The Evolution of Productive Organizations

Francisco Brahm*, Joaquín Poblete†‡

January 29, 2019

Abstract

We develop a cultural evolution model that illuminates the evolution of productive organizations, such as partnerships, guilds and modern corporations. The basic insight provided by the model is that productive organizations evolved because they favored the difficult-to-propel process of cumulative culture by virtue of being exclusive and facilitating social learning. Productive organizations make social learning and culture useful to society, playing a fundamental role in the adaptive success of the human species. The model also illuminates issues regarding adaptation and rigidity, the locus of innovation, secrecy and the origins of specialization. We test the model using a sample of premodern societies drawn from the Ethnographic Atlas. The empirical analysis provides supporting evidence for our predictions.

KEYWORDS: Organizations, Theory of the Firm, Knowledge, Social Learning, Cultural Evolution, Specialization.

---

*London Business School. E-mail: fbrahm@london.edu

†Pontificia Universidad Catolica de Chile, E-mail: joaco.poblete@gmail.com. Poblete acknowledges research support from the Instituto Milenio SCI Project No. P05-004F, Institute for Research in Market Imperfections and Public Policy, ICM IS130002.

‡We would like to acknowledge comments provided by Christoph Loch, Stelios Kavadias, Brian Silverman, Joel Peck, Antoinne Feylessoufi, Raicho Bojilov, Emilio Depetris-Chauvin, Marcus Feldman, Ricardo Alonso, Jose Tessada and Tim Waring; by seminar participants at the Judge Business School, Stanford Business School, the London Business School; and by participants at the 2nd Annual Meeting of the Cultural Evolution Society, Clare Hall Graduate Students Conference, Third Culture Conference at the University of Birmingham, the INFORMS Annual Meeting, the Winter CJBS PhD Conference, the Sixth Development Workshop at the Catholic University of Chile, the inaugural Strategy Science Conference, Barcelona GSE Summer Forum (Organizational Economics) and the Trento Summer School on "New thinking on the firm: Perspectives from law, economics, organization, and history".

1 Introduction

Organizations, defined as a stable and interacting collection of individuals with a common and specific goal (Scott, 2003), have played a crucial role throughout human history. Clubs, armies, universities and churches are some examples. One of such organizations is the "productive organization" (PO), whose goal is to produce goods and deliver services that satisfy the material needs of human populations (e.g., food, shelter, clothes, tools, transportation, healthcare, defense). Whether it is the societas in Roman times (Hansmann et al., 2006), the guilds in medieval times (Ogilvie, 2014), the partnerships in early Renaissance (Padgett et al., 2006), or some sodalities in semi-literate societies (Smith et al., 2016), goods and services of POs have consistently sustained villages towns, cities and states. In the last century and a half, the influence of modern firms such as large corporations cannot be overstated: their scale and reach dominate modern economic life (Chandler, 1990).

Organizations are varied, so let us contextualize POs at the outset. There is a vast literature in sociology describing and cataloguing organizations (Blau and Scott, 1963; Scott, 2003). First, organizations exclude social aggregates – collection of people in the same place – (e.g., public in stadium) and networks – fluid collection of connected individuals sharing an interest (e.g., collectors). Organizations are groups – a stable and interacting collection of individuals. But not all groups are organizations. "Primary groups" such as family, household, and friends do not have a specific goal; instead they serve an intimate and emotional supporting role. Organizations, in contrast, are a "secondary group" that exhibit a common and specific goal, often practical or utilitarian. Most organizations also display a high degree of formalization (e.g., roles, rules governing behavior, clear boundaries). POs are located here: they are a large subset of these "formal organizations"[1]. POs differentiate from another important type of formal organization, the "voluntary organizations", where members can freely join or leave (e.g., some charitable organizations, clubs, churches); POs, as we define them, have instead the *crucial distinction of restricted access*. Although the most common type of POs are for-profit companies producing goods and services, our definition does not preclude other POs providing more specific services such as defense (e.g., police, army), healthcare and education (e.g., schools, universities). However, we restrict our attention to companies, specially its pre-modern ancestors such as guilds, to focus the discussion. This allows to contrast with theories in economics, and it avoids dealing with the nuances that arise when discussing some specific services such as defense or

[1] In contrast, "informal organization" refers to the tacit consensus and norms that guide the goal-directed behavior of the group. All "formal organizations" contain informal organization within them; in contrast, informal organizations can exist with minimal, or even none, formalization (however, this is rarer).

education.

Extensive research in economics and business provides explanations for the existence of POs, focusing on incentives and governance. Under the umbrella of the "theory of the firm", several theories propose, in a nutshell, that firms are a way to avoid the potential hazards involved in the market exchange of goods, labor and assets (Coase, 1937). Under conditions of uncertainty and specificity, transactions among self-motivated agents become costly, favoring the use of a hierarchical organization to govern them. Different theories emphasize different costs of market exchange and thus different rationales for firms. For example, firms allow for ex ante investment incentives when parties could behave opportunistically (Hart and Moore, 1990); firms allow for the use of authority and fiat when transaction complexity requires constant coordination (Williamson, 1991); and firms solve moral hazard problems that stem from diversity of tasks in a transaction (Holmstrom and Milgrom, 1991), among others. Empirical evidence supports these these governance functions of firms (Lafontaine and Slade, 2007).

Recent research indicates, however, that the nature of firms is not only about protection from hazards in exchange or investment but also about being carriers and transmitters of culture, knowledge and intangible capital. Plenty of evidence is consistent with this view. Recent evidence from a comprehensive sample of US firms shows that vertically integrated companies exhibit a surprisingly small flow of physical goods and a significant flow of knowledge and intangible capital (Atalay et al., 2014); the literature on organizational learning shows that knowledge diffusion is enhanced within firms (Argote and Miron-Spektor, 2011); the existence of persistent performance differentials across firms has been related to firm specific know-how based on relationships and culture (Gibbons and Henderson, 2012); the central function of guilds in medieval Europe was the efficient transmission of skills and tacit knowledge (De La Croix et al., 2017); and modern partnerships (e.g., a law firm) essentially provide know-how to their members through effective training and mentoring (Morrison and Wilhelm, 2008). Common to these examples is the idea of knowledge and information transmitted among individuals via social learning (e.g., copying, teaching, and apprenticeships). Thus, a first challenge is to incorporate social learning – the basic building block of culture – into the theory of productive organizations.

A second challenge to mainstream theories of POs is that their evolutionary origin is not addressed and that, as a consequence, we lack an understanding of their historical role in the development of our civilization. Current theories of POs focus on modern firms such as corporations, without addressing their relationship with their historical "ancestors" (e.g., partnerships, guilds, Roman societas). In current theories, POs are a *de novo* phenomenon, without explaining how they went from nonexistence to current universality. In other words, we lack an understanding of the evolutionary and historically grounded mechanisms that made the POs to be selected, to gradually increase in frequency and to come to dominate the modern economic landscape[2]. Clarifying the historical role and evolutionary roots of POs is necessary to fully understand their nature and their contribution to our species's economic success.

In this article, we develop and test a theory of the origin of POs that addresses these challenges. First, our theory puts social learning at the center of the role played by POs. Using a cultural evolution model (Boyd and Richerson, 1995), we show that POs can improve the conditions that are necessary for knowledge, technologies and other cultural traits to accumulate over time. In particular, POs need to restrict access and facilitate social learning in order to favor cultural ratcheting. Given that cultural ratcheting is a key driver of cultural complexity and the adaptive success of our species, societies that adopt POs would be favored by cultural selection. This theory produces an account of POs based on knowledge and culture that does not require incentive and governance considerations. Further, by focusing on knowledge and cultural accumulation, we go beyond extant theories that consider knowledge but that focus on the optimal organization of knowledge activities (e.g., hierarchical firms as a natural response to problems of varying complexity and agents of varying capacity; see Garicano, 2000).

Second, our theory provides a logic for the rise and invasion of POs. In our evolutionary past, the first exclusive and social-learning-enhancing POs favored the conditions for cumulative culture, and as a consequence, the mechanisms of imitation and group selection would have operated in selecting these early POs, which then gradually invaded the landscape. It is likely that some neolithic sodalities were these first POs (at the outset of sedentary human settlements): they were exclusive and focused on transmission of knowledge, and they specialized in a craft or interest (Lowie, 1948; Anderson, 1971). This account provides an additional layer to our understanding of POs. In current incentive-based theories, the set of exchanges or investments requiring governance originate from an already-in-place and exogenous pool of knowledge and technologies. Instead, in our theory, POs arose and expanded because they favored the expansion of this pool.

The theory of the origin of POs applies straightforwardly to premodern POs, such as guilds, and long-standing POs, such as partnerships. Consistently, we test our theory in the context of premodern societies using the Ethnographic Atlas. To the extent that modern firms (e.g., corporations and limited liability companies) descend from these earlier organizations, our theory also provides insights into the nature of modern firms.

[2] As the following quote attests, this is also true for evolutionary theories of POs: “One kind of glaring omission in our 1982 book was the failure to think about evolution, and industry evolution in particular, in a historical context. This is a real head-thumper kind of realization after the fact." (stated by Sidney Winter in Murmann et al., 2003, p. 28)

1.1 Our argument and findings

Our model is based on standard theories for cultural evolution (Boyd and Richerson, 1985; Boyd and Richerson, 2005; Henrich, 2015). These theories, developed in Evolutionary Anthropology, study culture as an evolutionary system (which can also feed back into genetic evolution, hence its tag of "dual-inheritance theory"). The theory specifies microfounded mechanisms for the inheritance, selection and variation of cultural traits. Culture is defined as information that is acquired from other individuals via social learning mechanisms, such as imitation and teaching from parents (vertical transmission) or nonrelated peers (horizontal and oblique transmission). Information includes beliefs, norms, knowledge, skills, artifacts and technology. A central insight of this theory is that social learning – a fundamental behavioral trait of the theory – is able to generate cumulative culture. By means of diffusing knowledge and innovations in a society, social learning produces their gradual accumulation over time. This process has allowed the human species to adapt and conquer, in a very short span of evolutionary time, every environment in the globe. Often, this adaptive cultural ratcheting process occurs independently of people's awareness or immediate understanding (Henrich, 2015).

However, social learning is not automatically conducive to adaptive cumulative culture (i.e., a culture that increases fitness). An important class of cultural evolution models show that, although social learning is favored by selection, it does not directly lead to an increase in fitness (Rogers, 1988; Boyd and Richerson, 1995). The problem is that, under changing environmental conditions (which call for novel technologies or cultural traits after a change), if part of the society invest in understanding the current state of nature and developing the required technology (i.e., individual learning) and others simply copy someone else in the society at a lower cost (i.e., social learners), then social learners enjoy higher fitness and, over time, expand until their fitness equalizes to that off individual learners. The outcome is a society with culture (i.e., social learning diffuses the technology and the cultural traits) but whose fitness hasn't increased (as compared to the starting condition of only individual learners). Simply put, social learning generates culture, but it is non-adaptive. This result, known as Roger's paradox, has generated an important literature exploring the conditions that makes social learning a source of adaptive cumulative culture (Laland, 2017; Boyd and Richerson, 1995; Boyd et al., 2013). Some of these conditions are payoff biased social learning, selective learning (e.g., using social or individual learning depending on the circumstances), and traits/technologies that are not independent across environmental states (see Boyd et al., 2013 for more details).

In this paper, we show that POs allows societies to overcome Roger's paradox and thus facilitate adaptive cultural accumulation. The exclusiveness of POs allows society to benefit from the improvements in social learning that happen within the POs. We contend that this benefit is the key to understanding the origins and subsequent increase in frequency of POs. Once a “right" PO arises (exclusive and better at social learning), then cultural group selection will exploit its capacity to generate adaptive culture, leading to an increase in its frequency over time. There are several group selection processes at play: first, POs will be imitated by other groups or societies; second, the group or society with POs will atract migration from worse-off better societies; third, the group or society with POs produce more offsprings and therefore will expand geographically more rapidly; fourth, POs help in generating technological and numerical advantage in armed conflicts with other societies. Note that these selection processes have been shown to select and diffuse other group-level or society-level traits, such as routines, religion, or political insitutions (Henrich, 2015; Turchin, 2016).

To be clear, we do not model the long-run cultural selection processes. We do not model different groups or societies with different institutions, for example with POs of different types, competing against one another over time. In this paper we determine the conditions of POs that make social learning useful as a source of adaptive cultural accumulation. Then we rely on the four mechanisms of the cultural selection processes described above to do the “long-run lifting". This type of explanation is standard in evolution. In biology, an evolutionary explanation consists of clarifying how the trait favors reproduction and survival (e.g., colour of flowers attract bees which then help polonization), and then simply pointing that natural selection will gradually select that trait in the organism's population (e.g., flowers with no colours will be selected out)[3]. Similarly, given that it is well documented that adaptive cumulative culture and group selection are the crucial processes that generated the success and expansion of the human species (Boyd, 2017; Henrich, 2015; Turchin, 2016; Laland, 2017), it only suffices to demonstrate that the trait being explored favours the conditions for these processes to operate (or alternatively, that the trait will be harnessed by these processes) in order to claim that the trait is going to invade the social landscape. In our case, we show that POs directly support adaptive cumulative culture, which in turn, supports group selection.

The modelling approach is as follow. We use a workhorse cultural evolution model that is widely used to explore the conditions that make social learning adaptive (Rogers, 1988; Boyd and Richerson, 1995). In every period, a population of agents adopts an activity (or technology) that confers benefits but whose value is lost if the environ-

[3]Paleontology and phylogenetic trees complement this by building the “tree of life", that is, the documentation of the branching of the different species over time, including their timing, sequence and differential traits. In cultural evolution, the equivalent is: i) the tracing the evolution of artifacts (or institutions such as kinship structures) across time and place by archeologists, anthropologists and historians, and ii) the study of the branching of the different languages in the human species by linguists.

ment changes. These agents can have one of two strategies: individual learning, where agents always adapt their activity to the environment of the current period with a cost C, or social learning, where the agent bears cost c (less than C) in order to copy a random agent from the previous period. Consistent with prior literature, we show that, although favored by selection, social learning does not enhance the fitness of the population because social learning expands until its benefit is the same as that of individual learning. As social learners expand, they place a negative externality on the rest of social learners (i.e., social learners become increasingly liable to changes in the environment). However, we prove that by placing a portion of the population into a small group that has a lower cost of social learning equal to $\tilde{c}$ and whose access is restricted – that is, a PO is introduced – the fitness of this group, and therefore of the average fitness of the population, increases (particularly if the likelihood of environmental changes and $\tilde{c}$ decreases). As a corollary, POs end up populated exclusively by social learners – an interior equilibrium is reached outside of the firm – and thus, POs adapt slowly to environmental changes; consequently, invention in society is performed outside the POs. The former is consistent with the frequently documented rigidity and inertia of POs (e.g., Kodak); the latter is consistent with major inventions in history coming from scientists, lone entrepreneurs and inventors, and with recent evidence coming from large scale data about US manufacturing plants (Arora, Cohen and Cunningham, 2018). We explore several comparative statics in order to explore conditions that make POs more or less beneficial. Among these, we find that "secrecy" – the fact the social learners located outside a PO learn imperfectly from the agents within the PO – decreases the fitness of POs.

Intuitively, POs make social learning adaptive by using two mechanisms. First, POs improve the efficiency and the fidelity of knowledge transmission ($\tilde{c} < c$) (Lewis and Laland, 2012). Knowledge transmission is a cooperative act (Laland, 2017; Fogarty et al., 2011), and POs naturally generate conditions that favor cooperation among their members. This comes from many sources, such as population structure and assortment of cooperative types (Nowak, 2016; Rand and Nowak, 2013) or triggering deep-seated tribal instincts (Bowles, 2009). The second mechanism is that POs, by means of being exclusive, put a halt to the expansion of social learners, limiting the negative externality they generate. This restriction of access enables any improvements in the cost of social learning within POs to be adaptive and useful for society. Thus, instead of preventing hazards when agents interact (as in current theories of POs), POs mitigate the negative externality that the expansion of social learners generates.

An important contribution of our theory is that it provides a novel solution to Roger's paradox. Current models suggest that in order to increase fitness, social learning has to allow individual learning to perform better (Boyd and Richerson, 1995). We are the first to show that social learning can be beneficial to society without resorting to a positive impact on individual learning. Within exclusive POs, improvements in social learning are sufficient. Equally importantly, our theory also provides an explanation for the origin of specialization within POs, a condition that has been prevalent in these organizations. Contrary to extant explanations based on comparative advantages and economies of scale, we show that even in societies in which trade is absent, specialization within POs – that is, social learners of a PO share the same activity – is beneficial to the society. In a world with multiple productive activities, specialization within POs maximizes the social learning advantage of POs which in turn maximizes the fitness benefits that POs bring to society.

We test our theory using data from the Ethnographic Atlas and the Standard Cross Cultural Sample, as provided by the D-PLACE dataset (Kirby et al., 2016). We measure the presence of technologies in premodern societies (e.g., weaving, metal working, and pottery) and whether they were executed throughout the society or mainly by a small and exclusive group of people, that is, within a PO. We provide a lenghty discussion regarding the mapping between the PO in our model and our measurement. We use several measures of population as our dependent variable, which adequately capture the notion of fitness in evolutionary models and of progress in premodern societies (where progress would translate into population increase rather than increases in per-capita wellbeing). After controlling for a large set of covariates –geography (e.g., distance to sea, latitude), resource endowment (e.g., mammal richness), year of ethnography, agricultural intensity dummies (e.g., intensive), type of settlement dummies (e.g., nomadic), and region dummies (36 in total)–, we find supportive evidence for the main proposition of our model: the presence of technologies increases the fitness of the premodern societies only when they were executed using POs. We also explore the comparative statics of our model. We proxy environment changes using climate uncertainty at the society's location. Change in the cost of social learning is proxied using several measures (e.g., presence of apprenticeship in the society). Although the proxies are imperfect, we find supportive evidence for all the six measures we use. The results are robust to the use of cultural complexity as the dependent variable and to controlling for alternative explanations (i.e., presence of trade and political complexity). We address endogeneity concerns using two strategies: first, we show that selection on unobservables would need to be very large to overthrow our results; second, we use an instrumental variable estimation to correct for a potential problem reverse causality. Following previous research (Depetris-Chauvin and Ozak, 2017), we instrument the presence of POs using the migration distance from Africa. We also instrument for the presence of technology using the idea of kinship tightness (Enke, 2017). The IV estimations do not change our results (if anything, they become slightly stronger). Following Giuliano and Nunn (2018), we also execute an out-of-sample test and confirm a secondary prediction of our model: we find that modern societies with ancestors reliant on POs have a more intense use of tradition today (i.e., they use more social learning).

As a whole, we see our empirical exercise as a first attempt to bring the theory to data. It grounds the theoretical arguments and provides compelling correlations in line with the predictions. The accumulation of consistent evidence in several tests also serves to compensate for the potential identification weaknesses in any one of them.

1.2 Other related literature

Our paper is related to other strands of the literature regarding the nature of firms. Knowledge based accounts of POs have been put forward in the economics and management literature. These theories point at the importance of knowledge as a basis for POs, focusing mostly on the integration of different bits of specialized knowledge in order to apply it to current productive challenges (Grant, 1996) or in order to solve novel problems (Nickerson and Zenger, 2004). Also, they explore the benefit that hierarchies provide in matching problems of varied difficulty to agents with different levels of skills (Garicano, 2000), the substitution of knowledge transfer by authority and directives (Demsetz, 1988) or how the organization of a firm aims to reduce asymmetric information between the firm and its clients (Levin and Tadelis, 2015, Poblete, 2015) (for a discussion, see Spulber, 2009). We focus on a more basic and general issue – the transmission of knowledge among individuals. In this sense, we are closer to work on organizational learning (Argote and Miron-Spektor, 2011), knowledge replication (Winter and Szulanski, 2001), idiosyncratic culture as the root of performance in POs (Gibbons and Henderson, 2012), and knowledge transmission costs as the source of boundaries in knowledge-intensive firms (Espinosa, 2017). Our theory is also explicitly dynamic, formally modeling the transmission of knowledge and its evolutionary process.

An evolutionary take on POs is also part of extant research in economics, sociology and management (Nelson and Winter, 1982; Aldrich (1999); Hannan and Freeman, 1997; Levinthal, 1997). Although these theories successfully inform questions of change and adaptation, they do not address the question of what are the evolutionary and historical origins of POs. Why it is that POs have gone from nonexistence to domination of the modern economic landscape is not explained. By studying how POs impact the process of cultural evolution, we provide a first step in this direction.

Our paper also relates to the literature in economics that explores the long- term and historical determinants of economic development (Nunn, 2009; Spolaore and Wacziarg, 2013). Within this body of work, we relate more directly to research that studies the origin and evolution of institutions (North, 1991), particularly the branches that highlight the role played by culture (Alesina and Giuliano, 2015; Nunn, 2012; Mokyr, 2016; Tabellini, 2008) and that use a cultural transmission (Bisin and Verdier, 2011) or a cultural evolution approach (Giuliano and Nunn, 2018; Giuliano, 2016; Enke, 2017).

Our contribution lies in providing a model for the evolution of organizations, complementing the prevailing focus on institutions (e.g., state centralization). Our theoretical arguments regarding the origins of POs also complements the economic history literature that is empirically tracing the origin of guilds, corporations, and private limited liabilities companies (Dari et al., 2017; De La Croix et al., 2017; Lamoreaux, 1998; Guinnane et al., 2007; Hansmann et al., 2006). We believe that our theory (and its further developments) can inform the historical data.

2 Model and Predictions

We build on a workhorse model in the cultural evolution literature (Rogers, 1988; Boyd and Richerson, 1995). Mathematical proofs of all the results are presented in the appendix.

In every period, a continuum of long-lived agents adopt a technology that confers fitness, but whose value is subject to changes in the environment. There are N environmental states. In each period, the state may change with probability p. For every state, there is a unique technology that provides fitness. By a normalization, we can assume without loss of generality that the fitness of a technology is 0 unless it is tuned to match the state. Agents adopt a technology by using one of two behavioral strategies. An *Individual Learner* studies and understands her environment and is able to develop in each period a new technology tuned to the current state. This strategy has a cost C, which is bounded between 0 and 1. The second alternative is to learn socially. A *Social Learner* looks at what some other randomly chosen member of the population did in the previous period and simply copies its technology, incurring a cost $c < C$. This strategy is less costly because the agent does not need to understand the underlying state but rather simply copies what others do. In order for social learning to survive, we assume $(C - c) > p$.[4]

Let r_I be the share of individual learners and r_s the share of social learners in the population, and let q be the percentage of people with a tuned technology in the population. For any given pair of shares r_I and r_s, the expected ratio of tuned agents is given by $q^e(r_I, r_e) = \frac{r_I}{1-(1-p)r_S}$.[5] The fitness of an individual learner is $f_I = 1 - C > 0$ because she is always tuned to the state bearing the cost C. The fitness of a social learner is $f_S = (1 - p) \cdot q^e - c$. Social learners, given that they copy their behavior from others, sacrifice tuning if the state of the world changes or if inadvertently they

[4] Given that we model traits that can be transmitted among individuals, our theory is better suited for POs that display this behavior, such as guilds and partnerships. Modern corporations add to the mix the integration of different knowledge sets, which we do not formally model.

[5] To obtain this result, notice that the percentage q is governed by the difference equation $q(t) = r_I + q(t-1) \cdot (1-p) \cdot r_S$. The expected value is calculated by computing the steady state.

copy from an untuned member. The fitness of social learners is increasing in the share of individual learners because individual learners increases the chances of copying from tuned individuals.

We assume that in every period, a small proportion of agents adopt the strategy of other agents with higher fitness levels. This type of evolutionary dynamics is known as a quasi birth and death process and converges to evolutionarily stable strategies (ESSs).[6] Formally, a population with shares (r_S, r_I) plays an ESS if a small group of invaders using any alternative strategy achieves a strictly lower average fitness. Consistent with prior literature, we find that there exists a unique equilibrium in which both strategies are present with shares that depend on C, c, and p,[7] and the fitness of both types is $1-C$. (See the appendix for a formal proof.) Intuitively, in equilibrium, there cannot be only individual learners because in that case, the ratio of tuned population q^e approaches one and the fitness of social learners would be higher than that of individual learners. In the same way, there cannot exist only social learners because then, the ratio of tuned population q^e approaches zero, and the fitness of individual learners becomes greater than that of social learners.

Observe that because both behavioral types in equilibrium achieve the same fitness $1-C$, this implies that society as a whole does not benefit from the existence of social learning because the same average level of fitness can be achieved with individual learning only. The fact that social learning is selected but does not affect the fitness of the population is known as Roger's paradox (Rogers, 1988). This result has demonstrated to be robust to different specifications and assumptions, leading Boyd and Richerson to state that "*to improve the average fitness of the population, imitation must make individual learning cheaper or more accurate*" (Boyd and Richerson, 2005; p. 39). In what follows, we show that adding a PO in the society solves this paradox in a way that, we argue, is fundamentally different from other solutions proposed in the literature because it does not require improving the fitness of individual learning.

2.1 Productive organizations

We now introduce productive organizations into the model. Two characteristics describe a productive organization (PO). First, access to the PO is limited. A fraction of agents λ is located inside the organization, and this fraction is fixed. This means that even though additional members might want to be a part of the PO, membership is limited by the value of λ. This does not mean that λ cannot shift or evolve; we address this

[6] We could also assume that agents are short-lived and their reproduction rate depends on their fitness level. In either case, the equilibrium concept is ESS, and our results are the same.

[7] The share of individual learners is given by $r_I = \frac{p \times [(1-(C-c)]}{(1-p)\times(C-c)}$. Thus, individual learning increases with uncertainty and decreases with the cost advantage of social learning.

issue in detail in section 3.3. The definition of λ speaks to fundamental characteristic of POs, namely they have a boundary. Scott (2003) indicates that “organizations, as a condition of their existence, must maintain boundaries that separate them from their environments. In the absence of distinguishable boundaries, there can be no organization as we understand them" (p. 185).

Exclusiveness is a persistent characteristic that POs have had across history: Neolithic sodalities, Roman societas, medieval guilds, Renaissance partnerships and modern corporations all limit the possibility of becoming a member. Exclusiveness appears to be present from the first records of POs. For example, Apel (2008) describes the production daggers in Scandinavian society in the late neolithic and explains that “the production is consciously organized to keep the recipes of the technology exclusive to certain segments of the society". More general reviews confirm that exclusivity is a salient characteristic of the first non-kin-related goal-oriented organizations (sodalities). For example, in the revision by Anderson (1971) of Lowie (1948), he states that in his description of early sodalities, “he could find no common characteristics beyond the fact that they all excluded non-members."

Second, agents that belong to the same PO can learn from each other at a lower cost. When a social learner adopts her technology from another member of the PO, she has a cost $\tilde{c} < c$ (if she copies from outside, she bears a cost c.) This entails that for a specific piece of information to be transmitted, lower effort would be required; or alternatively, for a given amount of effort, the fidelity of the information transmission is higher. Theoretically, this assumption regarding POs is sustained because population structure favors cooperative behavior through assortment of cooperatives types or higher frequency of interactions (Nowak, 2016; Rand and Nowak, 2013). This, in turn, favors the emergence of teaching or mentoring, which in essence is a cooperative act (Laland, 2017; Fogarty et al., 2011; Dean et al., 2012). For example, when social learners imitate technologies inside the organization, they can be favored by the active transmission from the subject they are attending to and who might otherwise be passive. More generally, this assumption can also be sustained by pointing at the deep-seated tribal tendencies of humans beings, which make them prone to identify with their groups and to trust and help fellow members (Bowles, 2009).

Empirical evidence supports less expensive social learning within POs. There is considerable evidence in the management and economics literature showing that learning from others is more efficacious when learning from other coworkers of the organization, as opposed to from the outside (Argote and Miron-Spektor, 2011). Research regarding guilds and partnerships documents their role in improving knowledge transmission between its members (De La Croix et al., 2017; Morrison and Wilhelm, 2008). In archeology, Coto-Sarmiento et al. (2018) provides compelling evidence from three centuries of amphorae production in workshops in the Roman empire. Their analysis suggests that the variability of amphorae between workshops is mostly consistent with a process of high-fidelity social learning within workshops (i.e., master to disciples) instead of horizontal transmission or mobility between workshops.

As before, let r_I be the share of individual learners outside the firm, and $\tilde{r}_I$ be the share of individual learners inside the PO. In the same fashion, define r_s, $\widetilde{r}_s$, q, $\widetilde{q}$, q^e and $\widetilde{q}^e$. The fitness of an individual learner is the same outside or inside the PO, $\widetilde{f}_I = f_I = 1 - C > 0$. The fitness of a social learner outside the PO is

$$f_S = (1-p)[(1-\lambda) \cdot q^e + \lambda \widetilde{q}^e] - c \tag{1}$$

whereas a social learner inside the PO enjoys a fitness of

$$\widetilde{f}_S = (1-p)[(1-\lambda) \cdot q^e + \lambda \widetilde{q}^e] - [(1-\lambda)c + \lambda \widetilde{c}] = f_s + \lambda(c - \widetilde{c}) \tag{2}$$

In equilibrium (or equivalently, in the long run due to evolution), the expected fitnesses of both behavioral strategies outside the firm are equal. Provided there exist at least one social learner outside the PO (something that happens if the PO is not too big), there exists a unique equilibrium in which the average fitness of society is greater than that of individual learners. The result is stated formally in the next proposition.

Proposition 1 *If λ is sufficiently small, the existence of the PO increases the average fitness of the population.*

To understand the intuition behind this proposition, it is useful to compare the model with POs to the basic model of the previous section. In the basic model, social learners reproduce and grow, lowering the average fitness of the population q until the fitness decreases to the level of individual learners $1 - C$. With the introduction of a PO of limited size, this negative externality is put to a halt before all the benefits are diluted away, thus allowing society to benefit from social learning (figure 1A). A corollary of this result is that, inside the PO, there are only social learners, as any equilibrium with individual learners inside the PO would be invaded by social learners (figure 1D).

The PO has several other interesting effects. First, notice that as the PO is populated only by social learners, they bear a relatively larger risk of environmental change, and thus, the share of fitted population inside the firm $\tilde{q}$ is less than the share of the fitted population outside the firm q. This makes POs slower to adapt, and the mere existence of POs decreases the average level of tuning in the population $(\lambda\widetilde{q} + [1-\lambda]q)$.

To understand how POs affect the ESS in the population, observe that as the PO changes size (λ), two effects take place. On the one hand, social learners inside become relatively fitter, as the difference between inside and outside social learners is given by $\lambda(c - \widetilde{c})$ (see equation 2). On the other hand, the PO reduces the average tuning in society ($\lambda\widetilde{q} + [1-\lambda]q$) by reducing the fitness of all social learners (both inside and outside). Eventually, the second effect dominates, making the benefit of POs decrease with size (figure 1B).

Another mechanism at play is that the PO increases the share of individual learning outside the PO (figure 1C). The PO makes the social learners outside worse off because they now are “forced" to copy members of the PO, who, like themselves, are liable to environmental change. This situation generates an increase in the number of individual learners outside, benefiting the PO. As λ grows, social learners inside the PO gradually substitute social learners outside.

These results suggest that early POs – e.g., hunting bands, Roman societas, medieval guilds, and Renaissance partnerships – probably had a key historical role in promoting hard-to-propel cumulative culture by means of making social learning less expensive and more exclusive. Evolutionary benefits of organizations are typically justified by multilevel selection (Turchin, 2016), where stable organizational heterogeneity plus competition leads to cultural progress. Our model points to a more basic advantage: through cooperation, identity and trust, exclusive POs facilitate social learning, making culture useful for society. This account provides a clear evolutionary origin to POs, one that puts knowledge and cooperation at center stage. This contrasts with extant theories of POs, most of which are ahistorical and focus on incentives and effort, assuming that a cultural tradition or pool of knowledge is already in place.

Our results also highlight the issue of the locus of innovation. In our model, innovation occurs outside POs. POs may still generate innovations that decrease the cost of social learning (e.g., unique language, enforcement devices, or adjusting the technology for improved replicability), but they do not generate the radical innovations (or inventions) needed to track the state of nature. Over history, radical innovations have tended to happen outside POs, for example,by inventors, entrepreneurs, individual scientists (which then might set up a PO to exploit the innovation). Recent evidence shows that this is also the case for modern firms. Arora et al. (2018) shows that most manufacturing firms, when introducing a new product, do so by mainly imitating other firms (and then tweaking/differentiating the product slightly), and if they innovate (i.e., the product is new to the world), then they source the invention from external independent inventors, such as scientists, independent labs/inventors, and users.

Finally, proposition 1 is robust to PO-biased social learning, that is, social learning inside (outside) the PO is preferentially executed inside (outside) the PO (i.e., the likelihood of social learning inside the firm is greater for members than for nonmembers of the organization).

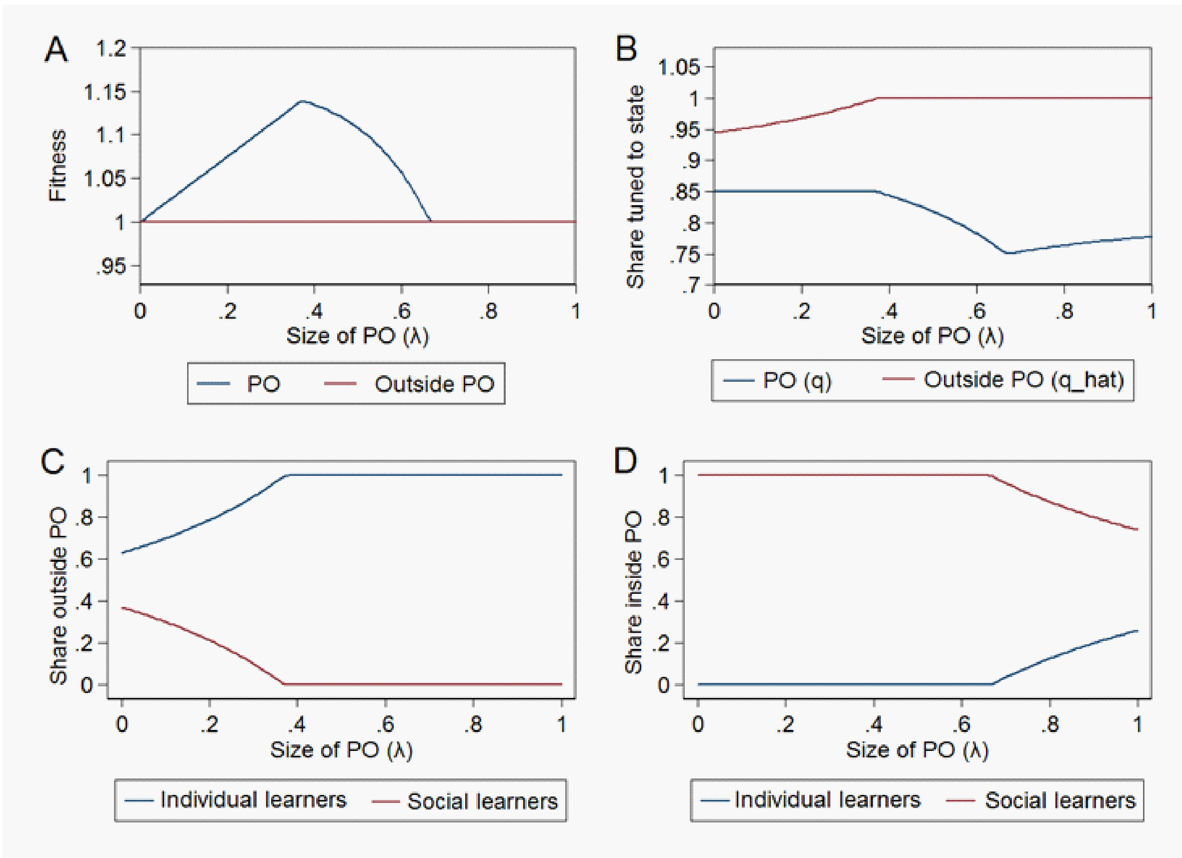

Figure 1: Equilibrium values of the model for different values of lambda. We use $C = 0.6$, $c = 0.45$, $\widetilde{c} = 0.3$, and $p = 0.1$. (A) Fitness inside and outside organization (we multiply fitness by 2.5 to obtain fitness equal or superior to 1). (B) Percentage of the population that has a technology tuned to the state. (C) Share of social and individual learners outside the organization. (D) Share of social and individual learners inside the organization.

2.2 Comparative statics

The model presents several interesting comparative statics, which are depicted in figure 2. Confirming the intuition that the fitness of the PO comes from its ability to facilitate social learning, we found that the fitness of the PO depends negatively on the cost of social learning inside the organization $\widetilde{c}$ (figure 2B). Empirically, this means that improvements in the ability of executing social learning inside organizations will impact positively their fitness. There is a large literature in organizational learning that provides evidence for this (Argote and Miron-Spektor, 2011). Less intuitively, we find that changes in the cost of social learning outside the PO (parameter c) do not translate into a monotonic change in the PO's fitness; instead, a decrease in c decreases the fitness of a small PO but increases the fitness of the PO if it is sufficiently large (figure 2A).

Another interesting prediction is that increases in uncertainty generate a decrease in the fitness of POs (figure 2C). Given that POs are populated entirely by social learning, it is easy to see that as the parameter p increases, then the PO will be increasingly liable to a change in the environment, reducing its expected fitness.

This result is opposite to what is predicted by theories of PO based on transactional cost economics (TCE), in which the value created by the firm increases with the uncertainty of the environment (Williamson, 1991; Tadelis and Williamson, 2013). Given this disagreement, it is interesting to assess the available evidence. Despite well-documented cases in which uncertainty favors the use of hierarchies (Forbes and Lederman, 2009), two meta-analyses of TCE show that uncertainty is slightly related to a decrease in the use of hierarchies in favor of the use of markets and hybrids, a sign of lower fitness of POs under higher uncertainty (Geyskens et al., 2006; Crook et al., 2013). The literature on industry and product life-cycles also provides consistent evidence: the size of firms is smaller at earlier stages of the cycle, when uncertainty is higher and product or technological standards are not yet defined (Klepper, 1996). We expand on the comparison with TCE in the empirical section below.

3 Extensions

In this section, we discuss four extensions to our basic setup. First, we study the impact of secrecy in the fitness of POs. Historically, many POs have been reluctant to share their knowledge with people outside the organization; for example, guilds were protective of their knowledge and techniques (Ogilvie, 2014). We show that secrecy is detrimental to the fitness of POs, but for a different reason that those highlighted in the literature on guilds (Ogilvie, 2014). Second, we extend our basic model by allowing for several competing technologies. We show that in the presence of multiple technologies,

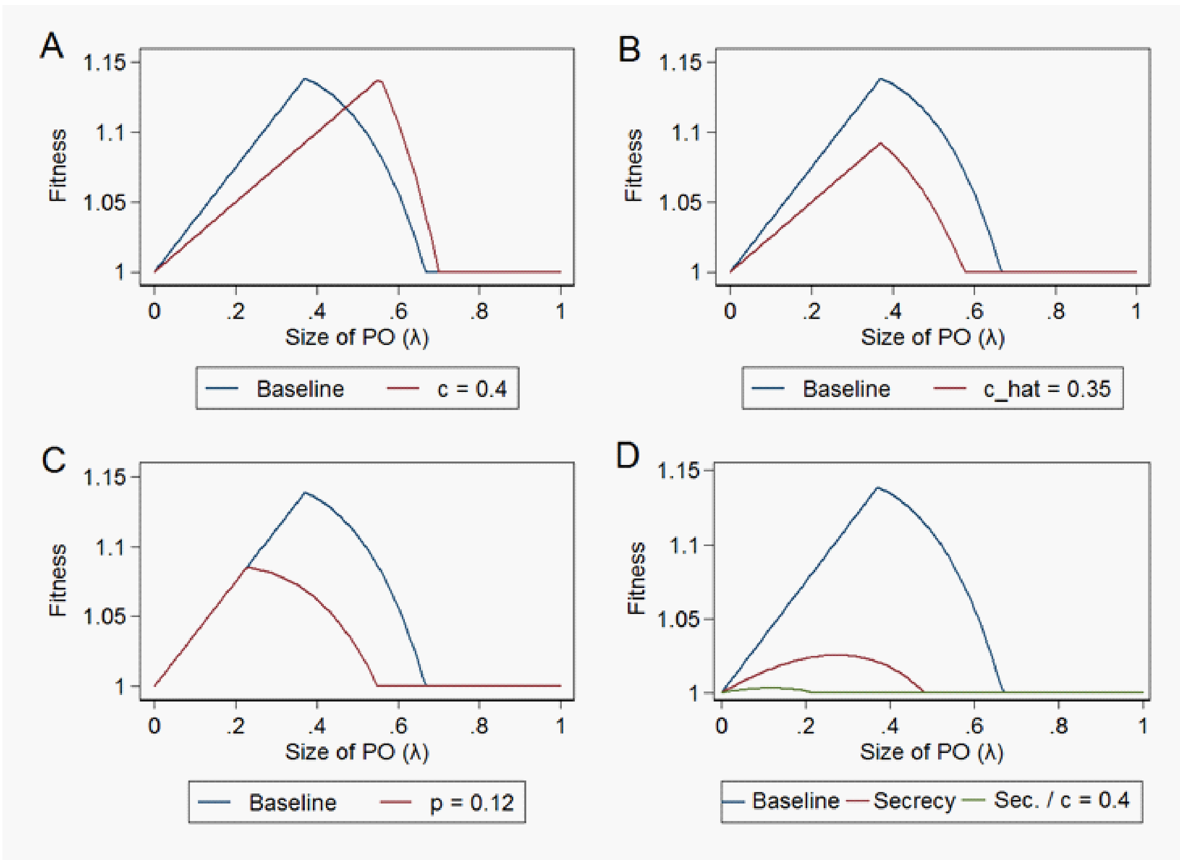

Figure 2: Comparative statics. For the baseline case, we set $C = 0.6$, $c = 0.45$, $\widetilde{c} = 0.3$, and $p = 0.1$ (as in figure 1, fitness is multiplied by 2.5). (A) We set $c = 0.4$. (B) We set $\widetilde{c} = 0.35$. (C) We set $p = 0.12$. (D) In the red line, we make the PO secretive, that is, social learners outside cannot imitate members of the PO. In addition, in the green line, we reduce the value of social learning outside of the PO to $c = 0.4$.

the PO specializes such that in equilibrium, all members of the PO adopt the same technology. Further, if many POs are allowed, then they specialize in different technologies. This result provides a novel explanation for specialization within and across organizations. Third, we briefly discuss the endogeneity of the size of the firm λ and its impact in our analysis. Finally, by adding the ideas of know-how and incremental innovation, we provide a broader interpretation of our model. This interpretation expands the positive impact that POs have on society.

3.1 Secrecy in productive organizations

In this subsection, we study the impact of secrecy on the fitness of POs. Secrecy mirrors the long standing concept of "boundary spanning" in organizational sociology (Scott, 2003). Given that setting a boundary is a sinequanon foundation of organizations – the lamda of our model – (Scott, 2003), a lot of attention has been devoted to the study of how organizations span the boundary, or, in contrast, why they might limit its transparency to outsiders. In terms of our model, the latter means that social learners outside the PO cannot imitate members of the PO. Consequently, outside the PO, we are back to the original case without a PO, in which the fitness of social learners outside the PO is

$$f_S = (1-p) \cdot q^e - c, \tag{3}$$

independent of the size of the PO. The fitness of social learners inside the PO is as before, given by

$$\widetilde{f}_S = (1-p)[(1-\lambda) \cdot q^e + \lambda \widetilde{q}^e] - [(1-\lambda)c + \lambda \widetilde{c}]. \tag{4}$$

In this case, it is not straightforward to determine whether f_s or $\widetilde{f}_S$ is higher. This is so because although inside the PO, individuals benefit from cheaper social learning ($c < \widetilde{c}$), individuals outside the PO benefit from copying relatively more tuned population ($q^e > \widetilde{q}^e$). In the case without secrecy, the PO imposes a negative externality on social learners outside the PO, such that the share of individual learners increases with λ (and social learners inside the PO gradually substitute the social learners outside the PO) (see figure 1C). In the case with secrecy, this negative externality is not present, and thus, social learners outside the PO copy more tuned agents on average (compared to social learners inside the PO).

It is relatively straightforward to show that in this setting, proposition 1 still holds, that is, provided the PO is sufficiently small, it increases the fitness of the population. However, society benefits less from secretive POs, and in order to be productive, their size needs to be smaller (see figure 2D). This is stated formally in the next proposition.

Proposition 2 *The positive impact of the PO on the fitness of the population decreases if the PO is secretive.*

Intuitively, the benefit of POs are now smaller because under secrecy, outsiders are not allowed to learn from the organization's members. Therefore, to make sure that in equilibrium, all members outside the PO have a fitness of $1 - C$, the share of social learners outside the PO increases, which in turn implies a negative externality for the PO because there are less individual learners to learn from. In other words, secrecy creates a relative disadvantage to social learners inside the PO because on average they learn from a less tuned pool than social learners outside the PO.

This situation also forces the sizes of POs to be smaller because this disadvantage is smaller when λ is small because then, the level of tuning inside the PO $\tilde{q}$ is closer to the level of tuning outside q (or alternatively, the relative tuning advantage of social learners outside the PO is lower).

Notice that, just like the case of uncertainty discussed above, the prediction of lower fitness and size of PO under secrecy runs opposite to what TCE would indicate (Williamson, 1991). Secrecy, as a symptom of weak appropriability regimes and threats of leakage, would be correlated in TCE with higher benefits of using POs, moving transactions away from market and increasing the size of POs. In the empirical section, we show which of these opposing predictions fares better for our data.

The finding that secretiveness decreases fitness appears to be at odds with the fact that historically, several POs have been secretive. For example, guilds tended to be protective of their knowledge and techniques (Ogilvie, 2014). This fact suggests that in order to evolve, secrecy in POs must have provided an additional benefit. We speculate that secrecy might generate a lower learning cost $\widetilde{c}$ because secrecy can galvanize the notion of "us vs. them" that promotes within-group identity, trust and cooperation (Bowles, 2009). It is also possible that the secretive organization evolved first, with the benefits of non-secretiveness being discovered later, perhaps in modern times.

An interesting finding is that secrecy is more detrimental (compared to non-secrecy) when the cost of social learning outside the PO is reduced (figure 2D, green line). This result can explain the trend towards more transparent and open organizations that has occurred in the last thirty years (e.g., Wikipedia). This trend has largely been a result of better communication technologies that generate a higher reduction in social learning costs outside POs (Benkler, 2006).

3.2 Specialization

We now add multiple technologies to our model. Suppose there are M technologies and as before, nature can change with probability p, rendering the technologies less useful. For example, a society living on an island with dry climate, and thus infertile soil, would exploit a myriad of sea technologies, such as net fishing, spearfishing, angling, shellfish harvesting, and boat building. If the environment changes, then the optimal means of executing each of these activities would change, making the techniques less beneficial. For example, a migration of larger fish might render current nets too weak, requiring stronger nets for effective net fishing.

Agents, as before, can be either social learners or individual learners, but now, their behavioral strategy also specifies an activity ($j \in J$) whose share in the population will depend on the replication dynamics. All activities are assumed to be ex ante equally productive, with fitness levels equal to the previous section. The behavioral strategies and proportion of tuned agents in the population now have a superindex $j \in J$ that specifies the activity. Social learning occurs randomly but is restricted to the set of people that executed the same technology j in the previous period. This requires the assumption that a social learner can identify people that executes the same activity that they do. Following our example, if a social learner uses net fishing, she will be copying agents in the population that executed net fishing in the previous period, some of which will be tuned – using the optimal net fishing technique for the current environment – and some of which will be not tuned – using a less optimal but still identifiable net fishing technique.

Replication dynamics in this model drives the fitness among social and individual learners and among all surviving activities to be the same. Because in this model there could be multiple (but qualitatively identical) equilibria, in order to make comparative statics, we focus our attention on the symmetric equilibria in which individual learners are uniformly distributed among activities. (This could be the result, for example, of uniform natural preferences of individual learners or some decreasing returns to scale in each activity).

The results of this model are straightforward and mirror the case with only one technology. Social learners are selected into the population, but the overall fitness of the society does not increase. Given our assumptions, individual and social learners will be distributed evenly across technologies. The total share of social learners will decrease with c and p.

We now allow for the existence of one productive organization of size λ. A social learner with activity j will now bear a cost $\widetilde{c}$ when imitating agents on activity j whom are sharing her PO, and cost c when learning from agents with activity j located outside her PO. Thus, inside the PO, social learners only copy agents that have their same technology. This captures the idea that is more difficult to learn across than within technologies[8].

To show our main result, define $x^j = r_S^j + r_I^j$ the share of agents that execute activity

[8] The results of this section also hold if we assume that social learners learn disproportionately from within the PO.

j outside the PO and equivalently define $\widetilde{x}^j = \widetilde{r}_S^j + \widetilde{r}_I^j$ the share of people outside the PO that executes activity j. The fitness of a social learner outside the PO is given by

$$f_S^j = (1-p)\left[\frac{\lambda\widetilde{x}^j}{\lambda\widetilde{x}^j + (1-\lambda)x^j} \cdot \widetilde{q}^e + \frac{(1-\lambda)x^j}{\lambda\widetilde{x}^j + (1-\lambda)x^j} q^e\right] - c. \tag{5}$$

Inside the PO, the fitness can be written as follows:

$$\widetilde{f}_S^j = f_S^j + (c - \widetilde{c}) \cdot \frac{\lambda\widetilde{x}^j}{\lambda\widetilde{x}^j + (1-\lambda)x^j}. \tag{6}$$

Observe that the advantage in fitness of social learners within the PO is increasing in the share of members in the PO that execute the same activity. Therefore, the evolutionary dynamics will drive social learners inside the PO to specialize in the same activity. This is stated formally in the next proposition.

Proposition 3 *Given a sufficiently small λ, the PO specializes in a specific technology.*

The intuition for this result is that the advantages of lower social learning costs increase when you have a larger group that can learn from each other and reap the benefit of cheaper social learning. The total costs of social learning will be reduced within the PO if everybody specializes in the same activity.

Although we don't develop the extension of the model here, we can add multiple POs to the society. We find that in this case society also benefits from their presence and that each PO will specialize in a specific technology.

The standard explanation for specialized productive organizations is that they allow to take advantage of economies of scale and comparative advantage by trading with other specialized POs. Our results suggest an alternative and likely complementary explanation. Even in societies that have not discovered trade (both internal or external) and in which comparative advantages are absent (i.e., all agents bear the same opportunity cost of doing any technology, or in our model, a constant c), specialization within productive organization will be beneficial. This is so because it maximizes the social learning advantage of POs, which in turn maximizes the fitness benefits that POs (of limited size) bring to society. In the empirical analysis, we provide evidence that specialized POs are beneficial even in the absence of trade.

It is interesting to point out that in standard economic theory, the benefits from firms driven by specialization and trade are increasing in the environmental uncertainty (Burgess and Donaldson, 2010), while the benefits from social learning we discuss in this paper are decreasing in environmental uncertainty (see figure 2C). Thus, it is possible to empirically test wether the benefit of having specialized POs is mostly obtained by social learning or trade by analyzing the impact of POs on fitness at different levels of environmental uncertainty. We discuss this point in more detail in the empirical section.

3.3 The size of the firm

So far, we have taken the size of the firm as exogenous. Although it is not difficult to endogenize size, we choose not to do it, in order to keep the message of the paper as simple as possible. Below, we provide a brief discussion regarding the main issues surrounding the size of POs.

First, observe that regardless of the size of the PO, people will always weakly prefer to belong to a PO. Therefore, to endogenize the size, it is sufficient to assume people will be admitted as long as doing it increases the fitness of the PO members. The model clarifies a crucial trade-off in this process: a small PO does not take full advantage of cheaper social learning, but a large PO loses track of environmental changes. Given that our model is continuous, it follows that there exists an interior size that maximizes the average fitness of its members (see figure 1A). Several mechanisms of modifying the number of members, will make the size of the PO to converge (via trial and error) to the size that maximizes the members' average fitness. These mechanisms can be purposeful (e.g., voting in a partnership) or exogenous (e.g., a deadly disease). In both cases, especially the latter, it is necessary to assume that PO members track the resulting fitness of a change and update λ as a result. It is important to notice as well that the PO, as it changes its size, retains its exclusivity condition. Endogenizing size means adjusting the size while keeping access restricted; otherwise, the size of the firm would be expanded until its benefits are completely diluted.

Second, given the continuity of our environment it is possible to show that the size of the PO that maximizes fitness (which is the natural candidate for a model with endogenous λ) is always small from a social perspective. To see this let $\widetilde{f}(\lambda) + (1 - C)$ be the fitness of agents inside the PO. At the level that maximizes fitness, it must be that $\widetilde{f}'(\lambda) = 0$. The overall fitness of the population is given by $1 - C + \lambda\widetilde{f}(\lambda)$; therefore, the derivative with respect to the size λ at the point that maximizes member's fitness is $\widetilde{f}(\lambda) > 0$, meaning that the society will benefit from increasing the size of the PO further.

Third, we deem likely that λ and $\widetilde{c}$ are connected. If POs can generate a lower cost of social learning through group identity, trust and cooperation, then it follows that a larger λ will probably increase $\widetilde{c}$, which in turn would reduce the size of the PO.

3.4 Extended interpretation

The interpretation of the model can be extended by acknowledging that in order to perform an activity (or apply a technology), both learning *and production* are required. One has to learn not only what to do – for example, learn that net fishing generates a good catch in the current environment – but also develop an understanding and capacity of how to do it – for example, know which are the best raw materials to produce the net, how to assembly it properly, how to use it best, when and where it provides more benefits, and so on. An individual learner has to figure out these two challenges, learning what to do and then refining how to do it; similarly, a social learner has to copy both elements and then execute properly.

By adding the production stage, we can introduce the idea of incremental innovation, defined as the process through which the execution of an activity can be improved bit by bit over time. Evidence from the organizational learning literature indicates that these accumulated improvements can be substantial (Argote and Miron-Spektor, 2011; Levitt, List and Syverson, 2013). Incremental innovation contrasts with innovation that is executed solely by individual learners, which can be equated to radical or disruptive innovation: they spend considerable effort to understand the new state of nature and then generate a *novel* and well-adapted technology. In addition, incremental innovations are not protected from changes in the environment: no matter how refined the technique has become, the environment can change in favor of a different technology.

By allowing agents to engage both in learning and production, the parameters C, c and $\widetilde{c}$ would now include not only learning costs but also the costs of production. The difference between C and c would also capture any advantages in production costs by social learners. If we assume that in every instance of social learning the learner can introduce incremental improvements on execution (which can then be passed down to other social learners), then the aggregate cost advantage generated by social learners can be very large.

Within this framework, POs can further generate benefits to society by providing an increased capacity to generate incremental innovations (which would be reflected in a larger difference between c and $\widetilde{c}$). This capacity can be positive and substantial if incremental innovation is complementary with the effectiveness of social learning. This is not unlikely, as it is easier to improve upon a well-replicated technology, which appears particularly true under specialization of POs. Although a full-fledged development of this idea is beyond the scope of this paper, this interpretation allows for a larger impact of POs: any improvement in incremental innovation they generate would be beneficial for society.

4 Empirical analysis

Our empirical analysis has five parts. First, we introduce the Ethnographic Atlas data and the variables measurement, and we show that the fitness of premodern societies is correlated with the presence of technologies only when the technologies are performed within POs. Second, we make the argument that there is a causal relation between POs and fitness by using instrumental variables and by using bounds that show that our results are not likely to the reverted by omitted variables. Third, we perform a comparative statics analysis to see how the benefit of POs depend on uncertainty and the cost of social learning. Fourth, we execute several robustness checks that allow us to show consistency across different dependent variables and rule out alternative explanations, such as the presence of trade. Finally, we replicate the results of Giuliano and Nunn (2017) in order to do an out-of-sample test of a separate prediction of the model: we show that use of POs by ancestors increases the persistence of tradition in modern populations.

The combination of this five empirical exercises provides compelling evidence that POs played an important role in making social learning adaptive as our theory predicts. Still, as indicated in the introduction, an important part of our empirical exercise is to provide a "proof of concept", that is, as a way to verify that the theory has practical potential and that it can be empirically productive.

4.1 Dataset and productive organizations

To test the predictions of the model, we use the Ethnographic Atlas (EA) (Murdock, 1967) and the Standard Cross Cultural Sample (SCCS) provided by the D-Place dataset (Kirby et al., 2016). The EA describes cultural practices for 1291 premodern societies, ranging from societies with complex agricultural economies and political systems to small hunter-gatherer groups. The societies are globally distributed, with especially good coverage of Africa and western North America. The SCCS is a subsample of the EA where additional information about societies is provided. We use the SCCS to measure several variables that are needed to test the predictions of the model. These datasets were created by coding the available information about societies that is present in the extensive ethnographic accounts in the anthropology literature.

The EA provides information about eleven productive activities (or technologies) in the society: metal working, pottery making, weaving, leather working, hunting, boat building, house construction, gathering, agriculture, fishing, and animal husbandry. The dataset identifies whether each activity was present in the society and, if so, whether it was "normally performed by many or most adult men, women, or both" or was "largely performed by a small minority who possess specialized skills". We identify the second condition as the addition of a PO to the execution of a specific activity. We measure PO in this way because the types of "small minorities" covered by the EA fit the requirements of our theoretical model[9].

The minorities in the EA can be of four types: senior age specialization (i.e., only men or women beyond the prime of their life), junior age specialization (i.e., only boys or

[9]In addition to providing a proxy of POs, the EA allows us to allign the empirical test with our evolutionary argument. We posit that the first POs evolved, in some ancestral time, because of their ability to foster the accumulation of culture. This argument about "origins" requires evidence coming from pre-modern societies.

girls before the age of puberty), craft specialization (which includes occupational castes where the rights to execute certain activity were inherited), and industrial specialization (i.e., specialization is removed from age or craft specialization and is executed using industrialized techniques). Aggregating across activities, craft specialization covers roughly 85 percent of the societies that have a minority executing the activity, industrial specialization accounts for eight percent (aprox.) and senior/junior specialization split the rest.

Our model requires that PO possess three characteristics in order to benefit society: easier social learning, small size and exclusivity[10]. In the discussion that follows, the lower costs of social learning of the PO of the EA become evident. Regarding small size, the very definition in the EA specifies a "small minority". Exclusivity requires more care to be mapped to the EA. Industrial specialization and senior/junior specialization comply with the exclusivity criteria. In the former, exclusivity is predicated on employment, and in the latter, it is defined by age. To understand exclusivity in craft specialization, we sampled the ethnographies in the EA. Roughly, there are three types of craft specialization. First, the original ethnographies describe organizations that could be described as "proto-guilds" –the most common type of craft specialization in the EA. Similar to medieval guilds, these organizations had experts, sometimes called "masters", and apprentices, which would come together regularly –or seasonally, in the case of fishing at high latitudes– in order to exchange work for teaching and to learn from each other. Apprentices typically needed to prove their capacity in order to fully access the community of experts, so access wasn't freely granted. Being a master often carried prestige in the society. Frequently, the right to execute a particular craft/activity was hereditary (e.g., the Chekiang society in China for fishing), generating occupational castes (or a specialized clan). This hereditary element compounded exclusivity because, even in this case, skill was also a frequent pre-requisite to enter the "proto-guild". The second type of "specialized minority" were the small and scattered workshops, where one or more skilful specialists, with the help of a handful of workers, would serve the needs of a portion of the society, typically the local town or region (for example, metalworking in the Riffian culture in Northern Africa). The third type is "attached specialists" where skilled crafstmen were appointed and funded by the rulers of the society (e.g., metalworking in the Inca Empire). Either by the selection of workers or their funding, the second and third types also seem to ensure exclusivity. All considered, even though there is heterogeneity in the "craft especialization" of the EA, the basic idea exclusivity in these organizations seems to hold ground[11].

[10]In what follows, the lower costs of social learning of the PO of the EA become evident. Therefore, we don't expand on it.

[11]However, in roughly a third or a quarter of the cases we sampled, the EA doesn't provide indication about the presence of a (exclusive) group of specialists; instead, it might indicate, for example, that

This characterization of pre-modern craft specialization in the EA –and our use of it as a proxy for PO– is consistent with the broader archeological literature. For example, in its review of the evidence Sterelny (2012) indicates that "craft expertise –the kind of skill sets that forager lives depend on– is fine-tuned at a generation and reliably transmitted across generations by this mode of *organized human learning environments*" (p. 35; emphasis added). When discussing the prominent example in Stout (2002), an ethnographic study of stone adze making in Irian Jaya, Sterelny indicates: "The social and informational organization of adze making is strikingly akin to a medieval guild. The apprenticeship system is quite formal. There is a master adze maker who has at least formal authority over the distribution of raw material to adze makers. Apprentices have to be accepted by a recognized master, and while apprentices are typically close relatives of their master, that is not sufficient. [....] The parallels with the formal, insitutionalized system of apprentice guilds could hardly be clearer." (p. 40-41). In metalworking, this guild like structure is prevalent in the literature (Rowlands, 1971). On top of this basic template, ethnographic studies provide additional evidence about the heterogeneity of craft production systems across societies (Costin, 2001, Rowlands, 1971).

There is a final correspondence between the POs of our model and the measure we use. In section 3.2 we indicate that our model can accomodate multiple POs which, in the presence of multiple technologies, would specialize in a particular technology. This is reflected our measure: the EA identifies are minorities that are specialized in a particular technology.

4.2 Baseline model

We computed two variables: the percentage of activities that are present in the society ("% presence") and the percentage of those activities that are executed within a PO ("% within PO"). In the dataset, there is missing information about the activities due, for example, to the fact that the ethnography did not study one or more productive activities. Only 263 societies had complete information about the eleven activities. The variable "% presence" is computed as the division of the count of activities that were present in the society over the count of activities for which we had available information[12]. The variable "% within PO" is computed as the division of the count

in each clan there was an specialist in a particular craft without specifying whether these specialists would come together as a group (or which were the rules of entry). This is referred to in the literature as "home production" (Costin, 2001). We believe that this might introduce a downward bias in our empirical analysis.

[12] If the ethnographic atlas indicates "missing data", then we would not consider that activity in the denominator. If the ethnographic atlas indicates "the activity is absent or unimportant in the particular society", then would not consider that activity in the count. The difference between these is that in the former, the original ethnography did not provide any indication regarding the activity.

of activities "largely performed by a small minority who possess specialized skills" over the count of activities that were present in the society. The relationship between "% presence" and "% within PO" is positive, with a correlation coefficient of 0.4 (see figure 4A). The variable "% presence" captures, partially, the cultural complexity of a society. A society with more activities has accumulated more culture over time.

To test the impact of the presence of activities and PO on the fitness of the individuals in society i, we use the following econometric model[13]:

$$Population_i = \beta_1 + \beta_2 \times \%Presence_i + \beta_3 \times \%Presence_i \times \%withinPO_i + Controls_i + Error_i$$

Population as a dependent variable captures the standard notion of fitness as reproductive success. It also captures the fact that in premodern Malthusian economies, progress translates into increases in population and not per-capita wealth (Spolaore and Wacziarg, 2013; De La Croix et al., 2017)[14]. We use "size of local communities", which is a categorical variable with 8 categories: 1 is "less than 50 people" 2 is "from 50 to 99 persons", 3 is "from 100 to 199 persons", 4 is "from 200 to 399 persons", 5 is "from 400 to 1,000 persons", 6 is "more than 1,000 in the absence of indigenous urban aggregations", 7 is "one or more indigenous towns of more than 5,000 inhabitants but none more than 50,000", and 8 is "one or more indigenous towns with more than 50,000 inhabitants". In figures 4B and 4C, we plot the "size of local communities" against the variables "% presence" and "% within PO".

As controls, we added geographical variables (absolute latitude, average temperature, distance to coast, and slope of terrain), resource endowment variables (amphibian richness, bird richness, mammal richness, and vascular plants richness), intensity of agriculture dummies (complete absence, casual, extensive/shifting, horticulture, semi-intensive, and intensive), region dummies (36 regions in our final sample), type of settlement dummies (e.g., nomadic, semi-nomadic, semi-sedentary, shifting fixed settlements, neighborhouds of dispersed family homestead, separate hamlets forming a community, established villages, and complex settlements) and year of the ethnographic record. From the 263 societies with complete information about the technologies, we lost some

[13] In this model, we do not include an individual term for "% within PO" because this variable is nested within the presence of technologies (e.g., when "%presence" is zero, then "%withinPO" is zero as well). If this individual term were included, then it would mean that even if "%presence" is zero, "%withinPO" can have an impact on population, and this would be contradictory. As a result, the marginal effect of "%withinPO" is scaled by the variable "%presence".

[14] The model assumes an exogenous and fixed set of long-lived agents, of which a small share in every period copies the most succesful strategy (known as quasi birth and death process). Thus, it could be indicated that our dependent variable doesn't map to our model. However, our model yields the same results if we assume that agents' life spans only one period. The number of offsprings of the two types would be proportional to their fitness, with a small advantage for the type with highest fitness. Therefore, in this alternative specification of the model, an increase in population represents an improvement in the fitness of the society.

Figure 3: Societies included in table 1

societies due the missing data for the dependent variable (54 societies) and in some controls (mostly, resource endowment), leading to a final sample of 173 societies. In figure 3, we display these societies on a geographical map. In the appendix we display detailed information about these societies.

The ordered probit estimates are presented in table 1. Assuming that there are no POs, the results presented in column 2 of table 1 show that moving from 0% to 100% presence of activities does not generate an increase in the local population. This is depicted by the blue line with circle markers in figure 4D. The presence of a wider set of activities in most of the adult population does not translate to a larger population. Although this result might seem surprising, it is consistent with Rogers' paradox, in the sense that culture does not necessarily leads to increased fitness. However, consistent with proposition 1 of the model, activities do increase the local population when POs are present in the society. This increase is economically and statistically significant and is depicted by the red line with triangles in the figure 4D.

Table 1: Impact of presence of technologies and PO on the size of local population

	Dependent variable: Size of local population	
	1	2
% presence	1.137 (0.818)	0.016 (0.846)
% presence x % within PO		4.298*** (1.278)
Geographic controls?	Yes	Yes
Resource endowment controls?	Yes	Yes
Year of ethnography?	Yes	Yes
Agriculture intensity dummies?	Yes	Yes
Continent dummies?	Yes	Yes
Type of settlement dummies?	Yes	Yes
Observations	173	173
Pseudo R Square	0.329	0.352

Notes: We execute ordered probit regressions. The dependent variable is the size of the local population. Robust standard errors are used in all regressions and are reported in parentheses. *** indicates p-value<0.01.

4.3 Endogeneity

As difficult as it is with this type of data, in this section, we address concerns about endogeneity in our key variables.

4.3.1 Omitted variables

The first threat to identification of causality is omitted variables. We executed a test that uses selection on observables to assess the extent to which selection on unobservables would need to be in order to overthrow the results (Oster, 2016). In table 2, we replicate column 2 of table 1 and columns 2, 5, and 8 of table 6 (see section 4.4.1 for a robustness check with different dependent variables). We report the "Oster delta" in the last two rows of the table, assuming maximum R-square values of 1 and 0.95 respectively. Given the inherent measurement error of ethnographic data, the assumption of 0.95 is a good benchmark for the test (and perhaps conservative). This test assumes a linear model, so we estimate columns 1 and 3 using OLS (in section 4.4.1, we use the more appropriate ordered probit estimation for these dependent variables). The results show that the Oster delta is on average 0.97 when an R-square of 0.95 is assumed. This means that selection on unobservables would need to be at least 0.97 times the selection on observables in order to overthrow our results. A delta of 1 is a good indication against the threat of omitted variables, particularly if a comprehensive set of controls is used (Oster, 2016).

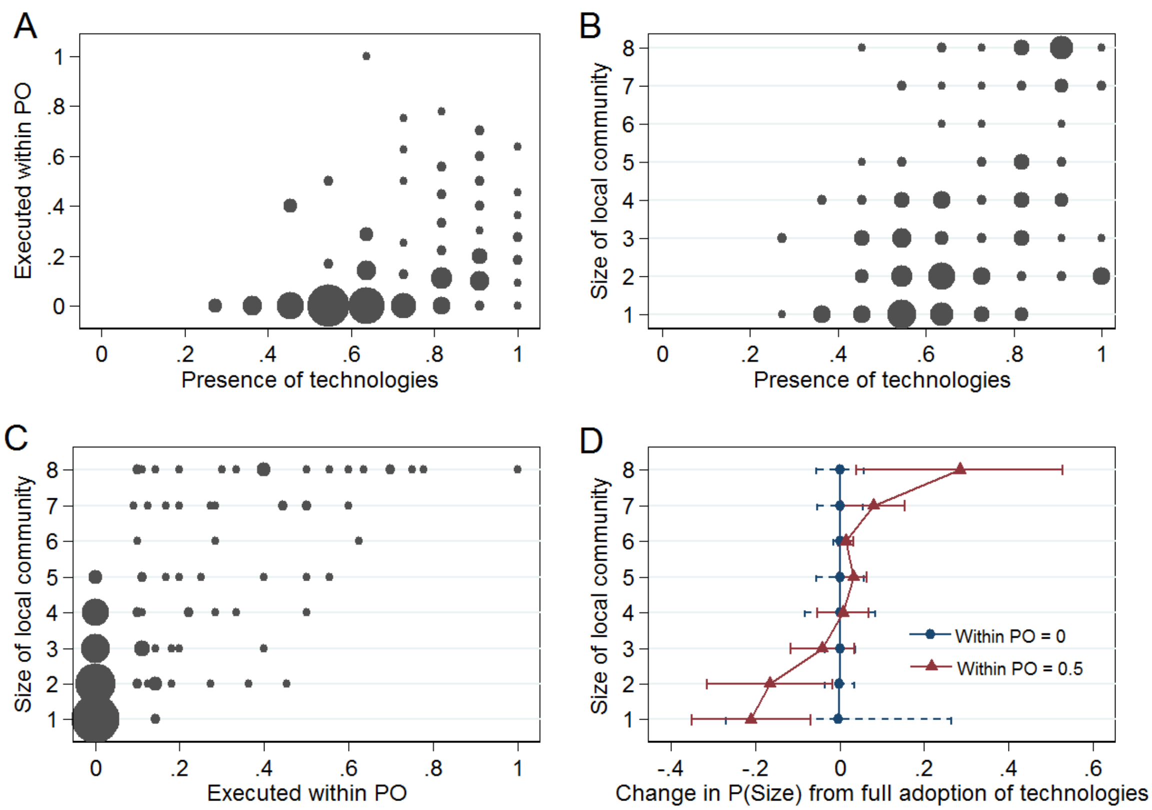

Figure 4: The impact of the presence of activities and POs on the size of local population. The figures use the sample used in table 1. (A / B / C) Scatter plots in which the sizes of the bubbles represent the frequency of societies. (D) Here, we plot in the second column of results in table 1. We evaluate how much the probability of each one of the eight size categories changes if the presence of activities increases from 0 to 1. We present the average of the marginal effects. To explore how this impact varies with POs, we set the variable "% within PO" equal to zero and equal to 0.5.

Table 2. Selection on observables versus selection on unobservables

	Dependent variable			
	1	2	3	4
	Local community size	Total population size	Population density	Cultural complexity
% presence	-0.143	-0.564	-0.669	-0.029
	(1.166)	(1.997)	(0.939)	(5.155)
% presence x % within PO	5.238***	5.818***	3.159***	14.887***
	(1.592)	(2.241)	(1.202)	(5.650)
All controls of table 1?	Yes	Yes	Yes	Yes
Observations	173	153	125	125
R-Square	0.771	0.819	0.409	0.888
Oster delta (R^2 max = 1)	0.41	0.52	0.71	0.93
Oster delta (R^2 max = 0.95)	0.51	0.69	0.98	1.69

Notes: Robust standard errors are used in all regressions and are displayed in parentheses. *** indicates p-value<0.01, ** indicates p-value<0.05, * indicates p-value<0.1. The Oster delta is computed for the interaction term "%presence x within PO"

4.3.2 Reverse causality

The second threat to identification is reverse causality. This threath can be present both in the presence of activities and in the use of POs. We analyze each one in turn.

An important proposition of cultural evolution is that a larger and more interconnected population would generate more cumulative culture (Henrich, 2015). In our case, this would translate into a higher presence of activities in the society, which could be channeled through the interaction with POs (assuming, for now, exogeneity in POs). This could generate an upwards bias in our baseline estimation of the impact of POs. To address this issue, we instrumented "% of presence" using two variables: "sex differentiation" (which we detail in section 4.3.2) and the index of "kinship tightness" developed by Enke (2017). For the first instrument, there is evidence that the presence of activities coevolved with sex differentiation in premodern societies (Haun and Over, 2013). For example, men specialize in large game hunting, while women specialize in gathering. Sex differentiation might affect population size if it affects fertility. We contend that conditional on the total amount of activity executed by women, and therefore controlling for the time restriction that differentiation might place on female fertility, the exclusion restriction should hold.

Kinship systems regulate the pattern of relatedness in society through family structure (e.g., independent nuclear families vs. extended families, post marriage residence in wife or husband's group vs. independent residence), marriage patterns (e.g., cousin marriage allowed vs. forbidden, polygamy vs. monogamy), and descent (e.g., unilineal vs bilateral descent group, presence of clans sharing a geographical location). Kinship tightness is a key variable affecting social organization of a society (Enke, 2017). Tight kinship (i.e., extended families, cousin marriage allowed, clans, polygamy, unilineal descent, etc.) generates high in-group bias, less cooperation with outsiders, strong conformism, and local institutions. The opposite occurs with loose kinship, with the consequence of being much more open to external groups. We argue that kinship tightness affects the presence of activities in society. A tight kinship system should increase the presence of basic and widely known activities through less reliance on sourcing activities from neighboring societies[15]. A "closed" society does not have an alternative but to provide the basic activities internally. A society with loose kinship, and that is therefore open, can source part of the basic activities from neighbors. The exclusion restriction for "kinship tightness" is sustained on the documented ancestral origins of kinship systems (Passmore and Jordan, 2017). Kinship systems can be traced back into the societies from which the focal society descent from. Thus, this element of societies can be treated largely as an exogenous variable, particularly when controlling for agriculture intensity and settlement type[16]. Furthermore, there are no a priori reasons to think that kinship tightness might generate larger or smaller populations through changes in fertility. For example, the polygamy-fertility literature is not at all conclusive. Accordingly, and consistently with Enke (2017), we do not find a relationship between kinship tightness and population in our data, conditional on covariates.

In the table 3, we present the instrumental variables estimations. In column 1, we present the first stage. As expected, both sex differentiation and kinship tightness are positively related to the presence of technologies. (Kinship tightness has a p-value of 0.16; if we drop sex differentiation, the p-value increases to 0.04.) By comparing the Cragg-Donald F-test of the first stage (reported in columns 2 and 3) with the values in Stock and Yogo (2002), we can conclude that our instruments are not weak. The Hansen-test indicate that the instruments are indeed exogenous, in line with the theoretical arguments laid out above.

In columns 2 and 3, we present the second stage, with and without the interaction with % within PO respectively. The results do not change from those of the table 1: the presence of technologies increases the local community size, but only when PO are in place. In columns 4 and 5, we use total population of the society as the dependent

[15] A society with high kinship tightness tends to be more isolated from neighboring societies. This would increase the need for having all the activities provided inside the society. A society with low tightness would be much more willing to provide some of the activities from abroad. Notice that this would hold for basic activities that have long being invented and diffused. For the case of innovation of newer and more complex activities, the impact of kinship tightness is detrimental to the adoption of innovations from other societies, as shown in Enke (2017) .

[16] There are arguments and evidence that indicates that kinship tightness evolved to optimally match the needs of agricultural subsistence, away from nomadism (see Enke, 2017).

variable (see section 4.4 below for the details of this alternative dependent variable). The results show that the presence of technologies has a positive impact independent of the percentage executed within PO. However, consistent with our prediction (and column 2 of table 6), this impact is larger when POs are in place. In column 5, if we assume absence of POs, then the impact of presence of technologies is statistically not different from zero.

Table 3. Instrumenting the presence of technology

Dependent variable:	1st stage. "% presence"	2nd stage Local community size		2nd stage Population size	
	1	2	3	4	5
% presence		4.793	2.958	12.982**	6.221
		(4.219)	(3.791)	(6.492)	(5.114)
% presence x % within PO			5.475***		10.865***
			(1.217)		(2.669)
Kinship Tightness	0.061				
	(0.044)				
Sex differentiation	0.119**				
	(0.053)				
Same controls as in table 1?	Yes	Yes	Yes	Yes	Yes
Region dummies and res.endow.controls?	No	No	No	No	No
Observations	194	194	194	160	160
Cragg Donald f-test first stage (p-value)		9.55***	9.62**	8.884**	8.085**
		(0.008)	(0.022)	(0.012)	(0.044)
Hansen test (p-value)		0.309	0.707	3.438*	3.843
		(0.578)	(0.702)	(0.064)	(0.146)

Notes: Robust standard errors are used in all regressions and are displayed in parentheses. *** indicates p-value<0.01, ** indicates p-value<0.05, * indicates p-value<0.1. In columns 3 and 5, we use "kinship tightness x % within PO" and "sex differentiation x % within PO" as instruments for "% presence x % within PO". We drop the controls of geographical region because the local geographical variation in our instruments is not high. Given that we rely on inheritance from ancestral societies, the societies occupying a particular region tend to share the several cultural traits from their common ancestor. We drop the resource endowment variables in order to avoid data loss and to avoid small sample bias in the IV estimation (results are consistent if we include these controls).

The third and main identification threat is endogeneity problems with the variable "% within PO". In particular, in case there is a minimum size for POs, large populations might make it easier to have POs. In addition, a common argument is that specialization is favored by the extent of the market. If any of these cases is correct, we might have a reverse causality problem which would generate an upward bias in our baseline estimation. We address this issue by instrumenting the presence of POs following the idea of Depetris-Chauvin and Ozak (2017). These authors explore the drivers of the presence of POs in premodern societies using the ethnographic atlas. They explore the extent to which diversity in the population of a society – measured by genetic diversity in the societies of the Atlas – drives the presence of POs. The theoretical argument is that a genetically diverse population has many different skills in place, which would lead to the creation of specialized groups. These authors instrument genetic diversity using the distance of the society from East Africa (specifically, modern day Ethiopia), which is the origin of the spread of the human species out of Africa (starting approximately 80,000 years ago). As the distance from Africa increases, the diversity within a society goes down, a phenomena known as serial founder effect (Ramachandran et al., 2005)[17]. The authors find substantial evidence in favor of their arguments: distance reduces genetic diversity, which in turn decreases the presence of POs.

In table 4, we follow these authors and use "Distance from Africa" as an instrument for "% within PO".[18] We measure the distance from Addis Adaba in east Africa to the focal society; for societies in America, we calculate the distance going through the Bering strait. We do not use the mediating variable of genetic diversity, and thus, we implement the "reduced form" model of Depetris-Chauvin and Ozak (2017)[19]. In column 1, we present the first stage. Consistent with Depetris-Chauvin and Ozak (2017), the distance from Africa reduces the presence of POs in societies. Although the Cragg-Donald test (reported in columns 2 and 3) indicates that the instrument is relevant, by comparing the values with Stock and Yogo (2002), we cannot rule out weakness in the instruments. To address this issue, in the second stage of columns 2 and 3, we use the limited information maximum likelihood (LIML) technique, which partially mitigates the problem of weak instruments. The results that we obtain with both dependent variables are consistent and supportive of our predictions. By comparing the values of the coefficient with those of table 1 and table 6, we also find that the coefficients exhibit increases in their size.[20]

[17] This also implies that cultural heterogeneity across societies would increase with the migratory distance to a common ancestor. This prediction is corroborated by Becker et al. (2018).

[18] We are using "distance from Africa" as an instrumental variable therefore our results cannott be interpreted as finding any causal effect of genetic diversity on any economic variable in the society.

[19] This reduced form allows for other mechanisms to impact the presence of POs. For example, given the nonrandomness of the migratory sampling process, it could also be the case that traits are lost. As migrant groups are typically small, the likelihood of loss increases due to drift.

[20] We also instrumented "% presence" and "% within PO" at the same time. We used the three instruments simultaneously following Wooldridge (2010, chapter 8). The results are consistent with table 3 and table 4. The coefficient for "% presence" is 1.32, and that for "% within PO" is 3.83.

Table 4. Instrumenting the percentage within PO

Dependent variable:	1st stage. "% presence x % within PO"	2nd stage Local community size	2nd stage Total population
	1	2	3
% presence	0.278***	-0.408	-1.316
	(0.082)	(1.223)	
% presence x % within PO		6.606*	17.70***
		(3.751)	(6.071)
% presence x Distance from Africa	-5.37e-06**		
	(2.73e-06)		
Same controls as in table 1?	Yes	Yes	Yes
Region dummies?	No	No	No
Observations	173	173	153
Cragg Donald f-test first stage (p-value)		5.228**	5.790**
		(0.022)	(0.016)
Hansen test (p-value)		n/a	n/a

Notes: Robust standard errors are used in all regressions and are displayed in parentheses. *** indicates p-value<0.01, ** indicates p-value<0.05, * indicates p-value<0.1. We use LIML in the estimations. We exclude region dummies because the variation of our instruments within regions is low.

An additional way to address the problem of reverse causality is that both theoretical arguments (i.e., minimum size of PO and extent of the market) would predict that the positive relationship between population and the presence of PO would be much stronger at higher levels of population. However, in figure 4D, we show that this is not the case. The impact of PO on the increase in population size is exerted throughout the different size categories of the variable "size of local community". For the case of the dependent variable of total population, we replicated column 2 of table 6 using a quantile regression estimation (and dropping the region controls, as they limit the estimation). In figure 5A we display the value of the coefficient related to "% presence" as it varies across the dependent variable. In figure 5B, we do the same for the interaction term "% presence x % within PO". In both cases, the graphs show that the positive impact of POs on population is exerted evenly across different population sizes, reducing the concern about reverse causality.

However, statistical significance is lost. The instruments retain their properties: for strength, they surpass the Stock and Yogo thresholds on strength; for exclusion, the Hansen test indicates that the instruments are valid.

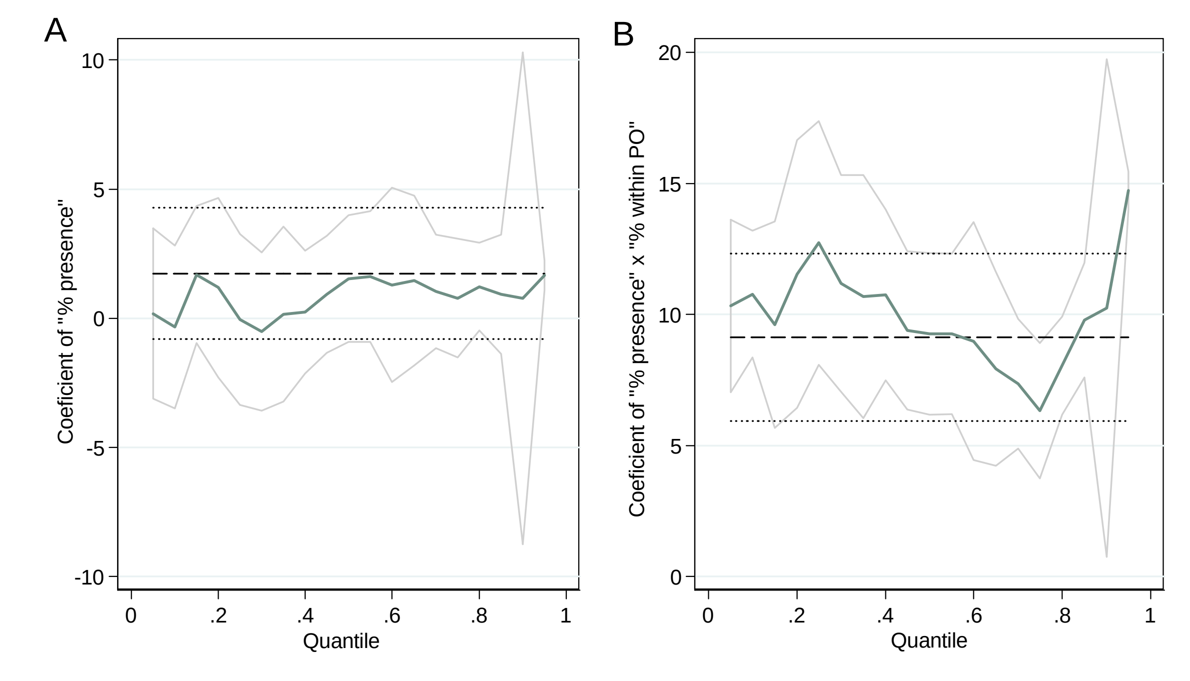

Figure 5: Coefficients in quantile regression. Dashed lines represent OLS coefficient, Continous lines represent the coefficients from the quantile regression.

4.4 Comparative statics

In this section, we test the comparative statics derived from our model, which are summarized in figure 2. The econometric model that we use is the following[21]:

$$
\begin{aligned}
Population_i \;=\; & \beta_1+\beta_2\times\%Presence_i+\beta_3\times\%Presence_i\times\%withinPO_i+\beta_4\times\%Presence_i\times Z_i \\
& +\beta_5\times\%Presence_i\times\%withinPO_i\times Z+\beta_6\times Z_i+Controls_i+Error_i
\end{aligned}
$$

In this model, we generate a triple interaction to explore whether the impact of POs is affected by the variable Z. We use different variables as Z in order to proxy for the different parameters p, $\widetilde{c}$ and c, in addition to the prevalence of secrecy. If the coefficient β_5 is positive (negative), then the impact of PO is enhanced (diminished) by the variable Z. Of course, the mapping between our proxies and the parameters of the model is not perfect. However, we believe that if *several* imperfect proxies confirm the comparative statics of the model, further confidence can be gained that the empirical findings reflect our model. As indicated above, the accumulation of consistent evidence serves to compensate the potential identification weaknesses in any individual test (i.e., bounds, IV, several comparative statics, robustness checks, out-of-sample test).

[21]For the same reasons explained in footnote 8, the interaction term between "%withinPO" and Z is not included. To appropriately saturate the model, all the remaining interactions are included.

4.4.1 Uncertainty

We first test the impact of environmental uncertainty, the parameter p of the model, by using climate unpredictability as the proxy. Climate data have already been successfully used to empirically test the parameter p in cultural evolution models (Giuliano and Nunn, 2018). The D-PLACE dataset reports "temperature unpredictability" and "precipitation unpredictability", which are measured using yearly data between 1901 and 1950, the period that has the largest proportion of ethnographies in the ethnographic atlas. The measure of unpredictability captures the extent to which temperature or precipitation patterns are predictable because these conditions are constant or whether they oscillate in a predictable manner (Colwell, 1974). We multiply these two measures to obtain our measure of climate unpredictability (if used individually, the results do not change). Consistent with the comparative statics of the model, the results from the column 1 of table 5 show that the impact of POs on population decreases when climate unpredictability is high. A joint t-test shows that the impact of PO is again highly significant and, importantly, moderated by climate unpredictability. This result is shown in figure 6A.

4.4.2 Social learning costs

We studied three variables that decrease the costs of social learning. First, the SCCS provides information about how rooted apprenticeship and teaching are in the society. "Apprenticeship" is a dummy variable that we computed from the variables v427 and v428 of the SCCS that measure the extent of guidance and/or formal schooling in late boys and girls, respectively. The dummy takes the value of 1 when either variables indicate that the society displays "predominant apprenticeship", or when "formal schooling is frequent and typical", and zero otherwise. Clearly, if schooling and apprenticeship are predominant in society, this will decrease both $\widetilde{c}$ and c. Given that a lower c has an ambiguous impact on the fitness of POs (see figure 2A), but a lower $\widetilde{c}$ has a unequivocal increase in the fitness of POs (figure 2B), we predict that "apprenticeship" should boost the impact of POs on population size. (It could also be argued that this dummy would be more tightly connected to a decrease in $\widetilde{c}$ than one in c because teaching and apprenticeship probably coevolved with POs.) Many hunter-gatherer societies –which lack POs– possess less teaching (relative to more advanced sedentary societies), and it is restricted mostly to kin (Hewlett and Roulette, 2016). This possibility would reinforce our prediction. The results are presented in column 2 of table 5 and are consistent with our prediction: the positive impact of POs on population is stronger if apprenticeship is predominant. We graph this result in figure 6B[22].

[22] It could be correctly argued that our measure of PO already captures apprenticeship (see discussion above). We counter this in two ways. First, there are many societies that have apprenticeships but

The second variable that reduces cost of social learning is “sex differentiation". The variables v44 to v54 of the EA provide information about the extent to which the eleven activities are executed by women and/or men. For each activity we coded a dummy that took the value of 1 in case the activity was executed by “males only or almost alone" or by “female only or almost alone". Then, we added these dummies and divided the result by the total number of activities that have available information. We label this variable “sex differentiation", and it captures the percentage of activities that are performed by either sex exclusively. There is plenty of evidence that social learning is facilitated by similarity, in which sex plays an important part (Henrich, 2015, p44; Fairlie et al., 2017,Losin et. al., 2012). Similar to “apprenticeship", this variable reduces both $\widetilde{c}$ and c, and therefore, our model predicts that more sex differentiation would lead to an increase in the impact of POs. This result is what we find in our estimations. In the third column of table 5, we obtain a positive and significant coefficient for the interaction term.

The third variable that reduces the cost of social learning is “Trust". We use the variable v335 of the SCCS, which measures the degree to which trust in inculcated in childhood in the society. This variable is ordinal, with 0 meaning “no inculcation or opposite trait" and 9 meaning “extremely strong inculcation". As with “apprenticeship" and “sex differentiation", high “trust" decreases both $\widetilde{c}$ and c, leading to the prediction of a higher impact of POs. The result is displayed in column 4 of table 5 and is consistent with the prediction from the model: POs have a larger positive impact on population when trust is high. This result is shown in figure 6C.

4.4.3 Secrecy

Finally, we analyze the impact of the variables “Honesty" and “Generosity". These are the variables v336 and v334 of the SCCS and are analogous to v335, namely, a categorical variable identifying inculcation of honesty and generosity in childhood. It is possible to identify these variable with a decrease in the degree of secrecy in the POs and therefore a boost in their fitness. Oftentimes, secrecy is related to selfish behavior, a desire to keep useful knowledge proprietary. The zeal to maintain secrecy could also benefit from dishonest behavior, deflecting requests to share knowledge with negation of its possession. It would also be possible to relate “honesty" and “generosity" to a decrease in the costs of social learning. When agents are generous and honest, it is very likely that communication and learning would improve. In both cases, the prediction from our model is clear: these variables should increase the fitness benefits of POs.

do not have POs (or that have a reduced number). Second, our measure of PO also capture crucial other elements of PO: size and exclusiveness. Therefore, we see that this interaction completes the third element we need in our POs, namely, a lower cost of social learning.

The results are presented in columns 5 and 6 of table 5 and are consistent with the prediction of our model. In figure 6D, we present the results for generosity.

Table 5. Heterogeneity in the impact of POs

	Dependent variable: Size of local population					
	1	2	3	4	5	6
% presence	-3.24 (3.62)	-0.536 (1.590)	0.125 (1.241)	1.112 (1.694)	1.364 (1.701)	0.973 (2.679)
% presence x % within PO	12.28# (8.56)	2.629*# (1.547)	-0.811# (4.684)	0.066 # (2.920)	0.690 # (2.999)	1.898 # (3.231)
Climate unpredictability	-3.25 (4.06)					
Clim. unp. x % presence	4.75 (5.03)					
Clim. unp. x % pres. x % within PO	-10.66# (11.38)					
Apprenticeship		-0.553 (1.129)				
Appren. x % presence		0.237 (1.556)				
Appren. x % pres. x % within PO		3.336*# (1.860)				
Sex differentiation			not included			
Sex differ. x % presence			-0.078 (1.216)			
Sex differ. x % pres. x % within PO			6.955# (6.468)			
Trust				0.250 (0.179)		
Trust x % presence				-0.275 (0.314)		
Trust x % pres. x % within PO				0.768 # (0.644)		
Honesty					0.313 (0.271)	
Honesty x % presence					-0.622 (0.392)	
Hon. x % pres. x % within PO					1.148*# (0.652)	
Generosity						0.574** (0.289)
Gen. x % presence						-0.517 (0.426)
Gen. x % pres. x % within PO						0.598 # (0.615)
Resource endowment control?	Yes	No	Yes	No	No	No
All other controls of table 1?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	173	136	173	101	84	79
Pseudo R Square	0.354	0.335	0.353	0.239	0.281	0.299

Notes: Robust standard errors are used in all regressions and are displayed in parentheses. *** indicates p-value<0.01, ** indicates p-value<0.05, and *** indicates p-value<0.1 . # indicates p-value<0.01 for joint two-tailed F-test of $dsize/d\%withinPO$. To avoid losing excessive observations by using the SCCS variables, the sample in columns 2, 4, 5, and 6 allows at least 6 technologies with available information.

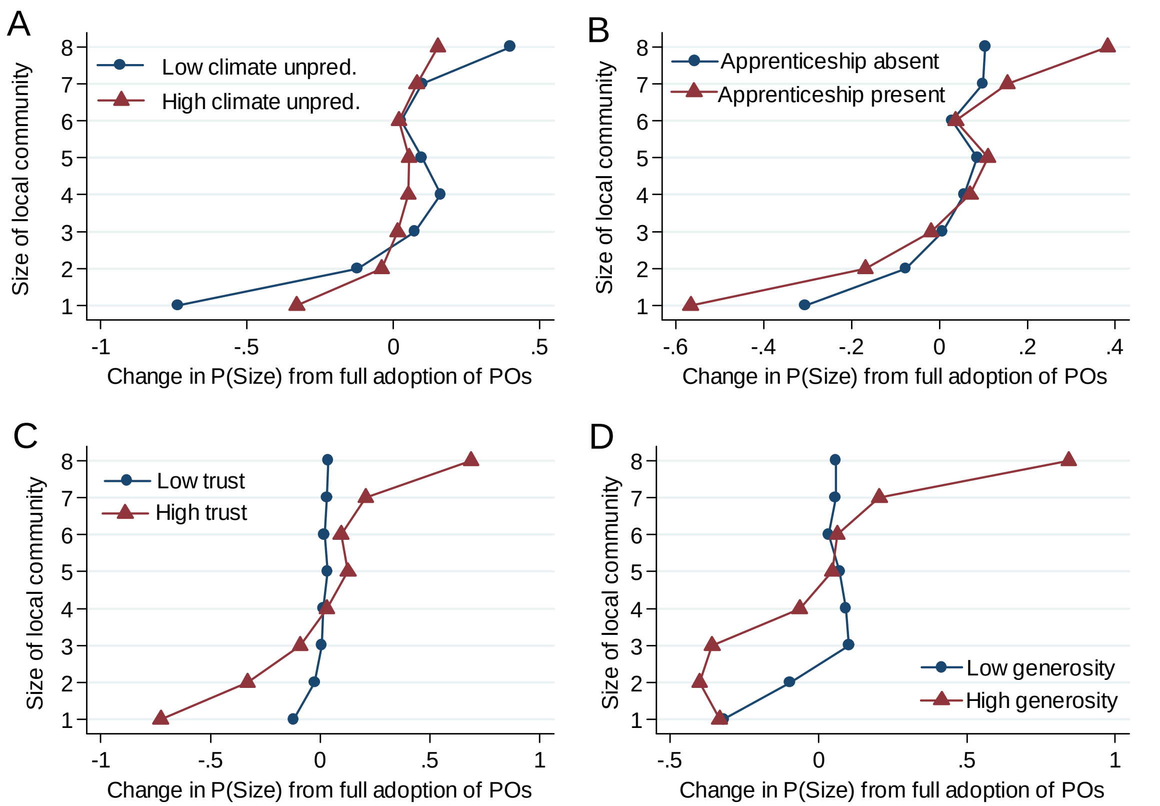

Figure 6: The impact of POs varies. In this figure, we analyze the predictions from the comparative statics of our model. We use the estimation of table 5. (A) We plot the average of the marginal effects of "% within PO", that is, the change in the probability of each one of the eight size categories as this variable goes from 0 to 1. To show how this impact varies with climate unpredictability, we set this variable to the minimum and maximum values in the sample used in the estimation. (B / C / D) Analogous to graph A. Confidence intervals are not displayed due to multicollinearity in the estimates (see the legend of table 5).

4.4.4 Comparison with transaction cost economics.

The results we obtain for the comparative statics – except for apprenticeship – are opposite to the ones that TCE would predict (Williamson, 1991; Tadelis and Williamson, 2013). TCE predicts that uncertainty would increase the benefits and the frequency of POs, as opposed to market exchange; trust and reputation would move transactions away from hierarchical POs and towards market or hybrid governance; the risk of leakage (and thus secretive behavior) would lead to heavier reliance on POs[23]. As discussed in sections 2.2 and 3.1 above, our theory provides opposing predictions, which, at least in our data, are supported.

Given the empirical support for the mechanisms proposed by TCE (Lafontaine and Slade, 2007), how can we make sense of this divergence? We speculate that this might indicate that when studying the evolution of POs, the explanatory logic changes. In our theory, we do not consider incentives and governance issues in order to focus on the impact that POs have on cumulative culture, that is, on the increase in the pool of useful knowledge and technologies. In contrast, TCE assumes a pool in order to focus on the governance of the transactions. Consistently, the unit of analysis in cultural evolution is the population, while in TCE, it is the PO and its transactions. Thus, these theories need not be contradictory; instead, they can complement each other by operating at different levels and time scales. For example, the advantage in social learning costs of POs might have its roots not only in identity and self-enforced cooperation but also in hierarchical governance devices that minimize the hazards of knowledge transmission. The exploration of these different explanatory logics, and how they interact, is a topic for further research.

4.5 Robustness checks

4.5.1 Robustness to alternative dependent variable

The results presented in table 1 are robust if the following three alternative dependent variables are used: "total population", "population density", and "cultural complexity". We address each one in turn. The variable "total population" is obtained from the Ethnographic Atlas and is a continuous variable that indicates the total population of the society. We use natural logarithms to normalize its distribution. In the columns 1, 2 and 3 of table 6, we display the results. To assess robustness of the comparative statics, in column 3, we include the interactions with climate unpredictability (other interactions were also robust to changes in the dependent variable across table

[23] Higher generosity and honesty could also be related to lower opportunism. Transaction costs economics would again predict that using POs would provide less benefits, the opposite of what we find.

6; these estimations are available upon request). The results show that total population increases with the presence of technologies, but only when these are executed by POs. The positive impact of POs on total population increases when the climate unpredictability is low.

"Population density" is the variable v1130 of the SCCS and is a categorical variable, with 1 equal to "less than 1 person per square mile", 2 equal to "1 - 4.9 persons per square mile", 3 equal to "5 - 24.9 persons per square mile", 4 equal to "25 - 99.9 persons per square mile", 5 equal to "100 - 499.9 persons per square mile", and 6 equal to "500 or more persons per square mile". The results presented in columns 4, 5 and 6 of table 6 show that our findings are also robust to the use of population density as the dependent variable.

Finally, "cultural complexity" is the variable v158.1 of the SCCS, where they sum the scores of 10 variables that capture the degree of cultural sophistication: writing and records, fixity of residence, agriculture, urbanization, land transport, money, density of population, political integration, social stratification, and specialization in metal working, weaving and pottery. In our case, we subtracted from the variable the last component of specialization. Overall, the results reported in columns 7, 8 and 9 of table 6 show that the results are robust to this alternative dependent variable.

Table 6. Robustness to other dependent variables

Dependent variable:	Total population			Population density			Cultural complexity		
	1	2	3	4	5	6	7	8	9
% presence	0.773	-0.564	-	0.789	-0.116	-6.213*	4.660	0.313	-8.208
	(2.253)	(1.997)	21.167***	(0.859)	(0.877)	(3.694)	(4.794)	(4.683)	(13.050)
			(5.947)						
% pres. x % within PO		5.818***	24.996#		3.584***	8.537#		18.910***	64.518***
		(2.241)	(14.409)		(1.101)	(5.423)		(4.897)	(18.426)
Climate unpredictability			-12.314**			-1.256			-2.943
			(5.659)			(4.509)			(14.188)
Clim. unp. x % presence			29.596***			8.808			11.606
			(8.307)			(5.519)			(20.434)
Clim. unp. x % pres.x % within PO			-27.183#			-7.028#			-
			(19.840)			(6.976)			61.207***
									(23.719)
All controls of table 1?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	153	153	153	128	128	127	128	128	127
Pseudo R-Square (R-Square)	0.801	0.819	0.835	(0.341)	(0.373)	(0.388)	0.807	0.828	0.835

Notes: Robust standard errors are used in all regressions and displayed in parentheses. *** indicates p-value<0.01 and # indicates p-value<0.01 for joint two tailed F-test of $\partial size/\partial \% withinPO$ in columns 3, 6 and 9. High multicollinearity – frequent in interaction models – requires a joint test. Columns 4, 5 and 6 use Ordered Probit, the rest OLS. For the columns 4 through 9, we use societies with information about at least 8 technologies in order to accommodate for the smaller sample size in these dependent variables (in table 6 below, we show that the results are robust to data stringency).

4.5.2 Alternative explanations

There are three main alternative explanations for our empirical results. We address each one in turn. First, it could be argued that the positive impact of POs is due to POs generating improvements in terms of the costs of individual learning, rather than the costs of social learning. To assess this possibility, we studied a model in which a PO decreases C instead of c. A model with this characteristic generates POs populated entirely by individual learners, and importantly, their benefits are increasing in the uncertainty parameter p. The latter implication is directly contradicted by our interaction with uncertainty: we find that uncertainty reduces the impact of POs on population (column 1 of table 5 and figure 5A). Of course, this mechanism could also be present, but the results show that the opposing mechanism, that is, POs decrease social learning costs, is stronger. The former implication – POs are populated by individual learners – is rebutted by simple perusal of organizational reality: in general, social learners dominate individual learners inside POs (and the opposite occurs in the market). Guilds are a good example.

A second alternative explanation to our empirical findings is related to trade. Specialized POs might have a positive impact on population because they are a marker for the presence of trade in the society, and trade could be the fundamental driver of larger populations and the key force behind the evolution of specialized POs. To test this alternative explanation, we use as controls several variables from the SCCS that proxy for the presence of trade in the society. Specifically, we use four variables: i) "intercommunity trade" (v1 of the SCCS) is a categorical variable that measures the extent to which intercommunity trade is a source of food (from "no trade" to "food imports present and contribute more than 50%"); ii) "presence of money" (v17) is a categorical variable with five categories (from "no media of exchange or money" to "indigenous coinage or paper currency"); iii) "presence of credit" (v18) is a categorical variable with four categories (from "Personal loans between friends or relatives" to "banks or comparable institutions"); iv) and "importance of trade" (v819b) is a continuous variable measures the percentage that trade contributes to subsistence in the society (i.e., food provision)[24].

In column 2 of table 7, we add "Intercommunity trade" as a control; in column 1, we use the same societies used in column 2. This allows us to cleanly assess the impact of the control variable on the impact of PO. The same is done in columns 4 and 3 for "money" and "credit" and in columns 6 and 5 for "importance of trade". Across these three comparisons, the coefficients decrease in size by an average of 16% but remain

[24]This variable is computed by the SCCS from using v1 and other five variables that provide categorical information on the extent that agriculture, fishing, gathering, animal husbandry and hunting contribute for subsistence (the mean is 8% with a maximum of 65%, a median of 5% and a 90th percentile equal to 25%).

statistically significant. The largest decrease is observed for the model that includes "money" and "credit". These reductions in the size of the coefficients indicate that some of the impact of POs is indeed generated through trade benefits, but that it is not the main mechanism. Instead, this result is consistent with our proposition 3, which states the origin of specialized POs is driven by the need to make social learning useful in society, without requiring trade as a force for its evolution. Thus, even in societies without trade, specialization within POs would be beneficial

A secondary means to assess the alternative explanation of trade is by exploring the interaction with uncertainty. The literature on trade has proposed and documented that trade (and thus the specialization that it drives) is particularly useful to mitigate the effects of shocks to local productivity, such as weather changes (Burgess and Donaldson, 2010). Therefore, if the benefits of trade are the key driver of the impact of POs, we should find a positive interaction of POs with uncertainty; however, we find the opposite in our results (column 1 of table 5 and figure 6A).

Table 7. Robustness to trade

	Dependent variable: Local community size					
	1	2	3	4	5	6
% presence	1.079	0.762	0.751	1.892	0.980	0.925
	(1.235)	(1.267)	(1.243)	(1.421)	(1.227)	(1.259)
% presence x % within PO	4.237***	3.897***	3.978**	2.630*	4.261***	4.072***
	(1.530)	(1.524)	(1.551)	(1.559)	(1.532)	(1.503)
"Intercommunity trade" dummies?	No	Yes	No	No	No	No
"Money" and "Credit" dummies?	No	No	No	Yes	No	No
"Importance of trade" control?	No	No	No	No	No	Yes
Resource endowment controls?	No	No	No	No	No	No
All other controls?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	130	130	124	124	131	131
Pseudo R-Square	0.367	0.379	0.368	0.422	0.367	0.375

Notes: We execute ordered probit regressions. Robust standard errors are used in all regressions and are displayed in parentheses. *** indicates p-value<0.01, ** indicates p-value<0.05, * indicates p-value<0.1. To avoid data loss, in all regressions, we use societies with information about at least 8 technologies and drop resource endowments controls.

The third alternative explanation for the origin of POs is that they emerge as a result of political complexity. The idea is that having a complex political organization in the society allows the society to better define, monitor and enforce POs. Thus, it might be that political complexity is really driving both the presence of POs and a larger population. In table 8, we show that our results are robust to adding “political hierarchy" as a control. This variable is a categorical variable from the Ethnographic Atlas that indicates whether the society has political authority and if it does, the reach of this authority (local chiefdoms, large chiefdoms, small states, large states). In these models, we replicate column 2 of table 1 and columns 2, 5 and 8 of table 6, but with the addition of the control of political hierarchy. Comparing the estimated coefficients with those of table 1 and table 6, the results shows that the coefficients are reduced by 23% on average (across dependent variables). This results indicates that this alternative explanation carries some weight, but not sufficient to overthrow our results. Of course, it could also be argued that political complexity is driven by POs in the first place. If that is the case, including this control would be biasing downward the true impact of POs.

Table 8. Robustness to political hierarchy

	Dependent variable			
	1 Local community size	2 Population size	3 Population density	4 Cultural complexity
% presence	-0.190 (0.861)	-0.253 (2.053)	-0.669 (0.939)	-0.020 (3.350)
% presence x % within PO	3.790*** (1.428)	4.851** (2.490)	3.159*** (1.202)	8.964** (4.360)
Political hierarchy dummies?	Yes	Yes	Yes	Yes
Region dummies?	Yes	Yes	No	No
All other controls from table 1?	Yes	Yes	Yes	Yes
Observations	171	151	125	125
Pseudo R-Square (Pseudo R-Square)	(0.358)	0.830	(0.409)	0.888

Notes: For “Population density” and “Cultural complexity”, we do not use region dummies, and we use societies with information in at least 8 technologies in order to accommodate for the smaller sample sizes for these dependent variables. Robust standard errors are used in all regressions and displayed in parentheses. *** indicates p-value<0.01. Columns 1 and 3 use OLS; columns 2 and 4 use Ordered Probit.

4.5.3 Data stringency

The results are also robust to being less restrictive regarding the available information about the activities (table 9). In many societies, there is information about only a portion of the activities. In columns 1 through 4, we change the minimum number of activities that have available information in the society, and the results do not change. In addition, in columns 5 and 6, we restrict the sample to regions that have at least 2 and 3 societies in them, leading to a loss of 7 and 19 societies, respectively. The results are robust to changing both of these information criteria.

Table 9. Robustness to available information on activities and regions

	Dependent variable: Size of local population (Ordered probit)					
	Number of activities				Societies per region	
	1	2	3	4	5	6
sample:	at least 7	at least 8	at least 9	at least 10	at least 2	at least 3
% presence	0.998 **	0.956*	0.912	-0.047	0.013	-0.246
	(0.472)	(0.525)	(0.590)	(0.647)	(0.828)	(0.832)
% presence x % within PO	3.230 ***	3.346 ***	3.201***	3.412***	4.185***	3.513***
	(0.715)	(0.759)	(0.837)	(0.978)	(1.243)	(1.261)
All controls of table 1 included?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	430	393	330	269	166	154
Pseudo R-Square	0.287	0.286	0.295	0.306	0.342	0.3

Notes: We execute ordered probit regressions. Robust standard errors are used in all regressions and are displayed in parentheses. *** indicates p-value<0.01, ** indicates p-value<0.05, * indicates p-value<0.1.

4.6 Testing the impact on the contemporary importance of tradition

The final empirical test we execute is drawn from Giuliano and Nunn (2017). These authors test an important implication of the baseline model: the use of social learning decreases with environmental uncertainty (see footnote 7). When the environment changes frequently, people increase the use of individual learning and decrease their reliance on inherited tradition (i.e., social learning). To test this idea, Giuliano and Nunn generate a measure of the environmental instability that the ancestors of a country's population were subject to. First, they create a mapping that breaks down a country's population according to their ancestry in different societies present in the ethnographic atlas. Second, they use the intergenerational temperature variability that was in place between 500 and 1900 in the regions of the country's ancestors to generate a measure of climate instability at the country level. Third, they analyze the impact of this measure on several measures that capture the use of tradition. Using several empirical approaches across countries, individuals, and descendants of immigrants, they find extensive support for an increase in the reliance on tradition if the ancestors lived in a stable environment.

We use the same idea but instead of relying on weather variability, we rely on the extent of use of PO in the premodern societies of the Ethnographic Atlas. Our model predicts that the introduction of POs leads to higher use of social learning in the population, particularly when their advantage in social learning is high[25]. Thus, countries with ancestors that used POs intensively would exhibit a higher use of tradition today.

In table 10, we replicate tables 3, 4 and 5 of Giuliano and Nunn's paper. In these tables, they analyze the impact of temperature variability on the persistence of cultural traits. Instead of using temperature variability, we use the presence of POs. In the columns 1, 2 and 3, we study how the presence of POs shifts the degree of persistence in female labor participation (FLP). The data on FLP are drawn from the World Bank Development Indicators. From column 1, we observe that the persistence is 0.32 between 1970 and 2012 (1 being the maximum and 0 the minimum). From column 2, we observe that this persistence varies systematically with the presence of POs. We find that POs generate a significant increase in the persistence of female labor participation (positive coefficient on the term "female labor participation in 1970 x % presence x % within PO"). The size of this correlation is not small: setting "% presence" at its mean, a one standard deviation increase in "% within PO" increases the coefficient of persistence by 0.17 (e.g., from 0.3 to 0.47), which equals half of the baseline coefficient in column 1 (we analyze the expression $\partial FLP12/\partial FLP70\partial \% withinPO \times St.Dev.\% withinPO$). In column 3, we control for all the interactions between FLP in 1970 and the other variables in the model. Although this addresses the impact that ancestral POs have on tradition persistence through their impact on development, it also generates high multicollinearity, which weakens the statistical significance of this model.[26]

In columns 4, 5, and 6, we repeat the analysis, but we now study the persistence of

[25] The exception to this prediction is in the case of no secrecy when the size of the PO is less than the point of maximum fitness (see figure 1A). In this case, an expansion of the PO reduces the share of social learners outside, leaving the total share of social learners unchanged at the population level. Thus, our prediction is true on average.

[26] In the models presented in table 10 we do not explore the heterogeneity of impact with respect to "sex differentiation". Sex differentiation in productive activities can impact female labor participation (or polygamy) in many different manners other than the channel that we care about, social learning within POs. Thus, the estimations would not be reliable. That said, we do find, particularly for polygamy, that sex differentiation boosts the positive impact of POs on reliance on tradition.

FLP in ancestors on FLP in 2012. Following Giuliano and Nunn, we use v54 of the EA, which measures the presence of females in preindustrial agriculture. We normalize this variable to make it between 0 and 1. Column 4 shows that the persistence is 0.164 and statistically significant. Compared to column 1, a lower persistence is expected due to the longer time span. The results displayed in column 5 and column 6 indicate that the presence of POs has a positive impact on the persistence of FLP from ancestry to modernity. However, we only obtain statistical significance in column 6. This could be due to fact that we explore the persistence from ancestry to modernity, and most of the interactions added to column 6 control for the impact of other EA variables on persistence. The effect size is large: setting "% presence" at its mean, a one standard deviation increase in "% within PO" increases the coefficient of persistence by 0.19 (e.g., from 0.1 to 0.29), which is slightly greater that the baseline coefficient in column 4. The significance is only at 10% because the multicollinearity is high in this model.

In columns 7, 8 and 9, we analyze the persistence of polygamy among ancestors on polygamy in 2009. Polygamy in 2009 is a dummy variable drawn from the OECD Gender, Institutions and Development Database, and we follow Giuliano and Nunn for it operationalization: it takes the value of 1 if having more than one spouse is accepted or legal. For polygamy among ancestors, we follow Giuliano and Nunn and use the variable v9 of the EA. We build a dummy that takes the value of 1 if there is the presence of polygamy among the ancestors and 0 if the society is monogamous. In column 1, we find a statistically significant persistence coefficient of 0.33. In columns 8 and 9, we find a positive impact of the presence of POs on the persistence of polygamy. However, due to high multicollinearity, significance is not present, but the effect size is large: using column 9 and setting "% presence" at its mean, a one standard deviation increase in "% within PO" increases the coefficient of persistence by 0.14 (e.g., from 0.2 to 0.34), which is approximately half of the baseline coefficient reported in column 7.

Although these tests are subject to confounders, the correlations that we document are consistent with the prediction of our model. Two features of this exercise provide additional confidence in our account of POs. First, the dependent variable are not drawn from the EA, making this an "out-of-sample" exercise. Second, we test a different prediction of our model that is not about fitness but rather about changes in the share of social learning (and therefore about the reliance on inherited culture).

Table 10. Impact of POs on the persistence of cultural traits

Dependent variable:	Female Labor Participation in 2012			Female Labor Participation in 2012			Poligamy in 2009		
Persistence variable “P":	Female labor participation in 1970			Female labor participation in ancestors			Poligamy in ancestors		
	1	2	3	4	5	6	7	8	9
P	0.324***	-1.338*	-1.673	0.164**	0.158	0.062	0.337**	0.497	2.815
	(0.123)	(0.708)	(1.062)	(0.068)	(0.302)	(3.921)	(0.146)	(0.899)	(4.537)
P x % presence		1.573**	1.624*		0.024	-0.290		-0.333	-0.434
		(0.756)	(0.895)		(0.369)	(0.467)		(1.024)	(0.969)
P x % pres. x % within PO		0.956**	0.662		0.174	1.032*		0.626	0.752
		(0.435)	(0.683)		(0.396)	(0.579)		(0.985)	(1.168)
% presence, % within PO and their required interactions?	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Controls of table 10?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
% presence x Sex differentiation?	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Interactions between P and controls of table 10 and “% presence x Sex diff."?	No	No	Yes	No	No	Yes	No	No	Yes
Region dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of ethnography?	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	74	74	74	149	149	149	96	96	96
R-Square	0.545	0.695	0.749	0.451	0.506	0.543	0.634	0.675	0.722

Notes: Robust standard errors are used in all regressions and displayed in parentheses. *** indicates p-value<0.01, ** indicates p-value<0.05, * indicates p-value<0.1. In models 2, 3, 5, 6, 7, and 8, there is high collinearity (this is frequent in models with interaction terms). Following Nunn and Giuliano (2017), in models 1 through 6, we added the square of ln(GDP per capita); in models 4 to 9, we added the year of the ethnography; and we excluded ethnographies that occurred before 1800 (many ethnographies are dated BC – these are drawn from historical studies). For FLP in 1970, we used the average of a five-year window around the year 1970 (the data are sparser in 1970 than in 2012).

5 Conclusion

In this article, we have developed a theory that explains the evolution of productive organizations (POs). We used a cultural evolution model to show that improvements in social learning within a PO can favor the hard-to-propel process of cumulative culture. Under this account, POs evolved because they facilitate the transmission of knowledge between individuals, particularly if the PO specializes in a specific activity or technology. If access to POs is restricted, as is typical, then this knowledge transmission advantage leads to higher fitness of societies and therefore to the selection and invasion of POs. We provide evidence from a sample of premodern societies that is largely consistent with the predictions of our model. The theory applies straightforwardly to premodern POs, such as guilds, and long-standing POs, such as partnerships; as a descendant of these older POs, our theory also yields insight into the origin of modern firms.

Our findings provide several main contributions. First, we are the first to show that social learning can be beneficial to society even if it does not generate a positive externality on individual learning (cf., Boyd and Richerson, 1995). Second, we provide an explanation for the origins of POs based on social learning, knowledge transmission and cultural accumulation. Mainstream theories of POs focus on governance and incentives, assuming a predetermined pool of knowledge and culture. Knowledge issues have been addressed, but mostly in relation to knowledge integration and problem solving (Garicano, 2000; Grant, 1996). Third, we provide a theory for the origins of specialization within POs that does not rely on trade and comparative advantage as the driving force; specialization evolves because it favors the social learning benefits of POs. Fourth, as our comparative statics and empirical results attest, our findings on the role of uncertainty, trust, and secrecy run counter to the conventional wisdom. This outcome suggests that an evolutionary lens changes the predictions that one would derive when the problem of knowledge and culture accumulation is not considered.

As with any trait that has been selected in a population, a full explanation of the nature of POs requires adding an evolutionary perspective to the mix. We need to consider not only the mechanisms that explain the inner workings and immediate benefits of POs, such as governance and protection from hazards, but also the evolutionary motive for why they might have increased in frequency in the first place. As a first step in this direction, we hope that this paper stimulates further research regarding this important evolutionary foundation.

We can point to several limitations in our paper, all suitable targets for future research efforts. First, although we test our model on premodern societies using a good proxy for POs, it would be interesting to test the model using data regarding guilds or early partnerships. There are interesting new datasets that could be used for this purpose (e.g., Comino et al., 2017). Second, the link of our theory with modern firms is derivative, mainly as descendant of early POs. Knowledge and technology in our model are transmitted across individuals, that is, we address accumulation of individual-level traits. However, modern firms combine specialized knowledge to generate complex technologies that are beyond the capacity of any single individual to produce or imitate. It would be very interesting to study how our model can be extended to study the evolutionary origin of modern firms. Third, the behavior in our model is simplified to copying by social learners and "radical innovation" by individual learners. The model can be enhanced by introducing incremental innovation: agents could improve the technology while the state of nature remains unchanged. We suspect that the introduction of this element will further expand the beneficial impact that POs have on society.

References

Alesina, A., Giuliano, P., 2015. "Culture and Institutions" *Journal of Economic Literature* 53 (4): 898-944.

Aldrich, H. 1999. *Organizations Evolving*. Sage Publications, London.

Argote, L., Miron-Spektor, E. 2011 "Organizational Learning: From Experience to Knowledge." *Organization Science* 22 (5): 1123-1137

Arora, A., Cohen, W., Cunningham, C. 2018 "Inventive Capabilities in the Division of Innovative Labor" *NBER working paper* No. 25051.

Atalay, E., Hortaçsu, A., Syverson, C. 2014. "Vertical Integration and Input Flows". *The American Economic Review* 104(4): 1120-1148.

Anderson, R., 1971 "Voluntary Associations in History". *American Anthropologist*, 73: 209-222

Becker, A., Enke, B., Falk, A. 2018 "Ancient Origins of Global Variation in Social Preferences" NBER Working Paper No. 24291.

Benkler, Y., 2006. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Bisin, A., Verdier, T., 2011. "The Economics of Cultural Transmission and Socialization" In *Handbook of social economics* (Vol. 1, pp. 339-416). North-Holland.

Blau, P., Scott, R. 1963. *Formal Organizations: A Comparative Approach*. Routledge and Kegan Paul, London.

Boyd, R. 2017. *A Different Kind of Animal: How Culture Transformed Our Species*. Princeton University Press.

Boyd, R., Richerson, P.J. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R., Richerson, P. J. 1995. "Why Does Culture Increase Human Adaptability?" *Ethology and Sociobiology* 16(2), 125-143.

Boyd, R., Richerson, P.J. 2005. *The Origin and Evolution of Cultures*. Oxford: Oxford University Press.

Boyd, R., Richerson, P. J., Henrich, J. 2013. "The Cultural Evolution of Technology: Facts and Theories." In *Cultural Evolution: Society, Technology, Language, and Religion*, ed. Peter J. Richerson and Morten H. Christiansen.

Bowles, S., 2009. "Did Warfare Among Ancestral Hunter-Gatherers Affect the Evolution of Human Social Behaviors?" *Science* 324 (5932): 1293-1298.

Burgess, R., Donaldson, D. 2010. "Can Openness Mitigate the Effects of Weather Shocks? Evidence from India's Famine Era 449-53." *American Economic Review* 100 (2): 449-453

Chandler, A. D. 1990. *Scale and Scope: The Dynamics of Industrial Competition*. Cambridge, MA, Harvard Business School.

Coase, R. 1937 "The Nature of the Firm." *Economica* 4 (16): 386-405.

Colwell, R. 1974. "Predictability, Constancy, and Contingency of Periodic Phenomena" *Ecology* 55 (5): 1148-1153.

Costin, C. L. 2001. "Craft Production Systems". In *Archaeology at the millennium* (pp. 273-327). Springer, Boston, MA.

Comino, S., Galasso, A., Graziano, C. 2017. "The Diffusion of New Institutions: Evidence from Renaissance Venice's Patent System" NBER Working Paper No. 24118

Coto-Sarmiento M., Rubio-Campillo, X., Remesal, J. 2018. Identifying social learning between Roman amphorae workshops through morphometric similarity, Journal of Archeological Science, 96: 117-123

Crook, R., Combs, J., Ketchen, D., Aguinis, H. 2013. "Organizing Around Transaction Costs: What Have we Learned and Where do we go From Here?" *Academy of Management Perspectives* 27 (1): 63–79

Dari-Mattiacci, G., Gelderblom, O., Jonker, J., Perotti, E. 2017. "The Emergence of the Corporate Form" *Journal of Law, Economics and Organization* 33 (2): 193-236.

De la Croix, D., Doepke, M., Mokyr., J. 2017. "Clans, Guilds, and Markets: Apprenticeship Institutions and Growth in the Pre-Industrial Economy." *The Quarterly Journal of Economics* Forthcoming.

Dean, L.G., Kendal, R.L., Schapiro, S.J., Thierry, B., Laland, K. N. 2012. "Identification of the Social and Cognitive Processes Underlying Human Cumulative Culture." *Science* 335 (6072): 1114-1118.

Demsetz, H. 1988 "The Theory of the Firm Revisited" *Journal of Law, Economics and Organization* 4 (1): 141-162.

Depetris-Chauvin, E., Ozak, O, 2017. "The Origins and Long-Run Consequences of the Division of Labor" Working paper.

Enke, B. 2007. "Kinship Systems, Cooperation and the Evolution of Culture." Working paper.

Espinosa, M. 2017. "Sourcing of Expertise and the Boundaries of the Firm: The Case of Lobbyists" Working paper, LSE.

Fairlie, R W., Florian Hoffmann, and Philip Oreopoulos. "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom." American Economic Review, vol. 104, August 2008, pp. 2567-2591

Fogarty, L., Strimling, P., Laland, K. N. 2011. "The Evolution of Teaching." *Evolution* 65 (10): 2760-2770.

Forbes, S., Lederman, M. 2009. "Adaptation and Vertical Integration in the Airline Industry" *American Economic Review* 99 (5): 1831-1849.

Garicano, L. 2000 "Hierarchies and the Organization of Knowledge in Production" *Journal of Political Economy* 108 (5): 874-904.

Geyskens, I., Steenkamp, J., Kumar, N., 2006. "Make, Buy, or Ally: A Transaction Cost Theory Meta-Analysis". *The Academy of Management Journal* 49 (3): 519-543.

Gibbons, R., Henderson, R. 2012. "Relational Contracts and Organizational Capabilities." *Organization Science* 23 (5): 1350-1364.

Giuliano, P. 2016 "Review of Cultural Evolution: Society, Technology, Language, and Religion edited by Peter J. Richerson and Morten H. Christiansen" *Journal of Economic Literature* 54 (2): 522-533.

Giuliano P, Nunn N. 2018 "Understanding Cultural Persistence and Change." Working Paper

Grant, R. 1996. "Toward a Knowledge-Based Theory of the Firm." *Strategic Management Journal*, 17 (S2): 109-122.

Guinnane, T. W. , Harris, R., Lamoreaux, N., Rosenthal, J. 2007 "Putting the Corporation in its Place" *Enterprise and Society* 8: 687-729.

Hannan, M.T., Freeman, J. 1977. "The Population Ecology of Organizations." *American Journal of Sociology* 82(5): 929-964.

Hansmann, H., Kraakman, R., Squire, R. 2006. "Law and the Rise of the Firm." *Harvard Law Review*, 119: 1335-1403.

Hart, O., Moore, J. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98 (6): 1119-1158.

Haun, D., Over, H. 2013 "Like Me" In *Cultural Evolution: Society, Technology, Language, and Religion*, ed. Peter J. Richerson and Morten H. Christiansen.

Henrich, J. 2015. *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating our Species, and Making us Smarter*. Princeton University Press.

Hewlett, B., Roulette, C. 2016. "Teaching in Hunter-Gatherer Infancy". *Royal Society open science* 3 (1): 150403

Holmstrom, B., Milgrom, P. 1994. "The Firm as an Incentive System." *The American Economic Review* 84 (4): 972-991.

Josefy, M., Kuban, S., Ireland, R. D., Hitt, M. A. 2015. "All Things Great and Small: Organizational Size, Boundaries of the Firm, and a Changing Environment." *Academy of Management Annals* 9 (1): 715-802.

Kirby, K., Gray, R., Greenhill, S., Jordan, F., Gomes-Ng, S., Bibiko, H., Blasi, D. et al. 2016. "D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity." *PloS one* 11(7): e0158391.

Klepper, S. 1996. "Entry, Exit, Growth, and Innovation over the Product Life Cycle." *The American Economic Review*, 86(3): 562-583

Kogut, B., Zander, U. 1992. "Knowledge of the Firm, Combinative Capabilities and the Replication of Technology." *Organization Science* 3 (3): 383-397

Lafontaine, F., Slade, M. 2007. "Vertical Integration and Firm Boundaries: The Evidence". *Journal of Economic Literature* 45 (3): 629-685.

Laland, K. 2017. *Darwin's Unfinished Symphony: How Culture Made the Human Mind*. Princeton University Press

Lamoreaux, N.R., 1998. "Partnerships, Corporations, and the Theory of the Firm" *The American Economic Review* 88 (2): 66-71.

Levin, J and Tadelis,S.,2005 "Profit sharing and the role of professional Partnerships" The Quarterly Journal of Economics, Volume 120, Issue 1: 131–171,

Levinthal, D. 1997. "Adaptation on Rugged Landscapes." *Management Science* 43 (7): 934-950.

Levitt, S.D., List, J.A.,Syverson, C., 2013. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant" *Journal of Political Economy*, 121 (4): 643-681.

Lewis, H. M., Laland, K. N. 2012. "Transmission Fidelity is the Key to the Build-up of Cumulative Culture." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367 (1599): 2171-2180.

Losin, E.R, Iacoboni, M., Martin, A., Dapretto, M. (2012) Own-gender imitation activates the brain's reward circuitry. Social Cognitive and Affective Neuroscience. 7: 804-810

Lowie, R. 1948. *Social Organization*. New York Rinehart and Co.

Mokyr, J., 2016. *A Culture of Growth: the Origins of the Modern Economy*. Princeton University Press.

Morrison AD, Wilhelm WJ. "The Demise of Investment Banking Partnerships: Theory and Evidence." *The Journal of Finance*. 63 (1):311-50.

Murdock, G. P. 1967 *Ethnographic Atlas*, University of Pittsburgh Press.

Murmann, J., Aldrich, H., Levinthal, D., Winter, D. 2003 "Evolutionary Thought in Management and Organization Theory at the Beginning of the New Millenium" *Journal of Management Inquiry* 12 (1): 22-40.

Nelson, R., Winter, S. 1982. *An Evolutionary Theory of Economic Change.* Belknap Press of Harvard University Press, Cambridge, Massachusetts

Nickerson, J., Zenger, T. 2004. "A Knowledge-Based Theory of the Firm – The Problem Solving Perspective." *Organization Science* 15 (6): 617-632.

North, D.C., 1991. "Institutions" *Journal of Economic Perspectives* 5 (1): 97-112.

Nowak, M. A. 2006. "Five Rules for the Evolution of Cooperation." *Science* 314 (5805): 1560-1563.

Nunn, N. 2009. "The Importance of History for Economic Development" *Annual Review of Economics* 1: 65-92.

Nunn, N., 2012. "Culture and the Historical Process" *Economic History of Developing Regions*, 27 (1): 108-126.

Ogilvie, S. 2014. "The Economics of Guilds." *The Journal of Economic Perspectives*, 28 (4): 169-192.

Oster, E., 2016. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* DOI: 10.1080/07350015.2016.1227711

Passmore, S, Jordan, F. 2017. "Shared Ancestry Shows Kinship is Conserved." Presented at the 1st Cultural Evolution Society Conference.

Padgett, J. F., McLean, P. D. 2006. "Organizational Invention and Elite Transformation: The Birth of Partnership Systems in Renaissance Florence." *American journal of Sociology* 111(5): 1463-1568

Poblete, J,. 2015 "Profit sharing and Market Structure" *International Journal of Industrial Organization*,39: 10-18

Ramachandran, S.; Deshpande, O.; Roseman, C. C.; Rosenberg, N. A.; Feldman, M. W.; Cavalli-Sforza, L. L. (2005). "Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa" Proceedings of the National Academy of Sciences 10 (44): 15942-15947

Rand, D. G., Nowak, M. A. 2013. "Human Cooperation." *Trends in Cognitive Sciences* 17 (8): 413-425

Rendell, L., Fogarty, L., Hoppitt, W., Morgan, T., Webster, M., Laland, K. 2011 "Cognitive Culture: Theoretical and Empirical Insights Into Social Learning STrategies." *Trends in Cognitive Sciences* 15(2): 68-76.

Rogers, A. R. 1988. "Does Biology Constrain Culture?" *American Anthropologist* 90 (4): 819-831.

Rowlands, M. J. 1971. "The Archaeological Interpretation of Prehistoric Metalworking" *World Archaeology* 3 (2): 210-224.

Scott, R. 2003. *Organizations: Rational, Natural and Open Systems.* Fifth Edition, Prentice Hall.

Smith, D.,Stebbins, R., Grotz, J (2016) *The Palgrave Handbook of Volunteering, Civic Participation, and Nonprofit Associations*, Pallgrave McMillan.

Spolaore, E., Wacziarg, R. 2013. "How Deep are the Roots of Economic Development?" *Journal of Economic Literature*, 51 (2): 325-69.

Spulber, D., 2009 "The Theory of the Firm" Cambridge University Press.

Sterelny, K. 2012. *The Evolved Apprentice: How Evolution Made Humans Unique.* MIT Press.

Stock, J., Yogo, M. 2002. "Testing for Weak Instruments in Linear IV Regression" NBER Working Paper No. 284

Stout, D. 2002 "Skill and Cognition in Stone Tool Production" *Current Anthropology*, 43 (5): 693-720.

Tabellini, G., 2008 "Institutions and Culture" *Journal of the European Economic Association* 6 (2-3): 255 - 294.

Tadelis, S., Williamson, O. 2013. "Transaction cost economics" Chapter 4 of the *Handbook of Organizational Economics* (eds. Robert Gibbons and John Roberts).

Turchin, P. 2016. *Ultra-Society: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth.* Beresta Books, LLC.

Williamson, O. E. 1991. "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives." *Administrative Science Quarterly* 36 (2): 269-296.

Winter, S., Szulanski, G. 2001. "Replication as Strategy." *Organization Science* 12 (6): 730-743.

Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Appendix 1: Proofs

Proof that there exists a unique ESS (r_I, r_s) **and in that equilibrium,** $f_s = 1 - C$. *The proof has two parts: we first show that there exists a unique Nash equilibrium; moreover, in this equilibrium,* $f_s = 1 - C$. *In the second part, we show that this equilibrium constitutes an ESS. There cannot exist other ESSs because the ESS is a refinement of the Nash equilibrium.*

Lemma 1: *In any Nash equilibrium, both strategies are played.*

Proof. Suppose not and remember that $f_I = 1 - C$. If $r_I = 0$, then $f_S = -c < f_I = 1 - C$.So r_I must be positive in any equilibrium. If $r_I = 1$, then $f_S = (1 - p) - c$. In order for f_S to be greater than $f_I = (1 - C)$, it is necessary that $(C - c) > p$. Since this condition will always be met according to assumption 1, then r_I can never be one because deviating to r_S would be profitable. ■

Lemma 2: *There is a unique Nash equilibrium* (r_I^*, r_S^*).

Proof. By lemma 1, in equilibrium, we must have $f_I = f_S$, but f_I is constant in r_I, and f_S is increasing in r_I. Therefore, there exists a unique r_I compatible with $f_I = f_S$, and thus, the equilibrium is unique. ■

Lemma 3: *The strategies played in the Nash equilibrium* (r_I^*, r_S^*) *are ESSs.*

Proof. Suppose that a population of size $\delta < 1$ with a share of individual learners $r_I \neq r_I^*$ invades. 1) If r_I is greater than r_I^*, then the resulting share of individual learners would be $\widetilde{r}_I = \frac{1}{1+\delta}(r_I^* + \delta r_I) > r_I^*$. This result implies that in the new equilibrium the fitness of social learners is $\widetilde{f}_s > 1 - C$, because the fitness of social learners is increasing in the share of individual learners. The average fitness of the invading population would be $1 - C + (1 - r_I)(f_S - (1 - C))$, while the average fitness of the existing population would be $1 - C + (1 - r_I^*)(f_S - (1 - C))$, which is strictly greater because $r_I > r_I^*$. 2) If r_I^* is greater than r_I, then the resulting share of individual learners would be $\widetilde{r}_I = \frac{1}{1+\delta}(r_I^* + \delta r_I) < r_I^*$. This result implies that in the new equilibrium, the fitness of social learners is $\widetilde{f}_s < 1 - C$ because the fitness of social learners is increasing in the share of individual learners. The average fitness of the invading population would be $1 - C - (1 - r_I)((1 - C) - \widetilde{f}_s)$, while the average fitness of the existing population would be $1 - C - (1 - r_I^*)((1 - C) - \widetilde{f})$, which is strictly less because $r_I < r_I^*$. ■

Because every set of ESSs is Nash, we know there exists a unique set of ESSs for this game. QED

Proof of Proposition 1: T*he structure of the proof is identical to the previous one, including a PO of size* λ, *for a sufficiently small* λ.

Lemma 4 *In any Nash equilibrium, both strategies are played outside the PO.*

Proof. Suppose not and remember that $f_I = 1 - C$. If $r_I = 0$, then $f_S = \lambda\widetilde{q} - c < f_I = 1 - C$ if λ is sufficiently small. So, r_I must be positive in any equilibrium. If $r_I = 1$, then $f_S = \lambda(\widetilde{q} - c) + (1-\lambda)(1-c)(1-p)$. In order for f_S to be greater than $f_I = (1-C)$ for every $\lambda > 0$, it is necessary that $(C - c) > p - pc$. Since this condition will always be met according to assumption 1, r_I can never be one because deviating to r_S would be profitable. ■

Lemma 2: *There is a unique Nash equilibrium* (r_I^*, r_S^*) outside the *PO*.

Proof. By lemma 1, in equilibrium, we must have $f_I = f_S$, but f_I is constant in r_I and f_S is increasing in r_I; therefore, there exists a unique r_I compatible with $f_I = f_S$, and thus, the equilibrium is unique. ■

Lemma 3: Inside the PO, all agents play "social learning".

Proof. The fitness of individual learning is $\widetilde{f}_I = 1 - C$, and the fitness of social learning inside the *PO* is given by $\widetilde{f}_s = f_s + \lambda(c - \widetilde{c}) = f_I + \lambda(c - \widetilde{c})$ (see equation 2). Because $\widetilde{f}_s > \widetilde{f}_I$, only social learning is played inside the firm. ■

Lemma 4: *The strategies played in the Nash equilibrium* $(r_I^*, r_S^*, \widetilde{r}_S = 1)$ *are ESSs.*

Proof. Suppose a population of size $\delta < 1$ invades. 1) If the invasion is outside the PO, the proof is analogous to that of lemma 3 of the previous proof and therefore omitted.If the invasion is inside the PO and $r_I \neq 0$, observe that average fitness of invading population is $r_I(1 - C) + (1 - r_I)f_S < f_S$, and thus, it has lower fitness than the population inside the PO. ■

Because every set of ESSs are Nash, we know there exists a unique set of ESSs for this game. QED

Proof of Proposition 2 *To prove this result, consider two societies with a productive organization of size λ, and assume that society a has secrecy and society b does not. Inside the firm (both with and without secrecy), all agents are social learners, and therefore, in either case, their fitness is given by*

$$\tilde{f}_s^i = [1 - \lambda][(1 - p)q^i - c] + \lambda[(1 - p)\tilde{q}^i - \tilde{c}]. \tag{7}$$

The recursive equations that determine the stock of knowledge inside the firm in the steady state is the same regardless of whether there is secrecy and is determined by

$$\tilde{q}^i(t) = [\lambda\tilde{q}^i(t - 1) + (1 - \lambda)q^i(t - 1)](1 - p). \tag{8}$$

From equations (7) and (8), it is clear that fitness in the firm is strictly increasing in the stock of knowledge outside the firm. Therefore, to show that secrecy is detrimental to the firm's fitness, we only need to show that secrecy is detrimental to the stock of knowledge outside the firm (i.e., that $q^b > q^a$).

To show this, it is useful to define the function $\tilde{q}^i(q^i)$ *, which is the stock of knowledge inside the firm, a function increasing in the stock of knowledge outside the firm, where* $\tilde{q}^i(q^i) < q^i$.

We know that

$$f_S^a = (1-p)q^a - c = 1 - C$$

$$f_S^b = (1-\lambda)(1-p)q^b + \lambda(1-p)\tilde{q}^b(q^b) - c = 1 - C$$

In contrast, suppose that $q(a) \geq q(b)$; *then,*

$$1 - C = (1-p)q^a - c > (1-\lambda)(1-p)q^b + \lambda(1-p)\tilde{q}^b(q^b) - c = 1 - C$$

is a contradiction. Therefore, it must be that $q(b) > q(a)$.QED

Proof of Proposition 3. *Observe from equation (6) that the fitness of a social learner that specializes in technology* j *can be expressed as*

$$\widetilde{f}_S^j = f_S^j + (c - \widetilde{c}) \cdot \frac{\lambda \widetilde{x}^j}{\lambda \widetilde{x}^j + (1-\lambda)x^j}. \tag{9}$$

Remember that in equilibrium, f_S^j *and* x^j *must be constant across* J *outside the productive organization. It is straightforward to check that there are two Nash equilibria inside the PO. 1) Either* $\widetilde{x}^j$ *is the same for every* j *in* J, *in which case* $\widetilde{f}_S^j$ *is constant in* j, *or 2) there is full specialization, and* $\widetilde{x}^j = 1$ *for* $j = j^*$ *and* 0 *otherwise.*

Observe now that the first equilibrium is not ESS because if a small population of size δ *that specializes in one activity* $\widetilde{j}$ *invades the PO, then the fitness of the invading population becomes*

$$\widetilde{f}_S^{\widetilde{j}} = f_S^j + (c - \widetilde{c}) \cdot \frac{\lambda \widetilde{x}^j + d}{\lambda \widetilde{x}^j + (1-\lambda)x^j + d}$$

The average fitness of the average population is

$$E_{j \in J}(\widetilde{f}_S^j) = f_S^j + (c - \widetilde{c}) \cdot \left[\frac{J-1}{J}\right] \left[\frac{\lambda \widetilde{x}^j}{\lambda \widetilde{x}^j + (1-\lambda)x^j + d}\right] + \left[\frac{1}{J}\right] \widetilde{f}_S^{\widetilde{j}} < \widetilde{f}_S^{\widetilde{j}}.$$

To see that a specialized Nash is ESS, notice that in a specialized equilibrium, the fitness inside the PO is given by

$$\widetilde{f}_S^{j^*} = f_S^{j^*} + (c - \widetilde{c}) \cdot \frac{\lambda}{\lambda + (1-\lambda)x^j},$$

Observe that an invasion of size δ *by any other technology will obtain a fitness of*

$$\widetilde{f}_S^{\widetilde{j}} = f_S^j + (c - \widetilde{c}) \cdot \frac{\delta}{\lambda + (1-\lambda)x^j + \delta},$$

which converges to $f_S^j < \widetilde{f}_S^{j^*}$ *as* δ *approaches 0, and therefore, the invading population obtains a lower fitness level.* QED.

Appendix 2: Descriptive statistics

Table A1. Descriptive statistics of continuous variables in table 1.

Variable	N	Mean	St. Dev.	Min	Max
Size of local population	173	3.38	2.33	1.00	8.00
% presence	173	0.67	0.18	0.27	1.00
% within PO	173	0.12	0.20	0.00	1.00
% presence x % within PO	173	0.10	0.16	0.00	0.64
Amphibian richness	173	15.64	15.87	1.00	102.12
Bird richness	173	202.72	109.90	2.75	528.85
Mamalian richness	173	73.12	41.77	2.75	196.00
Vascular plant richness	173	1817.97	879.41	2.75	4212.26
Distance to coast (km)	173	445.50	406.45	0.36	1697.74
Elevation	173	849.62	820.42	0.00	4345.04
Slope of terrain	173	2.73	2.90	0.00	14.13
Absolute latitude	173	31.71	18.95	0.00	71.00
Average temperature	173	13.18	9.81	-12.32	27.64
Year of ethnographic record	173	1880	212	-800	1965

Table A2. Presence and intensity of agriculture

Category	Number of societies
Complete absence	66
Casual	8
Shifting cultivation	38
Semi-intensive, horticulture-like	5
Intensive with techniques (e.g., fertilization, crop rotation)	38
Intensive, dependent upon irrigation	18
Total	173

Table A3. Type of Settlement

Category	Number of societies
Fully nomadic bands	19
Semi-nomadic (fixed winter settlement)	40
Semi-sedentary (shifting between fixed seasonal settlements)	27
Fixed settlements shifted every few years	3
Neighborhoods of dispersed family homesteads	18
Separate hamlets forming a community	6
Established Villages or towns	51
Complex settlements: Nuclear town with satellite homesteads/hamlets	9
Total	173

Table A4. Regions

Region	Number of societies
Southwestern U.S.A.	24
Northwestern U.S.A.	13
Eastern Europe	12
Indian Subcontinent	9
West Tropical Africa	9
Western Canada	9
Western South America	7
East Tropical Africa	6
Northern South America	6
Papuasia	6
Brazil	5
Eastern Canada	5
Malesia	5
Southern Africa	5
Subarctic America	5
Australia	4
Russian Far East	4
Southeastern Europe	4
West-Central Tropical Africa	4
Siberia	3
South Tropical Africa	3
Southern South America	3
Western Asia	3
13 Other regions	19
Total	173

Table A5. Year of ethnography

Year	Number of societies
800 BC	1
1520	1
1530	1
1750	2
1800 - 1819	2
1820 - 1839	1
1840 - 1859	18
1860 - 1879	28
1880 - 1899	23
1900 - 1919	24
1920 - 1939	31
1940 - 1959	36
1960 - 1965	5
Total	173