exposure of the area in which the society is located. In fact, various historians and anthropologists have been suggesting that castes represented a social device for reducing contagion in highly diseased environments. We built an index of geographical disease exposure conceived to capture only the component of disease exposure related to exogenous bio-climatic characteristics of the geographical area. Exploiting disease exposure as an instrument for castes, we identified a positive and causal effect of caste systems on technological sophistication and labor specialization. Indeed, our IV estimates of the effect of caste on technology confirm the positive effect of caste systems on technological sophistication and labor specialization in pre-modern societies.

References

ACEMOGLU, D., AND S. JOHNSON (2006): "Disease and development: the effect of life expectancy on economic growth," .

ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5), 1369–1401.

AIYAR, S., C. DALGAARD, AND O. MOAV (2008): "Technological progress and regress in pre-industrial times," *Journal of Economic Growth*, 13(2), 125–144.

ALESINA, A., AND E. FERRARA (1999): "Participation in heterogeneous communities," .

——— (2004): "Ethnic diversity and economic performance," .

ALESINA, A., P. GIULIANO, AND N. NUNN (2011a): "Fertility and the Plough," *American Economic Review*, 101(3), 499–503.

ALESINA, A., P. GIULIANO, AND N. NUNN (2011b): "On the origins of gender roles: Women and the plough," .

ALLAND, A. (1970): *Adaptation in cultural evolution: an approach to medical anthropology*. Columbia University Press New York.

Altonji, J., T. Elder, and C. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 113(1), 151–184.

Angrist, J., and J. Pischke (2008): *Mostly harmless econometrics: An empiricist's companion*. Princeton Univ Pr.

Ashraf, N., G. Fink, and D. Weil (2010): "Evaluating the effects of large scale health interventions in developing countries: The zambian malaria initiative," .

Ashraf, Q., and O. Galor (2011a): "Dynamics and Stagnation in the Malthusian Epoch," *The American Economic Review*, 101(5), 2003–2041.

——— (2011b): "The" Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," .

Ashraf, Q., and O. Galor (2013): "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103(1), 1–46.

Ashraf, Q. H., A. Lester, and D. N. Weil (2008): "When Does Improving Health Raise GDP?," (14449).

Bamshad, M., T. Kivisild, W. Watkins, M. Dixon, C. Ricker, B. Rao, J. Naidu, B. Prasad, P. Reddy, A. Rasanayagam, et al. (2001): "Genetic evidence on the origins of Indian caste populations," *Genome research*, 11(6), 994–1004.

Banerjee, A., L. Iyer, and R. Somanathan (2005): "History, social divisions, and public goods in rural India," *Journal of the European Economic Association*, 3(2-3), 639–647.

Barth, F. (1956): "Ecologic relationships of ethnic groups in Swat, North Pakistan," *American Anthropologist*, 58(6), 1079–1089.

Berreman, G. (1968): "Stratification, Pluralism and Interaction: a comparative analysis of caste," in *Ciba Foundation Symposium-Caste and Race: Comparative Approaches*, pp. 45–73. Wiley Online Library.

Birchenall, J. (2010): "Disease and diversity in Africa's Long-term Economic Development," .

Bleakley, H. (2006): "Malaria In The Americas: A Retrospective Analysis Of Childhood Exposure," (003185).

Bloom, D., D. Canning, and J. Sevilla (2003): *The demographic dividend: a new perspective on the economic consequences of population change*, vol. 5. Rand Corp.

Bockstette, V., A. Chanda, and L. Putterman (2002): “States and markets: The advantage of an early start,” *Journal of Economic Growth*, 7(4), 347–369.

Bose, P. (1916): *The Illusions of New India*. W. Newman.

Brahmajothi, V., R. Pitchappan, V. Kakkanaiah, M. Sashidhar, K. Rajaram, S. Ramu, K. Palanimurugan, C. Paramasivan, and R. Prabhakar (1991): “Association of pulmonary tuberculosis and HLA in South India,” *Tubercle*, 72(2), 123–132.

Cashdan, E. (2001): “Ethnic diversity and its environmental determinants: effects of climate, pathogens, and habitat diversity,” *American Anthropologist*, 103(4), 968–991.

Cervellati, M., and U. Sunde (2009): “Life expectancy and economic growth: the role of the demographic transition,” *Journal of Economic Growth*, pp. 1–35.

Cervellati, M., U. Sunde, and S. Valmori (2011a): “Disease Environment and Civil Conflicts,” (5614).

——— (2011b): “The Distribution of Infectious Diseases and Extrinsic Mortality Across Countries,” .

Cockburn, T. (1971): “Infectious diseases in ancient populations,” *Current Anthropology*, 12(1), 45–62.

Comin, D., W. Easterly, and E. Gong (2006): “Was the Wealth of Nations Determined in 1000 bc?,” .

Curtis, V., R. Aunger, and T. Rabie (2004): “Evidence that disgust evolved to protect from risk of disease,” *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(Suppl 4), S131–S133.

Curtis, V., M. De Barra, and R. Aunger (2011): “Disgust as an adaptive system for disease avoidance behaviour,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1563), 389.

Diamond, J., and L. Ford (2000): “Guns, germs, and steel: the fates of human societies,” *Perspectives in Biology and Medicine*, 43(4), 609.

Diop, C., and H. Salemson (1987): *Precolonial Black Africa: a comparative study of the political and social systems of Europe and Black Africa, from antiquity to the formation of modern states*. L. Hill.

Dumont, L. (1980): *Homo hierarchicus: the caste system and its implications*. University of Chicago Press.

Easterly, W., and R. Levine (1997): "Africa's growth tragedy: policies and ethnic divisions," *The Quarterly Journal of Economics*, 112(4), 1203.

Epstein, S. (1998): "Craft guilds, apprenticeship, and technological change in preindustrial Europe," *Journal of Economic History*, 58, 684–713.

Fearon, J., and D. Laitin (2003): "Ethnicity, insurgency, and civil war," *American Political Science Review*, 97(1), 75–90.

Fincher, C., R. Thornhill, D. Murray, and M. Schaller (2008): "Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism," *Proceedings of the Royal Society B: Biological Sciences*, 275(1640), 1279.

Fogli, A., and L. Veldkamp (2011): "Germs, social networks and growth," .

Freitas, K. (2006): "The Indian caste system as a means of contract enforcement," mimeo Northwestern University. Paper presented at the NEUDC conference 2006, Cornell University.

Gallup, J., J. Sachs, and A. Mellinger (1998): "Geography and economic development," .

Greif, A. (1993): "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition," *American economic review*, 83, 525–525.

——— (1994): "Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies," *Journal of political economy*, pp. 912–950.

——— (2006): *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge Univ Pr.

Guiso, L., P. Sapienza, and L. Zingales (2008): "Long term persistence," .

Hoff, K., P. Pandey, and W. B. D. R. G. I. C. Team (2004): *Belief systems and durable inequalities: an experimental investigation of Indian caste*, vol. 3351. World Bank, Development Research Group, Investment Climate.

Hutton, J. (1946): *Caste in India*. University Press.

Iliffe, J. (1987): *The African poor: A history*, vol. 58. Cambridge Univ Pr.

Lorentzen, P., J. McMillan, and R. Wacziarg (2005): "Death and development," .

Low, B. (1988): "Pathogen stress and polygyny in humans," *Human reproductive behaviour: A Darwinian perspective*, pp. 115–128.

Ludvico, L., and J. Kurland (1995): "Symbolic or not-so-symbolic wounds: The behavioral ecology of human scarification," *Ethology and Sociobiology*, 16(2), 155–172.

McNeill, W. (1976): "Plagues and Peoples. 1976," *Garden City, NY: Anchor P.*

Michalopoulos, S. (2008): "The Origins of Ethnolinguistic Diversity: Theory and Evidence," (11531).

Moreira, M. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71(4), 1027–1048.

Mountain, J., J. Hebert, S. Bhattacharyya, P. Underhill, C. Ottolenghi, M. Gadgil, and L. Luca Cavalli-Sforza (1995): "Demographic history of India and mtDNA-sequence diversity," *American journal of human genetics*, 56, 979–992.

Murdock, G. (1967): "Ethnographic atlas: A summary," *Ethnology*, 6(2), 109–236.

Murray, D., and M. Schaller (2010): "Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture," *Journal of Cross-Cultural Psychology*, 41(1), 99–108.

Murray, M. (2006a): "Avoiding invalid instruments and coping with weak instruments," *The Journal of Economic Perspectives*, 20(4), 111–132.

Murray, M. (2006b): "The bad, the weak, and the ugly: Avoiding the pitfalls of instrumental variables estimation," .

North, D. (1990): *Institutions, institutional change, and economic performance*. Cambridge Univ Pr.

Nunn, N., and L. Wantchekon (2009): "The slave trade and the origins of mistrust in Africa," .

Olsson, O., D. Hibbs, et al. (2005): "Biogeography and long-run economic development," *European Economic Review*, 49(4), 909–938.

Putterman, L. (2008): "Agriculture, diffusion and development: Ripple effects of the neolithic revolution," *Economica*, 75(300), 729–748.

Ramaiah, K., K. Kumar, and K. Ramu (1996): "Knowledge and beliefs about transmission, prevention and control of lymphatic filariasis in rural areas of South India," *Tropical Medicine & International Health*, 1(4), 433–438.

Sachs, J. (2001): "Tropical underdevelopment," .

——— (2003): "Institutions don't rule: direct effects of geography on per capita income," .

Santacreu-Vasut, E. (2009): "Diversity and Institutions: Theory and Evidence from the Spread of Industrialization," .

Shack, W. (1964): "54. Notes on Occupational Castes Among the Gurage of South-West Ethiopia," *Man*, 64(1), 50–52.

Shastry, G., and D. Weil (2003): "HOW MUCH OF CROSS-COUNTRY INCOME VARIATION IS EXPLAINED BY HEALTH?," *Journal of the European Economic Association*, 1(2-3), 387–396.

Siddiqui, M., S. Meisner, K. Tosh, K. Balakrishnan, S. Ghei, S. Fisher, M. Golding, N. Narayan, T. Sitaraman, U. Sengupta, et al. (2001): "A major susceptibility locus for leprosy in India maps to chromosome 10 p 13," *Nature genetics*, 27(4), 439–441.

Sokoloff, K., and S. Engerman (2000): "History lessons: Institutions, factors endowments, and paths of development in the new world," *The Journal of Economic Perspectives*, 14(3), 217–232.

Tabellini, G. (2007): "Culture and institutions," *CEPR Discussion Paper No. DP6589.*

Tamari, T. (1991): "The development of caste systems in West Africa," *Journal of African History*, 32(2), 221–50.

Thapar, R. (2004): *Early India: From the origins to AD 1300*. Univ of California Pr.

Tuden, A., L. Plotnicov, and T. L. K. C. L. of Congress. Hebraic Section) (1970): *Social stratification in Africa*. The Free Press New York.

Weber, M. (1958): "The Religion of India: The Sociology of Hinduism and Buddhism, trans," *HH Gerth and Don Martindale (Glencoe, Ill.: The Free Press, 1958)*, p. 207.

Weil, D. N. (2007): "Accounting for The Effect of Health on Economic Growth," *The Quarterly Journal of Economics*, 122(3), 1265–1306.

6 Appendix

6.1 Data Sources

The analysis is performed relying on two main sources of data: the Standard Cross Cultural Sample and the Ethnographic Atlas. When necessary, additional data were joined to the existing dataset through a mapping system.

6.1.1 Standard Cross Cultural Sample

Technological Sophistication: Craft Specialization SCCS variable $v153$, captures the degree of complexity and specialization in technological crafts. The variable ranges from 1 to 5: 5 is attributed to societies reported to have a variety of craft specialists, including at least smiths, weavers, and potters; 4 to societies reported to have specialized metalworkers or smiths, but to lack loom-weaving and or pottery; 3 wherever loom weaving is practiced but metalwork in absent or unreported; 1 to societies where pottery is made but metalworking and loomweaving are absent or unreported; and 0 to societies where metalworking, loom-weaving and pottery are all absent or unreported. *Source: Murdock, George P., and Caterina Provost. 1971. ETHNOLOGY 12:379-392.*

Technological Sophistication: All Technologies . Index of technological sophistication of the society along several dimensions: agriculture, artisanal crafts, writing and land transportation. Composed aggregating four index constructed through the recoding of information available in the SCCS. Agriculture index (SCCS variable $v3$ was recoded on a three point scale): 1 for societies where agriculture provide no food contribution; 2 for societies where agriculture contributes up to 50% of the food supply; 3 to societies relying primarily on agriculture. Craft index (constructed based on SCCS variable $v248$ and $v251$): 1 whenever neither metalwork nor pottery is present; 2 to societies which practice pottery (and not metal); 3 to societies which practice both pottery and metal. The presence/absence of pottery and metal was retrieved indirectly through information on the sexual division of labor; i.e. metal/pottery were considered absent if the activity was absent or not reported. Writing index (constructed based on SCCS variable $v149$): 1 if no writing and records are reported; 2 if non-written records and mnemonic devices are reported; 3 if true writing is reported. Land transportation index (constructed based on SCCS variable $v154$): 1 to societies where no technologies for land transportation is reported; 2 for societies which exploits packed/draft animals; 3 for societies with vehicles. The four indexes were aggregated into a single index, thus ranging from 0 to 12. *Source: variable* $v3$*,* $v149$ *and* $v154$*, Subsistence Economy and Supportive Practices George P. Murdock and Diana O. Morrow. 1970. ETHNOLOGY 9:302-330. Variable* $v248$ *and* $v251$*, Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998.*

Castes The variable CASTE takes value 1 if it is reported at least one type of caste stratification, and 0 if not. It is based on information retrieved from SCCS variable $v272$, which classify societies in 4 categories based on the forms of endogamous stratification they present, and namely 1) Caste distinction absent or insignificant; 2) Ethnic stratification; 3) Despised Occupational Groups; 4) Complex caste stratification. The baseline variable CASTE include all three types of caste stratification (2, 3 and 5). The variable CASTE1, include only complex caste stratification, and endogamous occupational groups. *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998.*

Latitude . Latitude of society centroid, measured in degrees. SCCS variable *latitude*. *Source: Standard Cross Cultural Sample (Murdock, George Peter and Douglas R. White. (1969)).*

Mean Temperature . Mean Temperature of the coldest month. SCCS variable $v188$, *Source: Walter, H., and H. Leith (1964) Klimadiagramm-Weltatlas, Jena: Gustav Fischer.*

Mean Rainfall . Average Annual Rainfall, coded from 0 to 8 based on annual mm of rainfall. SCCS variable $v929$. Source: World Meteorological Organization (1971) Climatological Norms (CLINO) for Climate and Climate Ship Stations for the period 1931-1960. Geneva.

Agricultural Potential . Index of agricultural potential of the land: sum of Land Slope, Soils and Climate Scales. SCCS variable $v929$. *Standard Cross Cultural Sample (Murdock, George Peter and Douglas R. White. (1969)).*

Sea Distance Distance of society centroid to the closest sea coast, measured in thousands of Km. Computed through a geographical mapping system, based on shorelines maps. *Source: Wessel, P. and Smith, W.H.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res., 101(B4): 87418743. http://www.soest.hawaii. edu/wessel/gshhs/*

River Distance . Distance of society centroid to the closest river, measured in thousands of Km. Computed through a geographical mapping system, based on shorelines maps. *Source: World Water Bodies Dataset. http://www.arcgis.com/home/item.html?id=e750071279bf450cbd510454a8 0f2e63*

Island Dummy . Dichotomous variables computed measuring the size of the land mass where the society centroid is located. All land masses with an area smaller than Australia were considered islands. *Source: Wessel, P. and Smith, W.H.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res., 101(B4): 87443. Available at: http://www.soest.hawaii.edu/wessel/gshhs/*

Ln(Timing of Transition to Agriculture) Natural logarithm of the timing to the transition to agriculture, from Putterman (2008). Since data are available only at country a level, each society was matched to the modern country based on geographical location. For some societies, timing of Neolithic transition of the modern country is not available, and namely: Bahamas, Eritrea, Fiji, French Polynesia, Kiribati, Marshall Islands, Micronesia, New Caledonia, Palau, Samoa, Solomon Islands and Vanuatu.

Distance to Technological Frontier Aerial distance (Haversine formula), in thousands of km, to the regional technological frontier in 1500 as in Ashraf and Galor (2011a), and namely London and Paris in Europe, Fez and Cairo in Africa, Constantinople and Peking in Asia, and Tenochtitlan and Cuzco in the Americas.

Migratory Distances Migratory distance, on a land path, from Adis Ababa. Computed following **?**. The distance of the society centroid from Adis Abeba is computed using the Haversine formula. In order to replicate the most likely migration pattern followed by early men, we calculated the distance from Adis Ababa of the path that connect several obligatory intermediate points, and namely: Cairo, Istambul, Phnom Phen, Anadyr and Prince Rupert.

Agricultural Intensity Index capturing the fraction of food provided through agriculture, SCCS variable $v3$ *Source: Subsistence Economy and Supportive Practices, George P. Murdock and Diana O. Morrow. 1970. ETHNOLOGY 9:302-330.*

Population Density Approximate number of people per square male, coded on a 7 point scale. SCCS variable $v64$ *Source: George P. Murdock and Suzanne F. Wilson. 1972. ETHNOLOGY 11: 254-295.*

Political Integration Number of political jurisdictions above the local. SCCS variable $v157$ *Source: Murdock, George P., and Caterina Provost. 1971. ETHNOLOGY 12:379-392.*

Disease Index Constructed recombining SCCS variables $v1253 - v1259$. Originally, variables coded on a three point scale: "Absent or not recorded", "Present, no indication of severity", "Present and serious, widespread, or endemic". In the main specification, in order to ensure exogeneity of the index, we recoded the variables on a two point scale with "Present, no indication of severity" and "Present and serious, widespread, or endemic" being equal to 1, and zero otherwise. The diseases include: Leishmaniasis, Trypanosomes, Malaria, Schistosomes, Filariae, Spirochetes and Leprosy. *Source: Low (1988).*

6.1.2 Ethnographic Atlas

Technological Sophistication: Artisanal Crafts The information regarding the presence of the craft was retrieved, indirectly, from variable $V44$, $V45$, $V47$. In order to replicate SCCS variables $v153$, the same index was constructed: the artisanal craft index is equal 1 for societies where no craft is practiced, to 2 for societies where only pottery is practiced, 3 for societies which have weavers (weavers only or weavers and potters) and 4 to societies which have metalworking and eventually one or both other craft specialization (either weaving and/or pottery), 5 to societies with all 3 artisanal crafts. *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998. Available at: http://eclectic.ss.uci.edu/drwhite/ worldcul/ EthnographicAtlasWCRevisedByWorldCultures.sav*

Craft Specialization Constructed recoding Ethnographic Variables $V55$, $V56$, $V58$, respectively for metal, pottery and weaving. The variable of CRAFT specialization was set equal to one for each observation of $V55, V56, V58$ having value 3 or 4, and 0 otherwise. For each craft, only societies where the craft was practiced (independently of having it performed as a specialized profession or not) are considered. The information regarding the presence of the craft was retrieved from variable $V44, V45, V47$, which describe the type of sexual division of labor of the society and are indirectly informative regarding the mere presence of the craft. Observations of variables $V44, V45, V47$ having value of 9 were considered to imply that the craft was not practiced. *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

Castes The variable CASTE takes value 1 if it is reported at least one type of caste stratification, and 0 if not. It is based on information retrieved from Ethnographic Atlas variable $V68$, which classify societies in 4 categories based on the forms of endogamous stratification they present, and namely 1) Caste distinction absent or insignificant; 2) Ethnic stratification; 3) Despised Occupational Groups; 4) Complex caste stratification. The baseline variable CASTE include all three types of caste stratification (2, 3 and 5). The variable CASTE1, include only complex caste stratification, and endogamous occupational groups. *Source: ETHNOGRAPHIC ATLAS. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

Latitude Latitude of society centroid, measured in degrees. Ethnographic Atlas variable $V106$. *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

Mean Temperature Mean global accumulated temperature (T¿0) in an area defined by a 100 km radius drawn from society centroid. Computed through a geographical mapping system. *Source: Global accumulated temperatures (Tmean¿ 0 C), Land and Plant Nutrition Management Service - AGLL - FAO-UN. Available at: http://www.fao.org/geonetwork/srv/en/main.home*

Mean Rainfall Average Annual Precipitation, coded on a scale based on annual mm of rainfall, in an area defined by a 100 km radius drawn from society centroid. Computed through a geographical mapping system. *Source: Climate Research Unit, Univ. of East Anglia. New, M.G., M. Hulme and P.D. Jones, 1999: Representing 20th century space-time climate variability. I: Development of a 1961-1990 mean monthly terrestrial climatology. J. Climate. 12, 829-856. Available at: http://atlas.sage.wisc.edu/*

Agricultural Suitability Average suitability index, ranging from 0 to 1, in an area defined by a 100 km radius drawn from society centroid. Computed through a geographical mapping system. *Source: C. Ramankutty, N., J.A. Foley , J. Norman, and K. McSweeney. The global distribution of cultivable lands:current patterns and sensitivity to possible climate change. Submitted to Global Ecology and Biogeography, March 2001. Available at: http://atlas.sage.wisc.edu/*

Mean Elevation Mean elevation, in meters, in an area defined by a 100 km radius drawn from society centroid. Computed through a geographical mapping system. National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colo. Available at: http://atlas.sage.wisc.edu/

Sea Distance Distance of society centroid to the closest sea coast, measured in thousands of Km. Computed through a geographical mapping system, based on shorelines maps. *Source: Wessel, P. and Smith, W.H.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res., 101(B4): 87418743. http://www.soest.hawaii.edu/wessel/gshhs/*

River Distance Distance of society centroid to the closest river, measured in thousands of Km. Computed through a geographical mapping system, based on shorelines maps. *Source: World Water Bodies Dataset. http://www.arcgis.com/home/item.html?id=e750071279bf450cbd510454a80f2e63*

Island Dummy Dichotomous variables computed measuring the size of the land mass where the society centroid is located. All land masses with an area smaller than Australia were considered islands. *Source: Wessel, P. and Smith, W.H.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. J. Geophys. Res., 101(B4): 87418743. http://www.soest.hawaii.edu/wessel/gshhs/*

Ln(Timing of Transition to Agriculture) Natural logarithm of the timing to the transition to agriculture, from Putterman (2008). Since data are available only at country a level, each society was matched to the modern country based on geographical location. For some societies, timing of Neolithic transition of the modern country is not available, and namely: Bahamas, Eritrea, Fiji, French Polynesia, Kiribati, Marshall Islands, Micronesia, New Caledonia, Palau, Samoa, Solomon Islands and Vanuatu.

Distance to Technological Frontier Aerial distance (Haversine formula), in thousands of km, to the regional technological frontier in 1500 as in Ashraf and Galor (2011a), and namely London and Paris in Europe, Fez and Cairo in Africa, Constantinople and Peking in Asia, and Tenochtitlan and Cuzco in the Americas.

Migratory Distances Migratory distance, on a land path, from Adis Ababa. Computed following Ashraf and Galor (2013). The distance of the society centroid from Adis Abeba is computed using the Haversine formula. In order to replicate the most likely migration pattern followed by early men, we calculated the distance from Adis Ababa of the path that connect several obligatory intermediate points, and namely: Cairo, Istambul, Phnom Phen, Anadyr and Prince Rupert.

Agricultural Intensity Index capturing the fraction of food provided through agriculture, Ethnographic Atlas variable $V28$ *Source: Ethnographic Atls. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

Political Centralization Number of political jurisdictions above the local. Ethnographic Atlas variable $V33$ *Source: Ethnographic Atlas. World Cultures revision by J. Patrick Gray, 1998. Available at: as above.*

Social Classes We recode Ethnographic Atlas variable $V66$ into an index which takes values: 1) for societies with no social classes; 2) societies with a dual stratification into a hereditary aristocracy and commoners, and landlord versus landless societies, and 3) societies presenting wealth distinction based on possession and distribution, and complex stratification correlated with differentiation of occupational statuses.

6.2 Descriptive Statistics

Table 9: Societies with Caste Stratification

	Afr	Med	Asia	Pac	N.Am	S.Am	Total
			Standard Cross Cultural Sample				
Caste	4	13	8	1	1	1	28
No Caste	24	12	24	30	32	31	153
			Ethnographic Atlas				
Caste	57	70	31	2	2	1	163
No Caste	286	47	68	148	270	97	916

Notes: The table reports the number of societies in each geographical area which present caste-like endogamous stratification. Societies are grouped into six geographical region: "Afr." stands for Sub-Saharan Africa, "Med." for Circum-Mediterranean, "Asia" for East-Eurasia, "Pac." for Insular Pacific, "N.Am" for North America and "S.Am" for South America.

Table 10: SCCS Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Technology: Artisan Crafts	3.091	1.413	1	5	186
Technology: All technologies	7.774	2.232	4	12	177
Castes	0.155	0.363	0	1	181
Latitude	14.656	25.162	-55.5	68.7	186
Mean Temperature	12.75	14.544	-28	28	180
Mean Rainfall	6.205	2.408	0	8	171
Mean Elevation	448.615	673.341	0	3822	182
Agricultural Potential	16.726	3.466	4	23	186
Sea Distance	0.367	0.415	0	1.579	
River Distance	0.231	0.564	0	5.068	186
Island	0.22	0.416	0	1	186
Ln(Timing of Neolithic Trans.)	8.237	0.554	5.991	9.259	171
Distance to Techn. Frontier	3.278	1.878	0.018	10.19	186
Migratory Distance	12.335	7.729	0.254	28.065	186
Agricultural Intensity	4.366	1.902	1	6	186
Population Density	3.761	1.977	1	7	184
Political Centralization	2.962	1.183	1	5	186

Table 11: Ethnographic Atlas Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Technology: Artisan Crafts	1.147	1.511	0	4	743
Specialized Potters	0.133	0.34	0	1	548
Specialized Leather Workers	0.152	0.36	0	1	446
Specialized Weavers	0.157	0.364	0	1	382
Specialized Metal Workers	0.961	0.194	0	1	435
Castes	0.151	0.358	0	1	1079
Latitude	14.522	22.16	-55	78	1267
Mean Elevation	704.515	637.652	0	4732.333	1224
Mean Rainfall	2.488	2.168	0	14	1196
Mean Elevation	7223.961	2753.316	122	10910.563	1184
Average Land Suitability	0.359	0.267	0	0.997	1157
Sea Distance	0.202	0.596	0	7.334	1267
River Distance	0.451	0.426	0	1.848	1267
Island	0.173	0.378	0	1	1267
Ln(Timing of Neolithic Trans.)	8.148	0.473	5.991	9.259	1205
Distance to Techn. Frontier	3.3	1.651	0	10.213	1267
Migratory Distance	10.145	7.244	0.11	28.039	1267
Agricultural Intensity	3.036	1.756	0	6	1267
Political Centralization	1.701	1.153	0	5	1267
Social Classes	1.582	0.624	1	3	1085
Disease Index	3.398	2.154	0	7	186

Table 12: SCCS Cross-correlation Table

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Artisan Craft	1.0000														
All Techn.	0.7539	1.0000													
	(0.0000)														
Castes	0.3909	0.4123	1.0000												
	(0.0000)	(0.0000)													
Latitude	0.0194	0.1364	0.0255	1.0000											
	(0.8182)	(0.1043)	(0.7628)												
Mean Temp.	0.1493	0.0321	0.1728	-0.6757	1.0000										
	(0.0751)	(0.7038)	(0.0391)	(0.0000)											
Mean Rain.	-0.0308	-0.2046	-0.1653	-0.3178	0.4036	1.0000									
	(0.7146)	(0.0142)	(0.0485)	(0.0001)	(0.0000)										
Mean Elev.	0.0870	0.1077	0.0184	-0.0871	-0.1004	-0.1052	1.0000								
	(0.3014)	(0.2005)	(0.8269)	(0.3011)	(0.2327)	(0.2111)									
Agri. Pot.	0.0909	0.1090	0.0712	-0.3279	0.4624	0.3585	-0.0283	1.0000							
	(0.2801)	(0.1950)	(0.3980)	(0.0001)	(0.0000)	(0.0000)	(0.7369)								
Sea Distance	0.0588	0.0812	0.1615	-0.0593	-0.0045	-0.1942	0.2159	0.0545	1.0000						
	(0.4852)	(0.3352)	(0.0539)	(0.4820)	(0.9574)	(0.0201)	(0.0096)	(0.5178)							
River Dist.	-0.0093	0.1307	0.2457	-0.1109	0.1356	-0.1932	-0.0808	-0.1410	-0.0237	1.0000					
	(0.9127)	(0.1197)	(0.0031)	(0.1872)	(0.1065)	(0.0208)	(0.3377)	(0.0930)	(0.7789)						
Island	-0.1849	-0.0403	-0.0717	-0.2132	0.1789	0.2734	-0.2237	0.0617	-0.3462	0.3303	1.0000				
	(0.0270)	(0.6328)	(0.3948)	(0.0106)	(0.0326)	(0.0010)	(0.0072)	(0.4639)	(0.0000)	(0.0001)					
Ln(Neolit. Trans.)	0.3735	0.4191	0.1592	0.2271	0.0742	-0.0393	-0.0367	0.0974	-0.0083	-0.1060	0.0483	1.0000			
	(0.0000)	(0.0000)	(0.0575)	(0.0064)	(0.3785)	(0.6412)	(0.6635)	(0.2472)	(0.9218)	(0.2077)	(0.5665)				
Front. Dist.	-0.4039	-0.3968	-0.1457	-0.1958	-0.0615	0.1418	-0.1795	-0.1671	-0.1931	0.2051	0.3794	-0.5271	1.0000		
	(0.0000)	(0.0000)	(0.0824)	(0.0191)	(0.4657)	(0.0912)	(0.0319)	(0.0461)	(0.0209)	(0.0140)	(0.0000)	(0.0000)			
Migr. Dist.	-0.4755	-0.4247	-0.3754	-0.1918	-0.0912	0.1335	-0.0905	-0.0171	-0.1405	-0.0952	0.0638	-0.1733	-0.0637	1.0000	
	(0.0000)	(0.0000)	(0.0000)	(0.0217)	(0.2785)	(0.1120)	(0.2822)	(0.8389)	(0.0942)	(0.2581)	(0.4488)	(0.0384)	(0.4496		
Agri. Int.	0.5127	0.6125	0.1727	-0.2020	0.4511	0.2562	0.0472	0.3588	0.0465	-0.0010	0.1122	0.2720	-0.2591	-0.2340	1.0000
	(0.0000)	(0.0000)	(0.0392)	(0.0156)	(0.0000)	(0.0020)	(0.5757)	(0.0000)	(0.5815)	(0.9905)	(0.1822)	(0.0010)	(0.0018	(0.0049)	
Pop. Dens.	0.5598	0.5828	0.2345	-0.0520	0.3511	0.1902	-0.0476	0.2630	-0.2095	0.0559	0.0960	0.3205	-0.1498	-0.4266	0.5989
	(0.0000)	(0.0000)	(0.0048)	(0.5372)	(0.0000)	(0.0229)	(0.5726)	(0.0015)	(0.0120)	(0.5076)	(0.2541)	(0.0001)	(0.0742)	(0.0000)	(0.0000)
Polit. Centr.	0.6146	0.7043	0.3020	0.0719	0.0439	-0.0588	0.1064	0.2232	0.0827	0.0592	-0.0970	0.2973	-0.1961	-0.4410	0.4597
	(0.0000)	(0.0000)	(0.0002)	(0.3936)	(0.6028)	(0.4857)	(0.2060)	(0.0074)	(0.3261)	(0.4823)	(0.2489)	(0.0003)	(0.0189)	(0.0000)	(0.0000)
Disease Ind.	0.4315	0.2527	0.2440	-0.4436	0.7031	0.3807	0.0841	0.3892	0.1124	0.0058	-0.0680	0.2081	-0.2412	-0.3262	0.5131
1.0000															
	(0.0000)	(0.0023)	(0.0033)	(0.0000)	(0.0000)	(0.0000)	(0.3180)	(0.0000)	(0.1812)	(0.9449)	(0.4195)	(0.0126)	(0.0037)	(0.0001)	(0.0000)

6.3 Preliminary Evidence

6.3.1 Robustness Checks

Table 13: Ordered Logit for Castes and Technological Sophistication

	Technological Sophistication			
	Artisanal Craft			
	(1)	(2)	(3)	(4)
CASTES	0.932**	1.037**	1.136**	1.591**
	(0.457)	(0.509)	(0.537)	(0.650)
	Change in Predicted Probability			
No artisan craft	-0.066	-0.280	-0.246	-0.077
At least Pottery	-0.077	-0.200	-0.192	-0.123
At least Weaving	-0.080	-0.082	-0.088	-0.159
At least Metalwork	0.114	0.341	0.327	0.175
Pottery, Weaving, Metalwork	0.109	0.220	0.199	0.184
Geography Controls		Yes	Yes	Yes
Location Controls			Yes	Yes
Ln(Timing of Neolithic Trans.)				Yes
Distance to Techn. Frontier				Yes
Migratory Distance				Yes
Migratory Distance square				Yes
Continent FE	Yes	Yes	Yes	Yes
Observations	150	150	150	150

Notes: The table reports ordered logit estimates. The upper part of the panel reports the coefficient estimates, the lower part of the panel reports changes in the predicted probability of having a certain degree of technological sophistication having castes, compared not to have them. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table 14: OLS for Castes and Technological Sophistication

	Technological Sophistication					
	Artisanal Craft			All Technologies		
	After 1500	After 1850	After 1900	After 1500	After 1850	After 1900
CASTES	0.920***	1.055***	1.078***	1.154**	1.381**	1.373**
	(0.280)	(0.299)	(0.326)	(0.507)	(0.564)	(0.555)
Geography Controls		Yes	Yes		Yes	Yes
Location Controls			Yes			Yes
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	158	137	94	151	132	89
R-squared	0.448	0.431	0.478	0.459	0.443	0.486

Notes: The table reports OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. In Column (1) and (4) we include the pinpointing date as a control, and we estimate the baseline equation over the whole sample. In Column (2), (5) we restrict the sample to societies observed after 1850 and in Column (3), (6) to societies observed after 1900. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

6.4 Identifying a Causal Relationship

Table 15: Castes and Technological Specialization

	Technological Sophistication					
	Artisanal Craft			All Technologies		
	(1)	(2)	(3)	(4)	(5)	(6)
CASTES	0.849***	0.807***	0.779***	0.927**	0.864**	0.811**
	(0.271)	(0.254)	(0.243)	(0.406)	(0.366)	(0.360)
Agricultural Intensity	0.308***	0.210***	0.173***	0.649***	0.519***	0.461***
	(0.050)	(0.060)	(0.062)	(0.050)	(0.064)	(0.065)
Population Density		0.221***	0.132**		0.311***	0.191***
		(0.063)	(0.060)		(0.083)	(0.071)
Political Centralization			0.270***			0.432***
			(0.082)			(0.104)
Ln(Timing of Neolithic Trans.)	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Techn. Frontier	Yes	Yes	Yes	Yes	Yes	Yes
Migratory Distance	Yes	Yes	Yes	Yes	Yes	Yes
Migratory Distance square	Yes	Yes	Yes	Yes	Yes	Yes
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes
Location Controls	Yes	Yes	Yes	Yes	Yes	Yes
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	149	149	149	143	143	143
R-squared	0.613	0.653	0.675	0.748	0.780	0.803

Notes: The table reports OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Agricultural Intensity is an index which capture what fraction of food is provided to agriculture. Population density measures the approximate number of people per square male, coded on a 7 point scale. Political Centralization captures the number of political jurisdictions above the local. The variable Ln(Timing of Neolithic Transition) is the natural log of the number of years elapsed since the area where the society is located went through the transition to agriculture. Distance to technological frontier is the distance, in thousands of km, from the closest regional technological frontier. Migratory Distance and Migratory Distance squared are, respectively, the migratory distance (in thousand of km on a land path) and square migratory distance from Adis Abeba. Geography control variables include: latitude, agricultural potential of the land occupied by the society, average temperature of the coldest month, average rainfall and mean elevation. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table 16: Assessing Bias From Unobservables

	Technological Sophistication					
	Artisanal Craft			All Technologies		
	(1)	(2)	(3)	(4)	(5)	(6)
Size of the Bias		5.6	28.8		2.1	3.4
CASTES	0.806***	0.683***	0.779***	1.050**	0.715**	0.811**
	(0.249)	(0.230)	(0.243)	(0.483)	(0.341)	(0.360)
Agricultural Intensity		Yes	Yes		Yes	Yes
Population Density		Yes	Yes		Yes	Yes
Political Centralization		Yes	Yes		Yes	Yes
Ln(Timing of Neolithic Trans.)			Yes			Yes
Distance to Techn. Frontier			Yes			Yes
Migratory Distance			Yes			Yes
Migratory Distance square			Yes			Yes
Geography Controls	Only Latitude	All	All	Only Latitude	All	All
Location Controls		All	All		All	All
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	149	149	149	143	143	143
R-squared	0.613	0.653	0.675	0.748	0.780	0.803

Notes: The upper panel of the table report the computed size of the bias from unobservables inferred from selection on observables. The lower panel of the table reports OLS estimates of the relative regressions. Regression in Column (1) and (3) are our "restricted regressions" and only include latitude and region fixed effects. Regressions in Column (2) and (4) are the first set of "unrestricted regressions" (R_u) and include country fixed effects, all geographical controls, all location controls, population density, agricultural intensity and political centralization. Regressions in Column (3) and (6) are the second set of "restricted regressions" (R_r) and include all controls from the first restricted regression, plus the Ln(Timing of Neolithic Transition), the distance to technological frontier, Migratory Distance and Migratory Distance squared. The bias from unobservables is calculated as $R_r/(R_u - R_r)$.

Table 17: Castes and Technological Specialization

	Technological Sophistication	Craft Specialization			
		Pottery	Leather	Weaving	Metal
	(1)	(2)	(3)	(4)	(5)
CASTES	0.380**	0.162***	0.354***	0.111*	0.019
	(0.156)	(0.050)	(0.102)	(0.056)	(0.016)
Agricultural Intensity	Yes	Yes	Yes	Yes	Yes
Political Centralization	Yes	Yes	Yes	Yes	Yes
Ln(Timing of Neolithic Trans.)	Yes	Yes	Yes	Yes	Yes
Distance to Techn. Frontier	Yes	Yes	Yes	Yes	Yes
Migratory Distance	Yes	Yes	Yes	Yes	Yes
Migratory Distance square	Yes	Yes	Yes	Yes	Yes
Geography Controls	Yes	Yes	Yes	Yes	Yes
Location Controls	Yes	Yes	Yes	Yes	Yes
Continent FE	Yes	Yes	Yes	Yes	Yes
Observations	632	480	407	319	366
R-squared	0.652	0.351	0.551	0.378	0.362

Notes: The table reports OLS estimates. The unit of observation is the Ethnographic Atlas society. The dependent variable in column (1) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (2), (3), (4) and (5), the dependent variable is a dichotomous variable indicating whether the existing craft is performed by specialized artisans or not. The explanatory variable of interest is the presence of castes, $CASTES$, which takes value 1 if at least one endogamous group is present and 0 otherwise. Geography control variables include: latitude, average agricultural suitability, average annual cumulated temperature above 0 degree, average precipitation and mean elevation, all measured within a 100 km radius of society centroid. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for society located on islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors, clustered at the linguistic family level (the total number of linguistic family cluster is 68), are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.

Table 18: OLS for Castes and Technological Sophistication

	Technological Sophistication					
	Artisanal Craft			All Technologies		
	(1)	(2)	(3)	(4)	(5)	(6)
Disease Index	0.229***	0.227***	0.148**	0.366***	0.363***	0.248***
	(0.076)	(0.077)	(0.064)	(0.111)	(0.112)	(0.091)
Geography Controls	Yes	Yes	Yes	Yes	Yes	Yes
Location Controls		Yes	Yes		Yes	Yes
Population Density			Yes			Yes
Continent FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	159	159	159	143	143	143
R-squared	0.433	0.446	0.574	0.413	0.414	0.622

Notes: The table reports reduced form OLS estimates. The unit of observation is the Standard Cross Cultural Sample society. The dependent variable in column (1), (2) and (3) is an index of artisanal sophistication which captures the type of specialized artisanal crafts performed in the society and ranges from 1 to 5. In column (4), (5) and (6), the dependent variable is an index of technological sophistication indicating the degree of sophistication of the society along multiple dimensions, i.e. crafts, agriculture, writing and land transportation. The explanatory variable of interest is an index of geographical disease exposure ranging from 0 to 7. Geography controls include: latitude, average temperature of the coldest month and average rainfall. Location controls are: distance of the society centroid from the closest river (in thousand Km), distance of the society centroid from the closest sea coast (in thousand Km) and a dummy variable for islands. Continent dummies for the six geographical areas in the sample, i.e. Sub-Saharan Africa, Circum-Mediterranean, East-Eurasia, Insular Pacific, North America and South America, are included in all specifications. Robust standard errors are reported in parentheses. *** stands for significant at the 1 percent level, ** at the 5 percent level, * at the 10 percent level.