THE FRIENDS WE MAKE
Networks, Culture and Institutions

Prateek Raj

A dissertation submitted in the fulfillment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY
of
UNIVERSITY COLLEGE LONDON

UCL School of Management
University College London

July, 2018

I, Prateek Raj, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Who are our friends and what is the nature of networks we make, can have macro-level implications on society.

There are relationship-based economies, and there are economies that rely on arm's length interactions. This thesis studies how relationship-based groups like merchant guilds and caste can persist due to a lack of incentives and information, or because of salience of social identities. It also explores the factors that can widen the social circle and give rise to impersonal and cosmopolitan social systems.

In the second chapter, I develop a theoretical model to show how difficult it is for relationship-based economies to be honest to and trusting of strangers and to transition into impersonal economies. I find a narrow set of conditions under which generalized trust in strangers can emerge and sustain in an impersonal economy.

In the third chapter, I explore the transition of guild-based economies of Northwestern Europe into market-based economies in the sixteenth century. I study the unique interaction of economic and technological factors that drove such a change. I find that cities in the sixteenth century, where monopolies of merchant guilds declined, were at the Atlantic coast and had significantly higher levels of printing in the fifteenth century.

In the fourth chapter, with Paola Sapienza and Luigi Zingales, I examine the effect of caste censuses of 1901 in making the caste identity salient today. We study its impact on networking and governance and find that in regions where district committees were formed to rank castes in 1901, households have a smaller social circle as they have fewer out-of-caste networks with professionals. They also have a poorer quality of government.

Overall, the thesis develops a detailed picture of how social structures historically evolve, and how trust, access to information and social identity are the forces that shape them.

Impact Statement

As business researchers, we are interested in helping businesses succeed. Much of the business-related research[1], uses business organizations from developed countries as subjects, where strong formal institutions support businesses. But, business prescriptions that use management theories that were developed in the context of the developed world can be insufficient for the developing world. How much can the current business research inform us about business in developing countries, where formal institutions are either not developed or are corrupt?

The key challenge in doing business in developing countries is going beyond the strict confines of the networks one has. These networks are the reliable way of doing business in places without credible formal institutions. But, this networked nature of business leads to an inefficient and unequal distribution of resources, as it is difficult for newcomers and especially small businesses to enter a market without previous contacts. So, incumbents who have many connections can seek rents at the cost of excluding potentially superior newcomers. Also, these incumbents can bribe the rule enforcers, and inhibit the running of impartial formal institutions.

This thesis helps policymakers understand the economic and cultural conditions under which doing business beyond networks is especially more difficult. It also helps them identify the conditions that can favor the emergence of business beyond networks, where businesses can genuinely enjoy a wide circle of opportunity.

Broadly, the thesis provides a synthesis of diverse literature about the fundamentals of organizing business in traditional and developing societies with weak institutions. By using 1) a variety of historical, empirical and modeling techniques, and 2) theories from multiple fields in strategy, organization studies, economics, political economy and history, the thesis takes a holistic view of how business gets done in a variety of institutional, cultural and technological settings.

[1] Among the highly cited recent management articles some notable exceptions that have a setting outside of the United States or Europe include Mair, Martí & Ventresca (2012) that focused on inclusive markets in Bangladesh, Marquis & Qian (2013) that studied corporate social responsibility in China, and Bloom et al. (2013) that studied the impact of improving management practices on firms in India.

Acknowledgement

I would like to thank my advisor Simcha Jong for his guidance, support, and mentorship since 2012. I would also like to thank the students and staff members of *UCL School of Management*, especially the Ph.D. Program Director Bilal Gokpinar, for their academic and administrative support. I am also grateful to Elias Papaioannou at London Business School (LBS), for his guidance during the early years of my Ph.D., which shaped the direction of my research.

I am grateful to Paola Sapienza for her guidance since 2015 at the Kellogg School of Management. Her feedback on my research has played a significant role in shaping this thesis. I am also grateful to Luigi Zingales for his academic guidance, support, and inspiration since 2016. Paola Sapienza and Luigi Zingales are the co-authors of the fourth chapter of my thesis, and I am thankful for the opportunity of working with and learning from them. I will like to especially thank the members of the *Stigler Center for the Study of the Economy and the State* at the University of Chicago Booth School of Business, for the financial and academic support since 2016 during the final years of my Ph.D.

My thesis is a work in economic and business history, and I am grateful to Joel Mokyr at Northwestern University, for his inspiration that motivated me to pursue research in economic history. I wish to thank the research community of the *Center for Economic History* at the Northwestern University, and the *Chicagoland Friends of Economic History* which exposed me to the highest standards of research in economic history.

I am grateful to Anil Doshi and Joel Mokyr for their detailed feedback on my thesis, and this final version is an improved draft based on their feedback.

My thesis is an outcome of interactions in several institutions. I have thoroughly enjoyed my time spent at University College London (UCL), London School of Economics (LSE), London Business School, Northwestern University and University of Chicago, and thank these communities for being so welcoming. I would like to thank Deirdre McCloskey for her feedback, especially on my writing and rhetoric in the second and third chapters of my thesis. I am also grateful to Achal Bassamboo for his friendly advice on research and support over the last three years. I am thankful to Ronald Burt, Jeremiah Dittmar, Avner Greif, Emily Kadens, Martin Kilduff, Sheilagh Ogilvie, Raghuram Rajan, Bo Rothstein, Padmanabh Samarendra, Francesca Trivellato, Bart Vanneste and JoAnne Yates for their fruitful feedback at different stages of my Ph.D., and whose works I cite in my thesis. I would like to thank Alexander Lee for sharing his data (Lee 2015) on caste petitions with me, and James Nye of the *University of Chicago Library* for his useful advice about historical data collection on India.

I am also grateful to Ran Abramitzky, Jesus Alfaro, Kenan Arifoglu, Ken Binmore, Louis Cain, Sendil Ethiraj, Carola Friedman, Enrico Forti, Tingliang Huang, Yasheng Huang, Soong

Moon Kang, Sebastiano Massaro, Ameet Morjaria, David Moss, Nicola Persico, Guy Rolnik, Carolyn Sissoco and Fabian Waldinger for their feedback on my papers, and on writing, presenting and conducting research. I will also like to thank my friends Mike Andrews, Abdus Samad Anwar, Tom Barson, Sayoni Das, Anuj Karpatne, Lakshmi Kaul, Nilam Kaushik, Madhav Kumar, Jung Won Lee, Aditi Oleman, Sanket Patil, Prateek Prasun, Leonardo Ruivo, Janice Sanchez, Yuta Takahashi, Raj Kamal Vasu, Melanie Xue and Yiling Zhao for their fruitful conversations about my research.

Last, but not the least, I will like to dedicate my thesis to my parents and my sister Priyanka, for their lifelong support. I will also like to thank my grandparents and family, especially aunt Toula and uncle Uday, for their support and blessings; my school and college teachers, especially Hemlata Prasad, Kiran Seth and Prem Vrat for their constant inspiration; and my friends in India, UK and the USA, especially Anjul, who inspired me and shared with me this incredible Ph.D. adventure. Above all, I would like to thank Goddess Saraswati, for her blessings and grace, which made this thesis and my Ph.D. journey possible.

Thro' many dangers, toils, and snares, I have already come;
'Tis grace hath brought me safe thus far, And grace will lead me home.

- Prateek Raj

Contents

List of Figures

List of Tables

Chapter 1

Introduction

For much of the human history, we have organized ourselves in close-knit groups. In India, *Jatis* are the traditional occupational and ethnic units around which the society has organized. In medieval Europe, people were members of craft and merchant guilds. These social groups have been sources of security and identity to people, and have historically played an essential role in the economy.

But in modern societies, people regularly go beyond their traditional groups and engage in impersonal interactions with people of diverse backgrounds, in cosmopolitan markets and cities. Such interactions in a wide circle in markets (De la Croix, Doepke & Mokyr 2017), and in cities (Lucas Jr 2017), were crucial for rapid diffusion of knowledge (innovation) and formation of productive partnerships (entrepreneurship). They ultimately helped in triggering the "hockey stick" (CORE team 2017) of exponential economic growth since the industrial revolution. Beyond economic growth, the nature of social interactions also influences the quality of institutions and democracy (Putnam, Leonardi & Nanetti 1994, Guiso, Sapienza & Zingales 2016). Tocqueville (1862) when writing about the democracy in the nineteenth-century United States observed:

> In democratic countries the science of association is the mother of science; the progress of all the others depends on the progress of that one.

If the nature of social interactions and its size and diversity is so central to progress, then it is critical to understand 1) when do traditional social systems like *Jatis* and guilds persist, and 2) when do modern impersonal and cosmopolitan social systems emerge? This thesis is dedicated to understanding this "science of association". I study the reasons why social systems persist and change, and focus on two prominent historical, social systems of the world- the guild system of Europe, and the caste system of India.

Academics have become interested in the historical trends that affect the social structure of exchange (Polanyi 1944, Banfield 1967, Malone, Yates & Benjamin 1987, Putnam, Leonardi & Nanetti 1994, Ogilvie 2011, Padgett & Powell 2012, Trivellato 2014). Yet, a tough question that evades an answer is how does the circle of social exchange expand, and what factors inhibit its expansion. Why are social and economic interactions in a wide circle - with strangers and with people of diverse identities - so difficult?

In this thesis, I study historical settings where economic agents decide to either stick to their traditional groups such as guilds or *Jatis*, or else they conduct exchange in a wider circle with strangers, in impersonal markets and with people of diverse backgrounds. These micro-level decisions, about the friends they make and the networks they form can have macro-level implications on the structure of society, and the emergent culture and institutions. The historical settings in this thesis are scenarios where the social structure of exchange transformed rapidly due to cultural, institutional and technological factors.

1.1 How honesty and trust evolves among strangers?

In the second chapter, I study the trade-off between relationship-based and impersonal exchanges, and ask: *when do traders undertake risky impersonal exchange with strangers over exchange through relationships?* I model an impersonal economy where impersonal exchanges are anonymous, there is no repetition, and traders derive no prosocial utility from honesty, in contrast to previous models. Instead, in the economy norms about how to deal with strangers are conceptualized as routines (Nelson & Winter 1982) that are sticky, and they evolutionarily compete (Boyd & Richerson 1988, Mokyr 2016) and contagiously diffuse (Shiller 2017).

I find that exchange through relationships is persistent, and impersonal exchange can emerge only under a narrow set of conditions. A threshold adoption of standardized routines that induce impartial conduct can constrain cheating among strangers (threshold condition). But, these routines can only sustain if the opportunities from impersonal exchanges are significantly more beneficial when compared to opportunities through relationships (utility condition).

To discuss the results, I consider the case of sixteenth-century Europe, which was a critical period when the circle of exchange expanded. I emphasize the role of horizontal communicators (printing press) in mass diffusing standardized routines like double-entry bookkeeping. I also discuss how identification to relationship-based groups like caste can influence the outcomes of the model.

1.2 How merchant guild networks declined in Northwestern Europe?

In the third chapter, I focus on the persistence and change of a key form of social organization in Europe- of merchant guilds. I study the impact of movable-type printing press on the decline of merchant guilds in the sixteenth century. Merchant guilds were arguably the most notable form of networked business organizations in medieval Europe (Ogilvie 2011). They began to decline in Northwest European cities like Amsterdam, Hamburg, and London in the sixteenth century, but not in other cities like Cologne, Lübeck, and Seville. As several institutional innovations, like the stock market or the joint stock company, can trace their origin to Northwestern Europe, the rise of the region has been a significant research theme (Weber 1905, North 1990, Cook 2007, Harris 2009, Padgett & Powell 2012, McCloskey 2016, Mokyr 1990, Rubin 2017). I believe that looking at this period through the lens of technological and organizational change can

bring a novel perspective to the literature. So, I collected historical data to study the historical puzzle.

I find that cities in the sixteenth century, where monopolies of merchant guilds declined, had significantly higher levels of printing in the fifteenth century, as they were early adopters of printing technology. Additionally, these cities were at the Atlantic coast, where traders had a large incentive to form new connections with unfamiliar traders.

The combination of the printing press and Atlantic coast, allowed for horizontal transmission of information - especially related to commerce - and under right conditions led to a decline of merchant guilds.

1.3 How caste identity became salient in India and influenced networking and governance?

In the fourth chapter, in joint work with Paola Sapienza and Luigi Zingales, I explore how social identity gets transmitted, and how it can influence networking and governance. Firstly, drawing from the classical work of Akerlof & Kranton (2000) and Bisin & Verdier (2001) we model how the social identity of caste and cosmopolitanism can be intergenerationally transmitted, and how institutions can influence this transmission in the long run by popularizing ideal caste related prescriptions. Further, drawing from the ideas of Cohn (1984) and Dirks (2001), we study if the colonial census of 1901 led to an increase in the salience of caste identity in India.

We find that the regions which formed district committees in 1901 to infamously rank castes, were not different on a range of metrics before 1901. Using this 1901 shock to the salience of identity as an experiment, we find that today these treated regions have a smaller social circle as they have lower levels of out of caste networking with professionals. These regions also have a lower quality of government and a higher presence of caste-based parties. We also find that immediately after 1901, the treated regions had a higher level of caste-based petitioning in subsequent censuses, often by organized caste groups.

1.4 Epistemology and Methodology

The three chapters focus on three key aspects of the social structure of exchange: trust, access to information and social identity in different historical settings. The theories developed in the thesis use mathematical models to highlight important tensions between different forces that determine the size of social exchange. For example, the trust model in the second chapter highlights the tension between risk and opportunity of impersonal exchange. Similarly, the model of social identity in the fourth chapter highlights the tension between individual preferences and socially and state prescribed activities and their long-term impact. Also, the theories I develop on trust and identity, borrow from fields like accounting, anthropology, law, network theory and sociology (Fanon 1967, Anderson 1991, Mayer, Davis & Schoorman 1995, Macy & Skvoretz 1998, Rousseau, Sitkin, Burt & Camerer 1998, Dirks 2001, Kadens 2014, Gong, McAfee & Williams 2016, Burt & Luo 2017, Hail, Tahoun & Wang 2018), and embed these concepts in socioeconomic utility models.

I conduct empirical and historical analysis to test the validity of the theories. The historical settings are such that the causal effects are identifiable. So in the third chapter, I use the unique geographical features of Europe to test whether information diffusion and incentives to trade played a role in the emergence of impersonal markets in the sixteenth century. Similarly, in the fourth chapter, I use the heterogeneity in the application of the unique 1901 caste censuses of India to study whether social identities can be made more salient by colonial prescriptions, and influence modern networking and governance. Additionally, I develop historical accounts to build a detailed picture of the historical periods I study. For example, I provide a historical narrative of the evolution of the social construct of caste in India (chapter four) and write a condensed account of the divergence between Hamburg and Lübeck in the fifteenth and the sixteenth century (chapter three and appendix D).

Overall, I study these social organizations at periods of significant transformation. I show how the emergence of impersonal exchange and markets requires incentives and norms that support exchange in diverse and cosmopolitan settings. In contrast, even in cosmopolitan cities like Calcutta and Madras, the popularization of prescriptions that draw distinctions between different social groups can lead to hardening of social identities, and inhibit the emergence of cosmopolitan cultures.

Chapter 2

How honesty and trust evolves among strangers?

> Go into the London Stock Exchange - a more respectable place than many a court - and you will see representatives from all nations gathered together for the utility of men. Here Jew, Mohammedan and Christian deal with each other as though they were all of the same faith, and only apply the word infidel to people who go bankrupt. Here the Presbyterian trusts the Anabaptist and the Anglican accepts a promise from the Quaker. - Voltaire (1694-1778)

Business exchanges are embedded in social relations (Polanyi 1944, Granovetter 1973, 1985, Coleman 1988, Greif 1994, Burt 1995, 2001, Kranton 1996, Uzzi 1996, Uzzi & Lancaster 2003, Saxenian 1996, Tsai & Ghoshal 1998, Gulati 1998, Fernandez, Castilla & Moore 2000). These relations are built on trust, and in some societies it is easier to trust strangers than in others. In societies where social exchanges are limited within familiar groups - like family, clan or networks like *jati*, *guanxi* and guilds - the civic or "bourgeois" (McCloskey 2016) norms of honest conduct with outsiders may not exist (Banfield 1967, Platteau 2000, Fafchamps 2011, Alesina & Giuliano 2015). Whether the circle of honest conduct is limited or wide has important consequences (Putnam, Leonardi & Nanetti 1994), for individuals, businesses and the economy[1].

At the individual level, when the circle of honest conduct is small, tightly knit groups (dense networks with high closure) are common and valuable, as such networks lower the risks associated with trust (Coleman 1988, Gargiulo & Benassi 2000, Burt 2001). In dense networks there are fewer opportunities for brokerage (Burt 1995), as there are fewer structural holes which may not only reduce the circle of opportunity (Gargiulo & Benassi 2000) but also the diversity of information (Reagans & Zuckerman 2001). Similarly, at the level of individual businesses, in India Bloom et al. (2013) found, little trust in non-family members by firm owners limits the growth of firms as they do not expand beyond the size that could be managed by close family members. At the level of the economy, Guiso, Sapienza & Zingales (2004) found that areas in Italy where strangers are generally trusted more (generalized trust) has more financial devel-

[1]Delhey, Newton & Welzel (2011) provides a detailed analysis on the size of the radius of trust in different countries.

opment, and "households invest less in cash and more in stock, use more checks, have higher access to institutional credit, and make less use of informal credit".

If a wide circle of trust is so central to the economy, we must know how does trust between strangers emerge. In this chapter, I study how impersonal exchanges[2] emerge in societies where generalized trust in outsiders is limited, and trade happens through relationships. Older theoretical models of generalized trust assumed that agents derive an intrinsic utility from honest behaviour (Tabellini 2008b), or a highly trusting environment does not attract opportunists (Guiso, Sapienza & Zingales 2008a). Alternatively, previous literature assumed that such trust exists because agents can repeat their exchanges (e.g., Ghosh & Ray (1996) and Nowak & Sigmund (2005)), punish the strangers (Sethi & Somanathan 1996, Bowles 2016), or sanction the entire community of the opportunist (Greif 2006). These channels effectively eliminate all forms of vulnerability associated with trusting strangers as cheaters are punished through some form of individual or communal identification.

But, scholars such as Kadens (2014) have questioned economic theories that assume that familiar traders act honestly to each other when threatened with a loss of reputation, or punishment. Even in close-knit societies, cheating and reneging of promises was tolerated (Kadens 2014) as long as traders had confidence in their overall networks. Moreover, in large economies with nascent institutions, cheaters cannot be perfectly identified, and exchanges with strangers always involve a degree of vulnerability, as *cheating is "central to commerce"* and is not an aberration (Kadens 2019 forthcoming). How does trust emerge in such cases when the possibility of cheating cannot be eliminated? The theoretical models where agents enjoy a positive utility from good behaviour (Tabellini 2008b, Greif & Tabellini 2017, Bisin & Verdier 2017), argue that societies with such initial prosocial preferences have passed on these values intergenerationally. I develop a theory about the emergence of honesty and generalized trust, where we do not need to assume that societies have intrinsically different preferences, nor do we need to assume that the possibility of dishonest behaviour is eliminated.

Central to the theory I develop is the notion that trusting requires taking a risk (Kadens 2019 forthcoming), and self-interested traders need to show *"willingness to be vulnerable"* (Mayer, Davis & Schoorman 1995, Rousseau, Sitkin, Burt & Camerer 1998) when they trust strangers. So, I consider a model impersonal economy where traders trade with anonymous partners, where their interactions cannot be repeated, and they enjoy no prosocial utility from good behaviour. Traders have different norms about how to deal with strangers, and these norms are embedded in their business routines (Nelson & Winter 1982). The first type of traders are dishonest towards strangers, and they favor familiar partners, while the other type is uniformly honest towards them. Traders become honest towards strangers if they adopt regular and predictable (Nelson & Winter 1982) i.e., standardized routines that reduce partiality in conduct, and make them honest and reliable uniformly, irrespective of familiarity with partners. Adopting processes that promote uniform standards of quality like ISO and TQM standards (Benner & Tushman 2003), or innovations like the adoption of double-entry bookkeeping for financial transparency can be some of the standardized routines. Norms about how to deal with strangers are replicators, and

[2]Impersonal exchange is also called arm's length exchange (Uzzi 1996, Rajan & Zingales 1998) or generalized exchange (Ogilvie 2011).

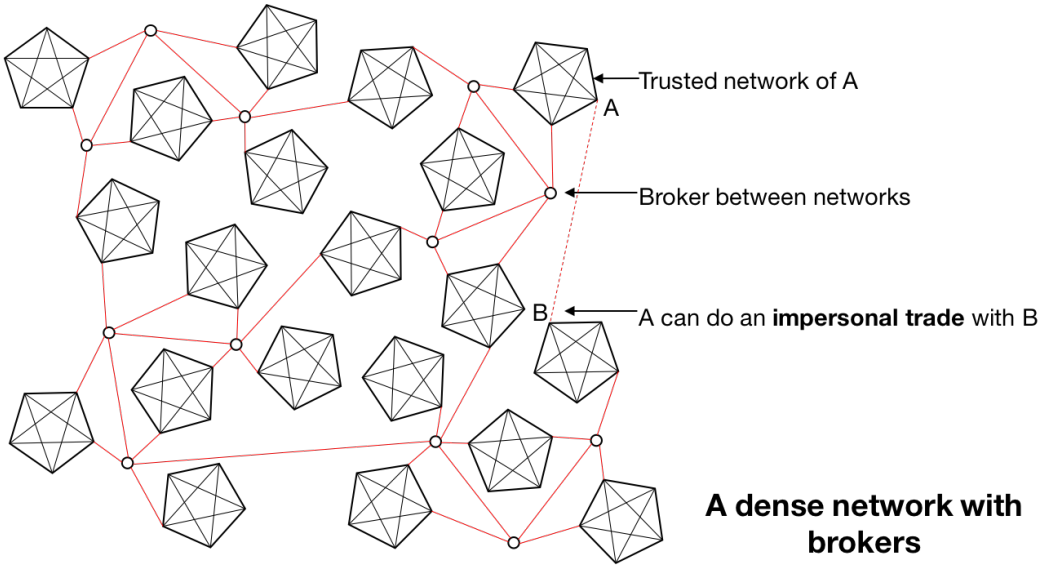

Figure 2.1: The figure represents an example network of individuals where an agent like A, has a trusted network. Whenever agents need to make an outward exchange beyond the trusted network (for example A exchanging with B), they can either do a relationship-based exchange, through trusted intermediaries (brokers) or else they can undertake a risky and arm's length exchange directly.

they are occasionally revised based on the average fitness of different ethical types.

Based on their past experiences with strangers, traders also decide whether to be impersonal or relationship-based. Impersonal traders trust strangers (like if A trusts B in Figure 2.1) and enter in a risky prisoner's dilemma game with them, without demanding any guarantees, as they are eager to make new deals (Kadens 2014). In contrast, relationship-based traders distrust strangers and do not play the prisoner's dilemma game. Instead, they get a steady payoff (like through A's trusted network of intermediaries in Figure 2.1), and if a stranger approaches them for exchange, they demand guarantees. They only conduct a one-sided prisoner's dilemma exchange with these strangers if strangers are willing to be on the side that takes all the risk in the exchange. In related literature (Dasgupta 2000, Gong, McAfee & Williams 2016, Kadens 2019 forthcoming), traders who do not trust strangers demand verification, which like a demand for guarantees reduces the risk in impersonal exchange, but at the same time reduces the circle of impersonal opportunities.

The setting models an impersonal economy with multiple types of impersonal exchanges, including one-sided prisoner's dilemma exchanges (also called the *fundamental problem of exchange* by Greif (2000)) between impersonal traders (who trust and are eager to make deals) and relationship-based traders (who distrust and demand guarantees). Honest impersonal traders can thrive in such one-sided exchanges by taking risks and making many successful impersonal exchanges (deals) *if* there are an ample fraction of honest relationship-based traders on the other side who bring large gains. As embedded routines are only occasionally revised due to being sticky[3], this stickiness creates the possibility that there are honest relationship-based traders who

[3]These routines can also be viewed as heuristics whose significance in everyday decision-making is well studied. Yamagishi et al. (2007) commented "Heuristics play a decisive role in the decision-making process that takes place in experimental games as well as everyday life, and the use of heuristics is often more adaptive than deliberate decision-making". Kandori & Obayashi (2014) studied the community unions of Tokyo (Tokyo Managers' Union) and found that cooperation sustained in such loosely knit organizations through "simple heuristic reasoning".

follow standardized routines that do not cheat with vulnerable impersonal traders in one-sided prisoner's dilemma exchanges.

When traders trade, too many honest and impersonal traders who trust and are eager to make deals, create space for dishonesty in trade. Too many dishonest traders give rise to distrust among honest relationship-based traders. Too many honest relationship-based traders who distrust, give rise to a demand for guarantees, which reduces cheating and gives rise to honest impersonal traders. Such a *cycle of trust, dishonesty, distrust, and honesty* can give rise to an impersonal equilibrium of exchange which co-exists with dishonest behaviour which is found to be common in historical (Kadens 2014, 2019 forthcoming) and modern markets[4] (Gong, McAfee & Williams 2016, Hail, Tahoun & Wang 2018). This equilibrium exists under two narrow conditions:

- **Utility condition:** the benefit from impersonal exchanges is significantly higher than the loss from being cheated, and the benefit from relationship-based exchanges.
- **Threshold condition:** a threshold of agents adopts honest norms towards strangers, i.e., they adopt standardized routines that induce impartial business conduct.

In the absence of either of these conditions, only relationship-based exchange persists, and the economy cannot transition to the impersonal equilibrium. I simulate these results to confirm the intuition. In the absence of the possibility of the emergence of the impersonal equilibrium, traditional institutions that cater to relationship-based exchanges are unwilling to reform. In an extension, I show how generalized institutions that serve all parties impartially can endogenously emerge when conditions are ripe for impersonal exchange.

The theory developed shows that incremental transmission of standardized routines cannot push a relationship-based economy to transition into an economy driven by impersonal exchange. So, traditional economies where such routines were transmitted vertically from one generation to another, could not develop sustainable forms of impersonal exchange. Instead, the emergence of impersonal exchange required a mass horizontal transmission of such standardized routines so that they diffused virally between peers[5]. I discuss the implications of the theory in the context of the rise of impersonal markets in sixteenth-century Northwestern Europe, where Atlantic trade increased the willingness to conduct trade with strangers, and horizontal transmission of practices like the double entry bookkeeping led to the adoption of standardized routines that may have constrained partial conduct.

[4]Gong, McAfee & Williams (2016) develop an evolutionary model of fraud cycles, where investors decide whether to verify the product sellers at a cost, and product sellers decide whether to produce low-quality products. Risk, which is taken by investors, is one-sided in the model. Hail, Tahoun & Wang (2018) develop a dataset of historical scandals and regulations from 1800. They find in the data "'vicious cycle' of scandals followed by bouts of regulatory activism".

[5]Rothstein (2011a) discusses the role of "big bang" reforms in fighting corruption, and provides a case-study of how corruption was eradicated in 19th century Sweden.

2.1 Modeling an Impersonal Economy

2.1.1 Relationship-based channels of exchange

Honest conduct in impersonal settings presents two challenges. Firstly, what enables *diffusion* of "bourgeois" (McCloskey 2016) or "civic" (Bowles 2016, Bisin & Verdier 2017) values, beliefs, norms, ideas or codes that make honest exchanges with strangers possible? Secondly, what *sustains* honest conduct, and stops cheating from flourishing in trusting environments, like in an impersonal market made up of unfamiliar traders?

Sociology literature has a long tradition of using agent-based modeling for studying the diffusion of norms and conventions (Macy & Willer 2002, Bianchi & Squazzoni 2015). Economists too have been interested in the formal modeling of diffusion and evolution of culture, and narratives (Boyd & Richerson 1988, Mokyr 2016, Shiller 2017)[6]. Using models of (vertical) cultural transmission between generations (Bisin & Verdier 2001), a branch of literature has explored the transfer of civic values of good conduct between generations (vertical transmission). In the literature that looks at values, values provide intrinsic benefits like the feeling of "warm glow" (Tabellini 2008b) or having a "city orientation" (Greif & Tabellini 2017) or "civic culture" (Bisin & Verdier 2017), beyond the pure monetary gains. The models show that small differences in initial norms between two societies can lead to persistent and diverging outcomes on how the societies are organized, and on the circle of honest or cooperative conduct in them (showing institutional and cultural persistence). But, the theories do not explain, how do societies develop different initial attitudes. One possible mechanism through which trust in others can change is by temporary shocks to the return to being trusting towards strangers (Guiso, Sapienza & Zingales 2008a). Such a shock increases the beliefs regarding the trustworthiness of others, which then persists. But, honest conduct is itself affected by the prevailing level of trust. One can expect that an economy where people start trusting strangers would also attract cheaters.

Another branch of literature has focused on how cheating can be contained from flourishing in cooperative environments. Cooperation in prisoner's dilemma settings can be sustained if there is some possibility of a repeated exchange or reputation building (Ghosh & Ray 1996, Nowak & Sigmund 2005), or if the cheaters are threatened by a vengeful and costly retaliation (Sethi & Somanathan 1996, Fehr & Gächter 2000, 2002)[7]. These methods are considered *private mechanisms* that individuals can privately use to constrain a partner from cheating. Alternatively, a *third party* (like ruler) can identify and punish cheaters by improving contract enforcement (or rule of law) (North 1990, Milgrom, North & Weingast 1990, Bowles 2016). Moreover, Greif (2006) proposed a *community mechanism* of community responsibility system that enables impersonal exchange between traders, as long as their community identities are known. In the community responsibility system, cheating by one member of a community

[6]Shiller (2017) discussed a formal epidemiological approach to study the diffusion of narratives, providing a framework for mathematically modeling social contagion. Similarly, Mokyr (2016) proposed an evolutionary "Market of Ideas", where culture can be conceptualized as a menu to which entrepreneurs could add, and customers could choose.

[7]Merchant guilds in Europe and kinship networks elsewhere (and Europe) were sustainable economic institutions as they provided an effective mechanism for monitoring and punishing opportunism (Greif, Milgrom & Weingast 1994).

follows sanctions on the entire community by the members of the community of the cheated partner. Using agent-based modeling, Macy & Skvoretz (1998) developed a *social network-based mechanism*, where strangers were embedded in a social network, and cooperation was sustained by recognizing signals of trustworthiness, and exiting from an exchange if the signal was fake. Similarly, Kadens (2011) emphasized the significance of locally and network embedded customs that merchants followed to conduct business exchange, where merchants were willing to tolerate dishonest behaviour or delayed payments as long as they had "confidence" that the reciprocal nature of their networks will pay them off in the long run (Kadens 2014).

The private, third party, community or social network-based mechanisms restrict cheating by identifying cheaters, and enforcing punishment on them (Axelrod 1984, Ghosh & Ray 1996, Nowak & Sigmund 2005, Sethi & Somanathan 1996, Fehr & Gächter 2000, 2002, North 1990, Milgrom, North & Weingast 1990), their communities (Greif 2006), or by not engaging with them in the first place (Macy & Skvoretz 1998, Bravo, Squazzoni & Boero 2012). Such identification of cheating is imperfect, and even close-knit medieval networks had instances of cheating and reneging of promises that was tolerated (Kadens 2014, 2019 forthcoming). This problem is especially even greater in large economies, where cheating cannot be completely negated. For example, the community responsibility system functioned particularly well in small medieval markets of Europe. But, it began to falter as the markets grew, and identities of traders became more difficult to assess (Greif 2006). Understanding impersonal exchange with a stranger in the absence of repeated exchange, without the clear knowledge of their identity, and in the absence of credible third-party enforcement becomes important, when we are interested in understanding emergence of impersonal exchange in settings where impartial legal institutions that serve all parties uniformly are not already well developed, and the economy does not have the prosocial values needed for trust.

2.1.2 Ingredients of impersonal exchange

Voluntary Prisoner's Dilemma

I now develop a model impersonal economy where traders face a dilemma, of whether to trust and be honest towards unfamiliar traders, who may be potential partners in trade. Here, in any given period traders are randomly paired with an unfamiliar trading partner, and the pair can play a game similar to a prisoner's dilemma, where both parties have incentives to cheat and not honor promises. Modeling trade in prisoner's dilemma-type settings has been common in literature (Milgrom, North & Weingast 1990, Kandori 1992, Ellison 1994, Macy & Skvoretz 1998, Dixit 2003, Tabellini 2008b). For example, Milgrom, North & Weingast (1990) modeled trade in cities and fairs of Medieval Europe, in a similar manner and wrote:

> With the exception of barter transactions, in which physical commodities are exchanged on the spot, virtually all economic transactions leave open the possibility of cheating. In the Champagne Fairs, where merchants brought samples of their goods to trade, the quantities they brought were not always sufficient to supply all the potential demand. Then, the merchants sometimes exchanged promises - to deliver goods of like quality at a particular time and place, or to make payment in a

> certain form. Promises, however, can be broken. To represent the idea that cheating may be profitable in a simple exchange, we use the Prisoners' Dilemma (PD) game as our model of a single exchange transaction.

Given the temptations to cheat, traders tend to be reluctant to partner with outsiders (like traders until sixteenth century in Europe or business people in developing countries (Bloom et al. 2013)) and prefer the relationship-based channels like guild or kinship ties. So, I am interested in incorporating the reluctance as well, and in understanding the mechanism of diffusion and sustenance of impersonal honest exchange in such relationship-based economies. So, in the economy, being matched in the prisoner's dilemma game is voluntary, similar to the model by Macy & Skvoretz (1998) which was a "prison with an option to exit"[8]. The key difference between this model and the model by Macy & Skvoretz (1998) is that, in the Macy & Skvoretz (1998) model agents decide to exchange with a stranger on the basis of the signal they receive about that stranger, while in this model they decide whether to exchange with a stranger on the basis of their experiences with other strangers in the past (Reputation of a stranger). Bowles (2016) in his discussion on trust in market-exposed economies (*Doux Commerce*) wrote of this past experience-based mechanism: "people learn from their market experiences that fair dealing with strangers is often profitable".

Norms as replicating routines

In the economy, traders can either have a norm of honesty or a dishonesty towards strangers (unfamiliar traders). Traders who are dishonest towards strangers favor familiar partners. But traders can become honest towards strangers if they adopt routines (Nelson & Winter 1982) like ISO or TQM quality standards or double-entry bookkeeping rules that standardize business practices in business, and reduce the scope for partiality. Introducing the notion of routines being replicators, Nelson & Winter (1982) wrote:

> In our evolutionary theory, these routines play the role that genes play in biological evolutionary theory. They are a persistent feature of the organism and determine its possible behaviour (though actual behaviour is determined also by the environment); they are heritable in the sense that tomorrow's organisms generated from today's (for example, by building a new plant) have many of the same characteristics, and they are selectable in the sense that organisms with certain routines may do better than others, and, if so, their relative importance in the population (industry) is augmented over time.

I would like to capture the idea that, *norms regarding honesty towards strangers are replicators*, that grow or shrink at a rate proportional to their relative "fitness" in the market. A simple

[8] An addition to the literature of evolution of cooperation has been the aspect of voluntary participation (Hauert, De Monte, Hofbauer & Sigmund 2002, Semmann, Krambeck & Milinski 2003), which can regulate free-riding behaviour. When engaging in social interaction with strangers is a choice, trust can be shown to arise when agents play public goods games in small groups. But, as group sizes increase, this cooperation falls as cheating rises. When cheating rises, more agents choose not to play, and non-playing behaviour replaces cheating behaviour. When non-players dominate the population, such a situation again creates a breeding ground for cooperation between small groups.

way to operationalize the idea is to assume that, traders occasionally review and compare the norms they have adopted regarding honesty towards strangers, described in section 2.1.4, and after the review adopt the norm that leads to better hard payoffs, i.e., payoffs that are earned and not just perceived. Unlike models that consider values like "warm glow" to be of intrinsic utility (Tabellini 2008b, Greif & Tabellini 2017, Bisin & Verdier 2017), in this impersonal economy traders are interested only in economic utility. While traders are concerned about their hard payoffs, they do not switch between norms in each period, i.e., *norms of traders are sticky*, and they only review the norms occasionally.

So, consider a unit mass of traders, each of whom is either honest or dishonest towards strangers, depending on whether they adopt the standardized routines or not. An honest (to strangers) trader has invested in standardized routines and has a norm $n = H$, while a dishonest (to strangers) trader has a norm $n = D$. The aggregate behaviour of the traders is described by a population state $p \in P$, where $P = \{p \in R^n_+ : p_H + p_D = 1\}$ is a simplex and p_H represents the proportion of traders who adopt standardized routines and are honest to strangers. Similarly, p_D represents the proportion of traders who do not adopt such standardized routines and are dishonest whenever they come across a stranger. Together honest and dishonest traders make up the whole population.

Multiple Types of Exchange

Let in each period traders have the option to approach strangers, and engage in an impersonal exchange with them. Such exchange has the potential for high reward, but also the risk of loss because of being cheated[9]. Alternatively, traders can choose a less rewarding option where they do not approach strangers, and so they are insulated from cheating by strangers. The safe but less rewarding option could be to choose relationship-based exchange and to exchange with familiar contacts like intermediaries of a merchant guild. But, relationship-based traders can still be approached by strangers who conduct impersonal trade. As relationship-based traders prefer safer but less rewarding option, they distrust approaching strangers. So, skepticism drives relationship-based traders to demand a guarantee of assured return, from any stranger that approaches them. Traders that are honest can provide such guarantees (pre-commit to honest conduct), and so they can approach and exchange with skeptical relationship-based traders. In the setting, the relationship-based traders can turn out to be dishonest, and especially when there are many honest impersonal traders, cheating becomes lucrative and thrives. So, along with the norms regarding honesty ($n \in \{H, D\}$), in each period the traders also choose an orientation, of being trusting of strangers or not. Traders change their orientation based on their past experiences, described in section 2.1.5.

Based on the model setup, in each period the traders have an orientation $o \in \{R, I\}$, where among the traders with norm n, p_{nR} engage in relationship-based exchange $(o = R)$ and p_{nI} engage in impersonal exchange $(o = I)$. If relationship-based traders are not approached by an honest impersonal trader who could offer them a guaranteed return, honest relationship-based traders (HR-type) earn m, while dishonest relationship-based traders (DR-type) earn

[9]Impersonal exchange because of its inherent vulnerability and risk resembles the notion of trust as "willingness to be vulnerable" (Mayer, Davis & Schoorman 1995, Rousseau, Sitkin, Burt & Camerer 1998).

Table 2.1: Different types of exchange

Focal Trader (F)	Partner Trader (P)	Type of Exchange	F's Payoff	P's Payoff
HR-type	not paired	Relationship-based Exchange	m	
DR-type	not paired	Relationship-based Exchange	m	
HI-type	HI-type	Simultaneous Exchange	c	c
HI-type	DI-type	Simultaneous Exchange	$-h$	$c+h$
HI-type	HR-type	Sequential Exchange	c	c
HI-type	DR-type	Sequential Exchange	$-h$	$c+h$
DI-type	HR-type	Sequential / No Exchange	0	m
DI-type	DR-type	Sequential / No Exchange	0	m

$g = m + i$. The value i represents the benefit an honest relationship-based trader forfeits when they adopt the impersonal routine[10]. *Impersonal traders exchange with strangers, and they get randomly paired with one of the traders of the population, with uniform probability*. Such pairing leads to different types of exchanges summarized in Table 2.1.

A few notes on the different types of exchange that emerge will be helpful:

- In a *relationship-based exchange*, traders refuse to get paired and trade through relationships and earn m.
- In a *simultaneous exchange*, paired traders play a prisoner's dilemma with payoff c on cooperation, $c + h$ on cheating, and $-h$ on being cheated.
- In a *sequential exchange* paired traders play a prisoner's dilemma where impersonal type moves first, i.e., relationship-based type demands a guarantee from impersonal type.

If an honest impersonal trader (HI-type) gets paired with another honest impersonal trader, both earn $c > g > m$ in a simultaneous trade. If the honest impersonal trader (HI-type) gets paired with an honest relationship-based trader, both earn $c > g > m$ in a sequential trade, where the impersonal trader moves first as the relationship-based trader demands a guarantee. But, if the honest impersonal trader is paired with a dishonest trader, impersonal or relationship-based, the honest trader loses h while the dishonest trader earns $c + h$[11]. Once again, the trades with impersonal traders are simultaneous, while trades with relationship-based traders are sequential.

If a dishonest impersonal trader (DI-type) gets paired with another impersonal trader it earns 0 if the paired trader is dishonest ($c + h$ if the trader is honest), while if it gets paired with a relationship-based trader, it earns 0, as the relationship-based trader refuses to exchange with the dishonest trader that cannot provide a credible guarantee of return. It is assumed that $i = 0$[12], i.e., adopting standardized routines is costless (and similar to choosing a norm) and

[10] In the baseline case of the model $i = 0$, i.e., adopting impersonal routines of impartial conduct is costless, but i in some settings can be high. For example, in parochial and fragmented societies if being partial and favoring towards insiders is a prescribed behaviour in a social group, then the traders who nonetheless adopt impartial routines can be penalized by the amount i for violating the prescriptions related to their *social identity* (see Akerlof & Kranton (2000) for a theory on social identity, and Mair, Martí & Ventresca (2012) for an example of market inclusion of women in Bangladesh.

[11] In section 2.3 a ruler sets the value of h, i.e., ruler can reduce cost/temptation of cheating h, by improving law and order and probability of detection.

[12] In subsection 2.2.5 I briefly discuss the impact of a higher i, where due to a range of reasons (cost of social identity or cost of actual adoption of standardized routines) honest traders forfeit a value $i > 0$.

$g = m$[13].

A few notes on the interpretation of the payoffs will be helpful:

- m is the payoff from relationship-based trading arrangements, and it is high if the arrangements are efficient. This measure can proxy the economic *power of intermediaries*, where a larger m implies more powerful intermediaries.
- c is the payoff from an impersonal exchange, and it is high when such opportunities are large. This measure can proxy the *size of potential economic opportunities*.
- h is the temptation/loss from a dishonest exchange, and it decreases as general contract enforcement improves. This measure can proxy the *quality of public institutions*, where a larger h implies poorer public institutions.

2.1.3 Payoff of different types

At any given time t in an economy of population $p = 1, p_H(t)$ traders are of honest ($n = H$) norm. The honest traders, alternate between orientations $o \in \{R, I\}$. $p_{HI}(t)$ honest traders behave as HI-type that exchange with strangers and they have impersonal orientation (and earn c on being paired with H-type traders and lose h on being paired with D-type traders), and $p_{HR}(t)$ honest traders behave as HR-type traders that have relationship-based orientation (and earn c on being paired with HI-type traders and m otherwise). Among dishonest traders, the model setup presents Lemma 1 on their choice of orientation.

Lemma 1 *There are no dishonest traders of impersonal orientation ($p_{DI} = 0$), because for a dishonest trader, having a relationship-based orientation is always preferred over impersonal orientation.*

Lemma 1 emerges because the only scenario where dishonest impersonal traders earn is if they are paired with honest impersonal traders. Dishonest impersonal traders are declined exchange when partnering with relationship-based traders, and they have no particular advantage over dishonest relationship-based traders in an exchange with other impersonal traders, given the uniform nature of random matching. Even if matching technology of impersonal traders was not uniform and was skewed to favor other impersonal traders, dishonest traders would still choose relationship-based orientation, unless the skew was too large[14]. In an extreme case

[13] Another way to interpret c and h is, h is the effort required to produce a result that benefits the partner by $c + h$. If both partners put the effort h both earn c, but, if one of them does not put in the effort, the effort-maker loses h, while the partner gains $c + h$. If neither put any effort, both earn 0. Pre-commitment or guaranteeing an assured return, can mean that in a pairing of relationship-based and impersonal traders, the impersonal trader moves first and puts effort h that benefits the relationship-based trader, and the relationship-based trader moves second and depending on norm $n \in \{H, D\}$ either puts in effort (honest) or does not (dishonest). The exchange resembles a prisoner's dilemma when played simultaneously between two paired impersonal traders, while it resembles an investment game when played sequentially between an impersonal trader and then a relationship-based trader.

[14] Skewed random matching: The payoff of a dishonest relationship-based trader is given as $P(HIpairsDR)(t)(c + h) + (1 - \beta P(HI - type)(t))m$, where $P(HIpairsDR)$ is its probability of being paired by an HI-type trader. The payoff of a dishonest impersonal trader is given as $P(HIpairsDI)(t)(c + h)$ where $P(HIpairsDI)$ is its probability of being paired by an HI-type trader. Now, in uniform random matching, $P(HIpairsDR) = P(HIpairsDI) = p_{HI}$. If there is a skew where impersonal traders are more likely to be paired with other impersonal traders, then the condition under which impersonal orientation is favored over relationship-based trade by a dishonest trader is if $P(HIpairsDR)$ is quite low compared to $P(HIpairsDI)$, i.e. $\frac{P(HIpairsDR)}{P(HIpairsDI)} < \frac{c+h-m/P(HIpairsDI)}{c+h-m}$.

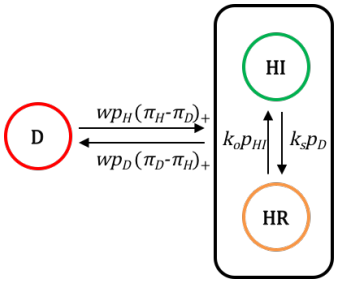

Figure 2.2: A trader transitions to a different ethical type with a rate proportional to the difference in average payoff (π_e) between types and the size of population of the opposite population $\rho_{ij}(t) = wp_j(t)(\pi_j(t) - \pi_i(t))_+$. An honest impersonal trader becomes relationship-based with probability k_s on encountering a cheater (with probability p_D). An honest relationship-based trader becomes impersonal with probability k_o on encountering an impersonal trader (with probability p_{HI}).

of skewing where impersonal traders were only paired with other impersonal traders, then that would mimic a pure simultaneous prisoner's dilemma game, and honest impersonal exchange cannot sustain in that scenario. The importance of uniformity highlights the critical role of the interaction between impersonal and relationship-based traders, which creates the sequential settings for exchange. In such sequential settings, honest and impersonal traders can thrive, under a limited set of conditions.

I consider the case of pure random matching where pairing is uniformly random, that does not discriminate between impersonal and relationship-based orientation. So, $p_D(t) = 1 - p_H(t)$ traders are of DR-type that are dishonest and choose relationship-based orientation (and earn $c + h$ on being paired with HI-type traders and $g = m$ otherwise). So, average utility earned by honest (H-type) ($\pi_H(t)$) and dishonest (D-type) ($\pi_D(t)$) traders at time t is given as:

$$\pi_H(t) = \frac{p_{HI}(t)}{p_H(t)} \times \underbrace{(p_H(t)c - (1 - p_H(t))h)}_{\text{HI earns } c \text{ with prob. } p_H, \text{ and } -h \text{ otherwise}} + \frac{p_{HR}(t)}{p_H(t)} \times \underbrace{(p_{HI}(t)c + (1 - p_{HI}(t))m)}_{\text{HR earns } c \text{ with prob. } p_{HI}, \text{ and } m \text{ otherwise}} \tag{2.1}$$

$$\pi_D(t) = \underbrace{p_{HI}(t)(c + h) + (1 - p_{HI}(t))m}_{\text{D earns } c+h \text{ with prob. } p_{HI}, \text{ and } m \text{ otherwise}} \tag{2.2}$$

2.1.4 Evolution of norms (Honest vs Dishonest)

In each period t, a fraction w of the population review their norms regarding honesty towards strangers, by randomly observing another trader from the population and comparing their payoff with the payoff of the observed trader.The probability that the trader changes its norm i to the norm j of the observed trader is proportional to the difference in the earned utilities of type j and i traders, at time t, where if type i traders are better off than the observed type j traders, type i traders never change their norm. Figure 2.2 illustrates this transition between D and H types.

The evolutionary dynamics[15] is given as:

[15]At the aggregate level the revision protocol can be written as: $\rho_{ij}(t) = wp_j(t)(\pi_j(t) - \pi_i(t))_+$ where ρ_{ij} is the probability that a trader with norm i revises to norm $j(T = i \to j)$, and $\pi_i(t)$ and $\pi_j(t)$ is the average payoff of

Evolutionary dynamics:

$$\frac{dp_H(t)}{dt} = -\frac{dp_D(t)}{dt} = p_D(t)\rho_{DH}(t) - p_H(t)\rho_{HD}(t) = \underbrace{wp_D(t)p_H(t)(\pi_H(t) - \pi_D(t))}_{\text{H revises with prob. } wp_D \times \text{ diff. of avg. utilities}} \tag{2.3}$$

Equation 2.3 captures the idea that norm is a replicator, i.e., market selects the economically "fitter" norm, unfiltered by any particular preferences of the traders.

2.1.5 Evolution of Orientation (Trust vs Distrust)

While traders revise their norms regarding honesty towards strangers, honest traders also revise their orientation. Let at period $t+1$ an impersonal honest trader become relationship-based with probability k_s on being cheated (paired with a U-type trader) in period t[16]. Similarly, let at period $t+1$ a relationship-based honest trader become impersonal with probability k_o on being cooperated (paired with an HI-type trader) in period t[17]. Figure 2.2 illustrated this transition between HI and HR-types. So behavioural dynamics is given as:

behavioural dynamics:

$$\frac{dp_{HI}(t)}{dt} = \underbrace{k_o p_{HR}(t)p_{HI}(t)}_{\text{HR turns HI with prob. } k_o p_{HI}} - \underbrace{k_s p_{HI}(t)p_D(t)}_{\text{HI turns HR with prob. } k_s p_D} + \underbrace{\frac{p_{HI}(t)}{p_H(t)}\frac{dp_H(t)}{dt}}_{\text{new H turns HI with prob. } f_{HI} = \frac{p_{HI}(t)}{p_H(t)}} \tag{2.4}$$

Note that in the economy, unlike norm that is replicated only occasionally, the orientation of honest traders is a transient choice that traders make each period based on their experience. So orientation is a behavioural rather than an evolutionary trait because honest traders would frequently alternate between being impersonal and relationship-based on their on-the-go estimate of the market environment.

2.1.6 The Relationship-Based and Impersonal Equilibria

The model developed has the following features that are suitable for studying impersonal exchanges:

- Firstly, being a one-shot prisoner's dilemma between strangers, the model is more conservative and does not consider regulating mechanisms like reputation and punishment.

the traders with norm i and j at time t. When w is small it implies switching between honest and dishonest norms is uncommon. The pairwise imitation protocol (Sandholm 2015) implies that rate of diffusion of norms is proportional to relative payoff. The proportionality constant $w > 0$ can be interpreted to regulate the timescale of imitation.

[16]Skepticism: Among honest traders, impersonal (HI-type) traders become relationship-based (HR-type) traders if they estimate that the number of dishonest traders in the population is large. Clearly in any learning protocol that an honest impersonal trader has, an experience of dishonest exchange increases the estimate and may change impersonal trader's orientation.

[17]Optimism: Among honest traders, relationship-based (HR-type) traders become impersonal (HI-type) if they estimate that the number of dishonest traders in population is small. Clearly, in any learning protocol that an honest relationship-based trader has, an experience of honest exchange decreases the estimate and may change relationship-based trader's orientation.

The model focuses on generalized trust in strangers- an important aspect of impersonal market exchanges (Fafchamps 2011, Macy & Skvoretz 1998).

- Secondly, unlike existing models the model does not consider values of prosociality (Tabellini 2008b, Greif & Tabellini 2017, Bisin & Verdier 2017) to be of intrinsic utility that supports cooperation. Accordingly, all traders, with their norms, attempt to maximize their material payoffs.

- Thirdly, the voluntary prisoner's dilemma setup can study the choice between the relationship-based and impersonal exchanges so highlighting the dynamics of relationship-based vs. impersonal exchange, an important aspect in understanding structural transitions (Fafchamps 2011, Ogilvie 2011, Greif & Tabellini 2017, Macy & Skvoretz 1998).

- Fourthly, by using an evolutionary model where traders make decision on-the-go as conditions evolve, I can understand the evolution of norms in large population more realistically where all traders operate in a market with competing norms. In the model, traders are calibrating both their norms as well as their orientation based on the conditions available to them, and so we can understand their behaviour both in equilibrium and in phases of transition.

Given the dynamics in Equations 2.3 and 2.4, we have Proposition 1[18].

Proposition 1 *There exist two sets of stable equilibria:*

1. *Relationship-based Equilibrium:* $p_H \in [0, p_H^{cr})$ *and* $p_{HI} = 0$*, represents a non-asymptotic equilibrium where* $p_H^{cr} = \frac{k_s}{k_s+k_o}$.

2. *Impersonal Equilibrium:* $p_H = 1 - \frac{k_o h}{k_s(c-m)-k_o m}$ *and* $p_{HI} = 1 - \frac{(k_o+k_s)h}{k_s(c-m)-k_o m}$ *represents an asymptotic equilibrium if* $w < \bar{w}$ *where* $\bar{w} = \frac{k_s c-(k_s+k_o)m}{h(c-f_{HI}(c-m))}$ *and* $f_{HI} = \frac{p_{HI}}{p_H}$*; distinct from (i) if* $\frac{k_s}{k_o+k_s} > \frac{h+m}{c}$.

3. *Condition for Transition: If system is close to equilibrium (1) i.e.* $p_{HI}(0) \to 0$*, and* $p_H(0) > p_H^{cr}$ *when* $p_H^{cr} > \frac{h+m}{c}$ *and* $w < \hat{w}$ *the system converges to the asymptotic equilibrium (2), and (1) otherwise.*

Figure 2.3 plots the movement of population with given level of honesty (p_H) and trust (p_{HI}). There are four regions that emerge from partitioning by curves $\frac{dp_H}{dt} = 0$ and $\frac{dp_{HI}}{dt} = 0$. In region 1, the level of honesty in the population is high, but there is some level of distrust. Such level of healthy distrust in an honest economy increases the levels of honesty ($p'_H > 0$) and trust ($p'_{HI} > 0$). In region 2, the level of honesty and trust are both high, and there is little distrust or caution in the population. A lack of healthy distrust in an honest economy decreases the levels of honesty ($p'_H < 0$) (while trust ($p'_{HI} > 0$) increases) as instances of cheating increase. In region 3, the level of honesty is low, and relatively the level of trust is high. A lack of distrust

[18]The proof of Proposition 1 is given in appendix C.1. Section B simulates the model impersonal economy, showing a shift from relationship-based to impersonal equilibrium, when condition for transition is favorable (Figure B.2).

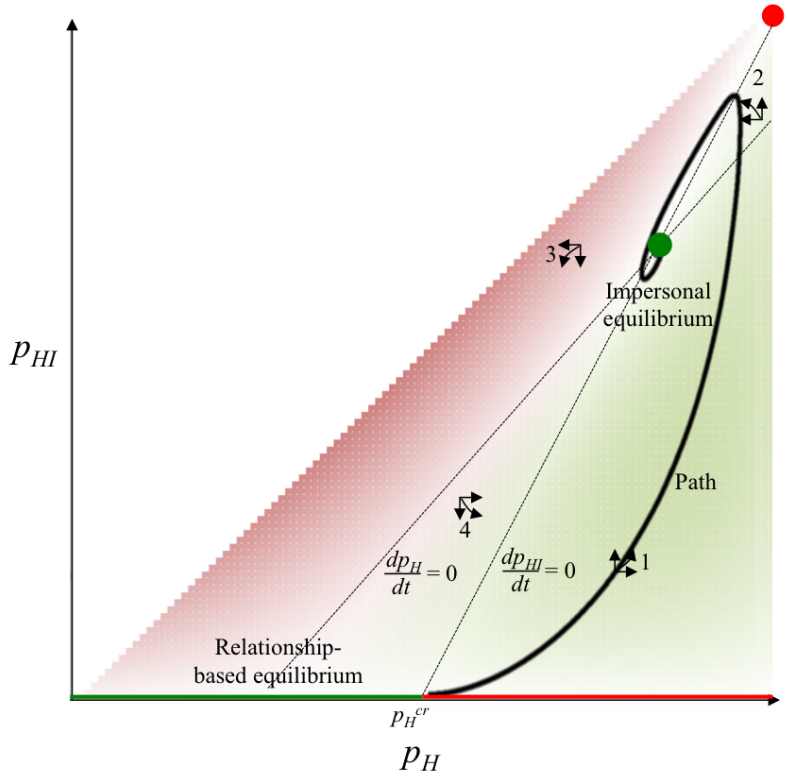

Figure 2.3: The graph plots the movement of population with given level of honesty (p_H) and trust (p_{HI}), as described by Proposition 1, with the relationship-based equilibrium ($p_H \in [0, p_H^{cr})$ and $p_{HI} = 0$), and the impersonal equilibrium (where curves $\frac{dp_H}{dt} = 0$ and $\frac{dp_{HI}}{dt} = 0$ intersect). In region 1, $\frac{dp_H}{dt}, \frac{dp_{HI}}{dt} > 0$ (green); in region 2, $\frac{dp_H}{dt} < 0$, $\frac{dp_{HI}}{dt} > 0$ (mixed); in region 3, $\frac{dp_H}{dt}, \frac{dp_{HI}}{dt} < 0$ (red); and in region 4, $\frac{dp_H}{dt} > 0$, $\frac{dp_{HI}}{dt} < 0$ (mixed). The path curve plots a path through which relationship-based economy moves to an impersonal economy. For illustration $w < \hat{w}$ is large so that a spiral convergence to impersonal equilibrium is illustrated.

in a relatively dishonest economy creates instances of cheating, and both honesty ($p'_H < 0$) and trust ($p'_{HI} < 0$) decrease as a result. In region 4, there are intermediate levels of honesty and the level of trust is low. A presence of a healthy level of distrust in a relatively honest economy increases the level of honesty ($p'_H > 0$) (while trust ($p'_{HI} < 0$) decreases). A population where $p_{HI} \to 0$ and $p_H(0) > p_H^{cr}$ (path curve), the population can move to the impersonal equilibrium, under favorable conditions defined in Proposition 1 which will be discussed in greater detail in next section 2.2.

As honest traders make on-the-go decisions regarding their choice of orientation (i.e., their orientation is not sticky unlike norms), based on experience, common biases like anecdotal thinking or aversion to risky exchanges, can affect the choice of orientation. The result of Proposition 1 holds regardless of nature of k_o and k_s. I will motivate the theory using the framework of *risk aversion* to highlight the importance of skepticism (a high k_s) (i.e., preference for non-risky relationship-based exchange) for sustaining impersonal exchange in the model impersonal economy.

If H-type traders prefer steady income gained from relationship-based exchange over the risks of impersonal exchange, they will be indifferent between being impersonal and

relationship-based, if $\frac{p_{HR}}{p_D} = \frac{(m+h)(1+\lambda)}{c-m}$[19], where $\lambda > 0$ for agents who dislike risk (or making a loss). At the impersonal equilibrium (2), $\frac{k_s}{k_o} = \frac{p_{HR}}{p_D}$, and so equilibrium (2) is supported and H-type traders are indifferent if $k_s = \alpha(m+h)(1+\lambda)$ and $k_o = \alpha(c-m)$ (where α is any constant of proportionality). So,

$$p_H^{cr} = \frac{(1+\lambda)(m+h)}{c+\lambda m+(1+\lambda)h}. \tag{2.5}$$

2.2 Theory: The two conditions

The impersonal economy developed in section 2.1 can be viewed as a race between dishonest and honest norms towards strangers. So, the norms evolutionarily compete (Mokyr 2016) and attempt to contagiously diffuse (Shiller 2017) in a competitive market. How do norms of honesty and dishonesty fare against each other? Proposition 1 shows that there exist two stable equilibria. In the first equilibrium, the population is dominated by traders who are dishonest to strangers, and all traders participate in the relationship-based exchange, and so the population of impersonal traders is 0 ($p_{HI} = 0$). Before the sixteenth century in Low Countries and England and rest of Europe until later, socio-economic exchange resembled the relationship-based equilibrium. Here exchange was confined within exclusionary and relationship-based networks, dominated by merchant guilds that wielded considerable self-serving economic and political power (Ogilvie 2011).

The second equilibrium is a polymorphic equilibrium where a positive population of impersonal traders exists ($p_{HI} > 0$) and honest traders alternate between impersonal and relationship-based exchange and coexist with the dishonest relationship-based traders. The equilibrium can be called the equilibrium of *impersonal exchange*. In impersonal equilibrium impersonal exchanges between strangers are possible, but dishonest traders invade an entirely honest and trusting economy, and only a polymorphic equilibrium exists (see section 2.2.3). In post sixteenth-century, traders in Low Countries and England began to exchange in impersonal markets outside of their traditional networks, and this resembled the equilibrium of impersonal exchange.

2.2.1 Impersonal Exchange Sustains under Favorable Incentives and Institutions

Proposition 1 shows that the equilibrium of impersonal exchange would not exist, unless the benefits from the impersonal exchange (c) considerably outweigh the benefits from non-exchange (m), and the loss of being cheated (h). It was also found that:

- the polymorphic equilibrium will fail to sustain itself unless honest traders have a behavioural dislike ($\lambda > 0$) for risky exchanges and so overreact to an experience of cheating and withdraw from the impersonal exchange and turn skeptical and relationship-based.
- also, if the rate of revision of norms is large ($w \geq \hat{w}$) and norms change rapidly, then Proposition 1 shows that impersonal equilibrium does not sustain itself.

[19] Honest traders are indifferent between being impersonal or relationship-based if their income from being impersonal ($p_H c - p_D h$) equals the income from being relationship-based ($p_{HI} c + (1 - p_{HI})m$), where h denotes loss payoff perceived by cautious traders.

It is assumed from now on that rate of revision of norms is small enough ($w < \hat{w}$) so that multiple equilibria exist.The conditions supporting impersonal equilibrium can be formalized as Corollary 1 of Proposition 1.

Corollary 1 *If* $c > \frac{1+\lambda}{\lambda}h+m$ *(utility condition), then two equilibria exist. In relationship-based equilibrium* $p_{HI} = 0$ *and in impersonal equilibrium* $p_{HI} > 0$*. Otherwise only relationship-based equilibrium exists. Utility condition is never satisfied if* $\lambda = 0$*, and the equilibrium does not exist if* $w \geq \hat{w}$*.*

For historical cases some observations from Corollary 1 are useful:

- If general contract enforcement in an economy is perfect such that $h = 0$ then, not surprisingly, as long as the payoff from the impersonal exchange (c) is higher than from traditional relationship-based exchange (m), such an impersonal exchange is sustainable.
- If payoff from impersonal exchange (c) is not highly lucrative compared to payoff through traditional relationship-based arrangements (m) and/or potential loss because of cheating in impersonal setting (h) was considerably high because of poor general contract enforcement then impersonal equilibrium is not sustainable[20].

For impersonal equilibrium to be sustainable (i) benefits from impersonal exchange (c) should be highly lucrative (ii) payoff from traditional relationship-based arrangements (m) should be relatively low and (iii) loss/temptation because of cheating in impersonal setting (h) should be relatively low (but not necessarily 0) because of better contract enforcement or because of significant gains from impersonal exchange.

2.2.2 Utility condition in Northwestern Europe

The North Italian city republics formed a unique form of non-kinship-based merchant networks, called merchant guilds. The merchant guilds helped in the expansion commerce of in Europe (Greif, Milgrom & Weingast 1994), but with time especially in beginning the of the fifteenth century, they turned more exclusive and monopolistic (Gelderblom & Grafe 2010, Ogilvie 2011). In the sixteenth century, in the Northwestern region of Europe (Low Countries and England), the merchant guild system began to decline, being replaced by impersonal markets with more impartial and impersonal institutions (Gelderblom 2004, Gelderblom & Grafe 2010). In the next chapter, I will discuss how the rise of these impersonal markets only happened in cities on the Atlantic coast, in cities like Antwerp which were major entrepôts attracting outsiders seeking business opportunities. There were significant benefits to trade with these relatively unfamiliar traders (high c) at the *Atlantic coast* which helped in creating the utility conditions needed for the emergence of impersonal markets.

While the sixteenth century Northwestern Europe may have had the favorable conditions for the emergence of more impersonal exchange, some pertinent questions emerge: Why did not impersonal trade emerge in the Iberian peninsula which also had access to the Atlantic coast?

[20]Consider the simulation exercise in Figure B.4 where a population dominated by honest and impersonal (HI-type) traders, becomes distrusting and relationship-based, as c is not significantly larger than m and h.

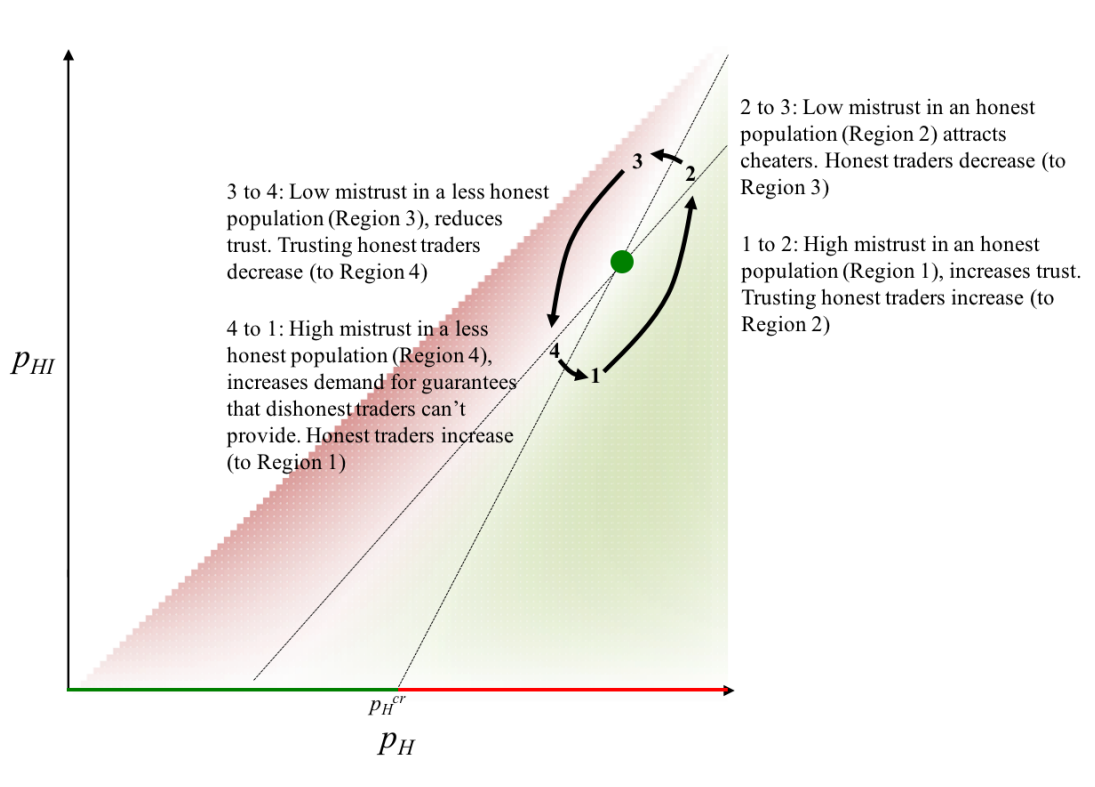

Figure 2.4: The graph shows the movement of population across regions, such that under supporting utility conditions too much trust gives rise to dishonesty (Region 2 to 3), which gives rise to distrust (Region 3 to 4), which in return gives rise to honesty (Region 4 to 1) and eventually trust (Region 1 to 2). The X-axis represents level of honesty (p_H) and the Y-axis represents level of trust (p_{HI}).

2.2.3 Relationship-Based Equilibrium Persists If Standardized Routines Do Not Diffuse

When the utility conditions are supportive, a cyclic rise and fall of honesty and trust can emerge in the modeled impersonal economy which is illustrated in Figure 2.4. This cyclic nature though will be visible only for high levels of $w < \hat{w}$ in the model.

The following is the cyclic nature of trust and honesty in the economy under supporting utility conditions:

1. **Rise of dishonesty:** When many honest traders do impersonal exchange (Region 2 of Figure 2.4), they trust, are willing to take risks, and are eager to make new deals (Kadens 2014). Such willingness to trust creates opportunities for traders to cheat. So, dishonest traders increase (p_D increases) and honest traders begin to get cheated.

2. **Rise of distrust:** As honest traders get cheated, they become skeptical and withdraw from impersonal exchange, and they become relationship-based (p_{HR} increases) (Region 3 of Figure 2.4). The skeptical relationship-based traders demand guarantees of assured returns.

3. **Rise of guarantees and trust:** As the demand for guarantees rises, honest traders that engage in impersonal exchange, and are eager to make new deals with them, can approach the skeptical traders by providing them with guarantees. The dishonest traders cannot

offer such guarantees to skeptical traders, and so they have fewer impersonal exchanges. The demand for guarantees makes honest impersonal exchange lucrative (Region 4 of Figure 2.4), and skeptical traders also switch from being relationship-based to participating in impersonal exchange (turn optimistic) (p_{HI} increases) (Region 1 of Figure 2.4).

So, a contagion of skepticism and withdrawal[21] if coupled with a reverse contagion of optimism and participation[22] under favorable utility conditions keeps cheating in check.

Cyclic nature of trust and honesty has been recently reported in the literature (Singleton & Singleton 2010, Stevenson & Wolfers 2011, Gong, McAfee & Williams 2016). For example, a textbook in fraud auditing by Singleton & Singleton (2010) describes the fraud environment as a "swinging pendulum". It describes how after the Enron and the WorldCom frauds, in 2002 the pendulum was close to the extreme end of caution on the part of US companies, and auditors that applied the stiffest requirements. But, the housing and the real-estate boom of the decade led to increased fraud, leading up to the financial crisis of 2008-09, and a return of caution in markets. Gong, McAfee & Williams (2016) develop a related model (where buyers need to trust sellers) explaining this cyclic nature of fraud, along with providing additional empirical evidence. They find that fraudulent offenses like counterfeiting/forgery and false-pretenses/swindle/confidence-game have a non-annual cyclic pattern, unlike nonfraudulent offenses like assault and burglary that have a distinct annual cyclic pattern. Similarly, Stevenson & Wolfers (2011) find a cyclic pattern in trust in business institutions (based on Gallup polls) which follows business cycles, which is consistent with the Financial trust survey by Sapienza & Zingales (2012) that reported a decline in trust following the 2008 financial crisis.

But, rising skepticism does not always pave the way for rising optimism and honesty, even if the utility conditions are favorable. Critically, in our impersonal economy if there are too few honest traders in the economy, then the positive contagion of optimism gets dominated by the negative contagion of withdrawal. This leads to a permanent decline in honesty and trust as the economy does not (re)enter the virtuous Region 1 of Figure 2.4. So, for any transition from the relationship-based equilibrium to the impersonal equilibrium to be feasible, there needs to be a minimum threshold of traders who adopt standardized routines that reduce partiality in conduct[23]. The result can be formalized in the Corollary 2 of Proposition 1.

[21] *Contagious Equilibrium* (Kandori 1992) is an equilibrium notion that sustains cooperation in iterated randomly matched prisoner's dilemma games where agents start to defect on coming across an agent with a strategy of defection. Such a switching has a contagious effect on the population, and the threat of such contagious outbreak of opportunism defers players from not acting dishonestly. While contagious equilibrium could provide a possible resolution to cooperation, but it is too vulnerable to break down. Extensions (Ellison 1994) that focus on the resolution of such vulnerability rely on simultaneity, i.e., players stop defecting (a form of collective punishment), and start cooperating in tandem. But, here again, achieving such a high level of coordination may not be practical in large populations.

[22] *Upstream Reciprocity:* A growing literature has pointed towards upstream indirect reciprocity or generalized reciprocity as one of the mechanisms through which individuals interact with strangers (Fowler & Christakis 2010, Nowak & Sigmund 2005). Individuals are willing to cooperate with strangers if they have experienced cooperation in the past (contagion of optimism). Similarly, they do not cooperate when they have been deceived in the past (contagion of skepticism). This behaviour has been explained using behavioural-imitation (Fowler & Christakis 2010, Tsvetkova & Macy 2014) and gratitude or gossip (Nowak & Sigmund 2005). When expectations can be observed, Reuben, Sapienza & Zingales (2009) found people reciprocated expectations and an expectation to be honest increased honest conduct. Even when the reputation of individuals can be visible, compared to reputation rewarding mechanism (downstream reciprocity) the mechanism of paying it forward (upstream reciprocity) has been found to have stronger and more lasting effects in sustaining generalized reciprocity (Baker & Bulkley 2014).

[23] Consider the simulation in Figures B.1 and B.2 where in a population dominated by distrusting relationship-

Corollary 2 *Conditions for transition are given as:*

- *When utility condition of Corollary 1 is satisfied, and number of initial adoptees of standardized routines that reduce partiality in conduct $p_H(0)$ is greater than a critical threshold $p_H^{cr}(\lambda, c, m, h)$ given in Equation 2.5 (threshold condition), then the economy transitions from relationship-based equilibrium to impersonal equilibrium and $p_{HI} > 0$. Factors that trigger the initial adoption will be called horizontal communicators.*
- *System stays in a punctuated and relationship-based equilibrium otherwise.*

So, regardless of the external conditions, there is a necessity of mass adoption of standardized routines, and in the absence of such mass adoption, economies stay in the punctuated equilibrium of relationship-based exchange. Punctuated equilibria are resistant to incremental changes, and so are persistent (Gould 1972, Mokyr 1990, Romanelli & Tushman 1994). For the sixteenth century Europe the theory implies that relationship-based guild centric economic organization of much of Europe was persistent, and transition to impersonal equilibrium was possible only if there was mass adoption of standardized routines that supported impersonal trade.

2.2.4 Horizontal transmission and Mass Adoption of Standardized Routines

Mass adoption of new norms, ideas, and routines requires greater *horizontal transmission* between horizontal contacts like peers and intellectual/cultural leaders, and lesser reliance on *vertical transmission* by parents, teachers, and superiors. Here, horizontal connected agents like opinion leaders, focal institutions or peer-to-peer groups can act as horizontal communicators. But, note that Corollary 1 suggests that horizontal communicators can only be persuasive in the presence of enabling utility conditions.

When enabling utility conditions exist, significant advances in information technology - that can enable a rapid horizontal transmission and adoption of new norms, ideas, and routines - can catalyze transition out of punctuated equilibrium. During the sixteenth century, did Low countries and England develop unique characteristics, that made such horizontal transmission of standardized routines possible? In the next chapter, I argue that the high penetration of *Printing* helped in the mass diffusion of information. Such information transmission helped in the diffusion of practices promoting standardized routines, like the double-entry bookkeeping in the region, that could have helped in the transition away from relationship-based guild equilibrium to impersonal market equilibrium.

An important feature of the theory developed is that the norms should be sticky ($w < \hat{w}$). Adopting and implementing a system of double-entry bookkeeping had some unique features that made them sticky:

1. **Training in *Ars Mercatoria*:** Implementing systems like the double-entry bookkeeping required training of traders in doing business with a set of rules and standards which were to be generally applied in all business transactions

based agents, honest and impersonal agents emerge ($p_{HI} > 0$), because the initial population of honest agents is above the critical threshold. Additionally, Figure B.3 shows the cycles of trust and distrust, where the population of impersonal agents rises and follows as discussed in the previous paragraph.

2. **Emphasis on Measurement:** Implementing systems like the double-entry bookkeeping increased emphasis on keeping records and measuring transactions (McCloskey (2016), Chapter 29 and 34), which was especially true after the adoption of the Hindu-Arabic number system in the late fifteenth century Europe (Durham 1992).

3. **Increasing formalization:** Trade before the emergence of impersonal markets occurred through guilds that were informal organizations (Kieser 1989). Adoption of routines such as the double-entry bookkeeping after the late fifteenth century helped in the formalization of business.

4. **Hurdles in bending formal rules:** Traders who adopted double-entry bookkeeping could conduct fraud nonetheless, yet, the existence of an overarching system of record keeping process meant that a fraudulent trader was met with a greater procedural complexity as it involved altering many "regular and predictable" formal organizational routines. This is much like a businessman of an ISO certified firm nonetheless producing a lower quality product, by not following the prescribed procedures.

These features of recordkeeping and adoption of accounting rules imply that the decision of implementing double-entry bookkeeping introduced organizational routines that were sticky. So, a trader who had implemented double-entry bookkeeping but also wanted to commit fraud towards selective customers (which was common in the sixteenth century period, and something the polymorphic impersonal equilibrium will allow), could not switch between the two modes of doing business seamlessly (implying w is small). So, double-entry bookkeeping could reduce partiality in doing business by embedding standardized routines that induced (most) traders to conduct a standards-based business that maintained records.

Corollary 2 thus implies that two economies (Atlantic coastal regions: Spain and Portugal vs. Low countries and England) with similar initial conditions can have divergent outcomes, because of the ability or disability of a critical fraction of traders to adopt standardized routines supporting impersonal trade. This could be due to higher levels of information diffusion in one region (Northwestern Europe) over another (like Spain and Portugal) due to better information technology (like printing press), that popularizes better business practices (like double-entry bookkeeping).

In the past and elsewhere for much later, technologies that could mass diffuse business practices did not exist. This limited the possibility of a sustained emergence of impersonal exchange from relationship-based exchange.

2.2.5 Identity, parochialism and honesty

The model has assumed that social groups are meshes of relationships that are sources of steady income (m), and agents make a choice between earning this steady income or conducting a risky exchange with strangers. But, groups are also sources of social identity (Akerlof & Kranton 2000) and leisure. Chapter four of the thesis will discuss the significance of social identity in greater detail, describing how social and state-sanctioned prescriptions about social categories, can induce conformity to social prescriptions, and deviation from individual tastes.

In the model economy, traders who conduct or are interested in conducting honest exchanges with strangers may be perceived by their existing relationships and groups to be less committed towards them (for not exhibiting prescriptions of in-group favoritism (Balliet, Wu & De Dreu 2014)) and more cosmopolitan. So, such impartial and cosmopolitan traders may risk being excluded from specific group benefits by identity-based traders. One of the ways in which these traders - who are honest to strangers, and form bridges with outsiders - may be punished is through character assassination (Burt & Luo 2017).

A historical example of such exclusion in parts of India was the notion of *kala paani* or "black waters". Traveling to foreign lands across seas was feared by members of some *Jatis* and communities, as they feared that travellers might deviate from prescribed community norms such as about diet and lifestyle and get ritually polluted. So, reentry into the community required ritual purification, if not excommunication. In the model economy, this implies that the traders who are identity-based and partial towards familiar contacts and stick to existing mesh of relationships, may have access to extra benefits i and earn a utility $m + i$ as a steady income, higher than the steady income m earned by honest relationship-based traders who are honest to strangers and willing to conduct impersonal exchanges (and hence risk ritual pollution).

In the model economy, if i increases, it increases the population of dishonest traders, because dishonest traders are earning extra utility i not enjoyed by honest relationship-based traders[24]. This implies the following corollary:

Corollary 3 *Increasing i, i.e., increasing identity benefits available exclusively to identity-based and partial (D-type) relationship-based traders, reduces the scope in which impersonal exchange can emerge.*

2.3 Endogenous Emergence of Impersonal Institutions

Now that we have explored the micro-foundation of honest exchange in impersonal setting, we can consider how powerful actors (rulers and elites) would react to it. We saw how relationship-based equilibrium is a persistent equilibrium. So, conditions before the transition to an impersonal equilibrium can be considered to be a stasis. In the absence of the possible conditions supporting a transition out of the stasis, would rulers/governments support impersonal and impartial institutions like impersonal marketplaces, networking venues, and impartial courts? Such open marketplaces and institutions were developed in Northwestern European cities like Amsterdam, Antwerp (Gelderblom 2013) and Hamburg (Dollinger 1970); while in cities like Lübeck opening of markets faced significant resistance from incumbents (Dollinger 1970, Lindberg 2009). Can the model impersonal economy help us explain this divergence?

I consider the specific setting of medieval and early modern Europe and introduce a ruler and merchant guild elites as actors in the model economy. These actors (rulers and elites) could alter the value of payoffs (m, h), observing population's initial levels of adoption of standardized routines $(p_H(0))$, and the payoff available in impersonal exchange (c). The ruler was interested in

[24]When $i = 0$, in the neighborhood of stable equilibrium where $p_D = \frac{k_o h}{k_s(c-m)-k_o m}$ the relative benefit of being an honest trader is increasing in p_D. When i increases the relative benefit reduces as π_D increases while π_H stays constant. This shifts the relative benefit curve in the neighborhood rightwards, and the equilibrium p_D increases with increasing i.

maximizing its tax revenue, while the guild elites were interested in maintaining their monopoly, granted to them by the ruler in exchange for revenue.

A ruler of a relationship-based system had a choice of removing monopolies of merchant guilds and establishing institutions that supported impersonal markets, or of maintaining them. If the ruler maintained merchant guild monopolies, then if at all any impersonal trade occurred, it occurred in informal (unsanctioned) markets. In impersonal markets sanctioned by ruler the ruler will improve law and order (reduce h), while in an unsanctioned informal market the ruler would attempt to make it worse by punishing those who conduct such impersonal exchanges despite the monopoly restriction (increase h). So, loss from cheating in the informal markets (h^i) was higher than the loss in ruler sanctioned impersonal markets ($h^i = \beta h > h$).

At the same time, the merchant guild elites were the beneficiaries of the relationship-based system, as merchant guilds were at the center of commerce in the relationship-based economy. Merchant guilds were the intermediaries/brokers in many business exchanges, and elites could endogenously influence the payoff (m) from the relationship-based channels. If brokerage charges of the merchant guilds controlled by elites were expensive, the payoff from relationship-based channel was lower. I assume elites could set relationship-based payoff within a range $m \in [m^{min}, m^{max}]$. Under what observed conditions (impersonal payoff (c) and initial adoption (p_H^{cr})) would ruler choose to remove monopolies, given that merchant guild elites would alter relationship-based payoff (m) to block such a transition?

To remove the monopolies granted to merchant guilds, the ruler must have anticipated that impersonal trade beyond relationship-based networks would be feasible. As impersonal trade was only supported under special utility conditions (Corollary 1) and in the presence of horizontal communicators (Corollary 2), the satisfaction of the two conditions was a necessary requirement for the ruler to remove monopoly privileges of merchant guilds. In the absence either of the two conditions, 1) all traders preferred relationship-based exchange, which merchant guilds dominated, and 2) ruler and merchant guilds elites continued their mutually beneficial nexus, where a) the ruler provided monopolies to the merchant guilds while in return b) the merchant guild elites provided ruler with revenue.

Merchant guilds elites would alter relationship-based payoff (m), to block the possibility of transition to impersonal equilibrium in which merchant guilds lost all power. Also, the merchant guild elites would prefer to keep a low relationship-based payoff for others, so that they did not have to part away with their share of wealth. How much payoff would guild elites set to block the transition to impersonal equilibrium?

When payoffs are endogenized, Proposition 2 emerges (see Proof in appendix C.2).

Proposition 2 *Choices of ruler and merchant guild elites are:*

- *A necessary condition for ruler to remove merchant guild monopolies is when merchant guild elites have increased relationship-based payoff to the maximum ($m = m^{max}$) and if $p_H(0) > \frac{(1+\lambda)(m^{max}+h)}{c+\lambda m^{max}+(1+\lambda)h}$ and $c > \frac{(1+\lambda)}{\lambda}h + m^{max}$. Otherwise ruler maintains monopoly.*

- *If ruler maintains monopoly, merchant guild elites set minimum relationship-based payoff $m = m^{min}$ if $p_H(0) \leq \frac{(1+\lambda)(m^{min}+\beta h)}{c+\lambda m^{min}+(1+\lambda)\beta h}$ or $c \leq \frac{(1+\lambda)\beta}{\lambda}h + m^{min}$. Otherwise merchant*

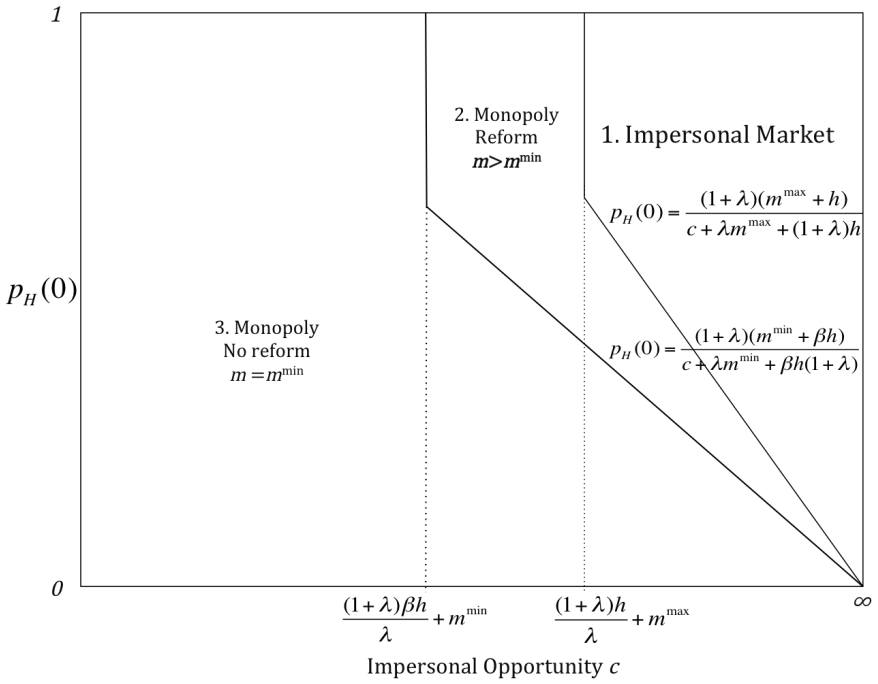

Figure 2.5: The figure describes conditions in which different systems of exchange emerge. The X-axis represents the level of impersonal opportunity c, while the Y-axis represents the level of initial adoption of standardized routines of impartial conduct $p_H(0)$.

guild elites set $m \in (m^{min}, m^{max})$.

Figure 2.5 represents the different conditions. Proposition 2 and Figure 2.5 implies the following.

- **Monopoly and no reform:** If, the number of traders who initially adopted standardized routines of impartial conduct, was small (p_H^m is small), then unless impersonal exchange (c) was exceptionally lucrative, traders were unable to trigger a transition to impersonal equilibrium. So, merchant guilds could set relationship-based payoffs (m) to the minimum (*exploit* by setting $m = m^{min}$ in Region 3 of Figure 2.5), as regardless of the payoff traders would continue to rely on relationship-based channels dominated by merchant guilds.

- **Monopoly and reform:** If payoff from impersonal exchange was high, and there was a large population of initial adopters of standardized routines (p_H^m is high), conditions could be favourable for transition even if ruler had not removed monopolies, as trade could occur in suitable informal settings[25]. Under the above conditions elites would attempt to make impersonal exchange relatively less lucrative, by raising the payoffs from relationship-based exchange (*reform traditional institutions* by setting $m > m^{min}$ in Region 2 of Figure 2.5) such that conditions became unfavorable to transition once again. But, there were limits to how much elites could raise the relationship-based payoffs.

[25]It could be that informal markets were risky and loss from being cheated h^i was high, in which case impersonal exchange is not feasible without ruler removing monopoly.

- **Impersonal market:** If (i) impersonal exchange became highly lucrative, and (ii) there was large population of initial adoptees of standardized routines, such that the elites did not have the resources to reform and raise relationship-based payoff high enough, then the relationship-based institutions could be said to have become inefficient, and the economy could then transition to the equilibrium of impersonal exchange (Impersonal market in Region 1 of Figure 2.5).

In the sixteenth century Europe, closeness to the sea in cities like Antwerp, Amsterdam, and Hamburg was favorable for long-distance trade, which made impersonal exchange lucrative, and so closeness to sea could proxy for higher payoffs from impersonal exchange (c). Additionally, the booming long-distance trade of the sixteenth century at the Atlantic coast could be the disruptive opportunity for such cities, that rendered guilds inefficient for reforms. In contrast, a city like Lübeck did not enjoy the disruptive opportunity of the Atlantic coast and had powerful local elites who profited from the Baltic trade, who were willing to block any transition to impersonal markets (Dollinger 1970, Lindberg 2009). Similarly, penetration of printed books increased horizontal transmission of new norms, ideas and standardized business practices like double-entry bookkeeping in cities like Antwerp and Amsterdam, and high printing penetration could proxy for high initial adoption of standardized routines (p_H^{cr})[26]. In such cities, there was scope for impersonal exchange between "footloose" traders (Gelderblom 2013), and rulers began to attract them by establishing more impersonal institutions, often ignoring the requests of merchant guild elites for new monopolies. The next chapter will discuss these dynamics in greater detail.

The insight provided in Figure 2.5 is not limited to medieval and early modern Europe alone, and can be abstracted to modern economies as well, where economies trapped in punctuated equilibrium of relationship-based exchange could also have exploitative elites/brokers (like *Mafia*), in nexus with politicians. It's only when suitable conditions exist that enable sustained impersonal exchange to occur (presence of favorable utility conditions and horizontal communicators), that politicians improve impersonal and impartial institutions, and stop favoring elite brokers who lose dominance in such an impersonal environment[27].

2.4 Conclusion: Persistence and Change

In this chapter, I developed a theory of diffusion and sustenance of trust between strangers and attempted to answer the following questions. Firstly, under what conditions do the norms that

[26] In the next chapter I plot in Figure 3.9 the 50 largest cities of fourteenth, fifteenth and sixteenth-century Europe (81 in total). On the X-axis, a city's closeness to sea (which is measured as the inverse distance from the sea (Opp_a)) is plotted. On the Y-axis Printing Penetration in the fifteenth century (which is measured as ($PrintPentr_a$)) is plotted. The cities had varying types of economic institutions in the sixteenth century. Impersonal cities were those where rulers no longer were granting guild monopolies. Cities undergoing reform were those where guilds were giving concessions to locals. Relationship-based cities were those where guilds or other feudal arrangements continued to hold dominance. One observes in appendix Figure 3.9, as expected in Proposition 2 and Figure 2.5, that impersonal markets emerged in places close to the sea and with high printing penetration.

[27] A relevant paper considering nexus between economic elites and politicians is Rajan & Zingales (2003) that discussed the political economy of financial development, and provided a mechanism for opening of financial markets where financial and business elites resisted such development as they would lose their dominance.

enable impersonal exchange diffuse? I found that, when the impersonal opportunities are lucrative (utility condition in Corollary 1), a critical threshold of initial adopters follow standardized routines that reduce partiality in conduct (popularized by a horizontal communicator in corollary 2), and identity specific benefits for partial traders are not too large (Corollary 3) then the norms that enable impersonal exchange can diffuse.

Secondly, what stops cheating from flourishing in trusting environments? I found that cheating can be contained in trusting impersonal environments, even without private, third party, community or social network-based monitoring and enforcement. This happens if the honest impersonal traders turn skeptical and distrusting of strangers, and become relationship-based on being cheated. As honest impersonal traders turn relationship-based, cheating becomes less lucrative. With more relationship-based traders, the demand for guarantees increases, making being honest and impersonal lucrative once again. So, contagions of skepticism and optimism - that reflect the general *reputation of a stranger* - can keep cheating in a trusting environment under control.

In the literature related to trust among strangers, papers that consider the evolution of generalized trust (Tabellini 2008b, Greif & Tabellini 2017) consider traders to be gaining special utility ("warm glow","city orientation") from honest conduct. In the theory I develop, traders gain no special utility from being honest or trusting. Instead of prosocial preferences, in the model, honest traders should be cautious and prudent ($\lambda > 0$) in their market exchanges. When λ increases the number of trusting traders reduce, but, there is a broader range of utility conditions (lower c) for which transition is possible. By not giving prosocial preferences to traders, the evolution of generalized trust does not depend on changing intrinsic preferences, i.e., individual traders across different equilibria have same intrinsic preferences.

Also, in the model trust is endogenously generated, in contrast to Guiso, Sapienza & Zingales (2008a) model of transmission of trust between generations, and honest and dishonest traders coexist in a polymorphic equilibrium. This polymorphic coexistence is dependent on the degree of stickiness of norms, where if norms were revised very frequently ($w \geq \bar{w}$) then this equilibrium would not exist. Like in the models by Tabellini (2008b), Greif & Tabellini (2017) and Guiso, Sapienza & Zingales (2008a), the presence of strong traditional relationship-based institutions, which make the need for impersonal exchange less lucrative, makes a transition to impersonal equilibrium less likely. When considering the evolution of cooperation, in contrast to Sethi & Somanathan (1996) where cooperation is sustained through punishment, or to Ghosh & Ray (1996) and Nowak & Sigmund (2005) where it is sustained by repeated exchange, here trust is sustained by temporary withdrawal from trusting strangers when cheaters begin to dominate the system. The mechanism is in the spirit of the voluntary public goods game literature by Hauert, De Monte, Hofbauer & Sigmund (2002) and Semmann, Krambeck & Milinski (2003), and to model by Macy & Skvoretz (1998) where agents maintain trust by having an exit option. But, compared to Macy & Skvoretz (1998) in this model, the decision to engage in a prisoner's dilemma is dependent on the general trust traders have on strangers (reputation of a stranger), based on their experiences in the past.

So, trust developed in the model is not an outcome of intrinsic or educated prosociality, nor of private, third party, community or social network-based regulation. Here honest conduct

competes with dishonest conduct, and the equilibria are an outcome of the competition. Also, because of the set up of the model, it can capture the *intragenerational* transmission of norms, ideas, and routines within a generation, which is as important to cultural evolution as diffusion (Mokyr 2016) between generations.

Moreover, in this chapter, I developed an endogenous model of the emergence of impersonal institutions. I found that, as long as the elites of relationship-based institutions were efficient, they could block the emergence of impersonal exchange by reforming themselves. So, under such conditions, rulers would prefer to be in a nexus with relationship-based elites, and the economy will have a relationship-based system of exchange with inadequate provision of impersonal and impartial (generalized) institutions. But, if elites started losing dominance, because (i) impersonal exchange became highly lucrative, and (ii) standardized routines inducing impartial conduct became popular, the ruler could promote impersonal and impartial institutions by opening markets, impartial court systems and improving law and order.

The model I develop relaxes assumptions such as a preference for prosociality, yet a limitation of the model is that it assumes an anonymous nature of impersonal trade. In reality, social interactions are indeed expected to be repeated, although it is reasonable to assume that there are only a limited number of interactions that two strangers will have with each other. If interactions are not one shot, but limited, the potential benefit c from impersonal exchange can be considered to proxy the expected benefit from repeated interactions between two strangers. So, the model I develop complements the existing work on trust and cooperation (Macy & Skvoretz 1998, Tabellini 2008b, Greif & Tabellini 2017, Guiso, Sapienza & Zingales 2008a, Sethi & Somanathan 1996, Ghosh & Ray 1996, Nowak & Sigmund 2005), where relaxation of some critical assumptions about vulnerability (Mayer, Davis & Schoorman 1995, Rousseau, Sitkin, Burt & Camerer 1998) illuminates some previously unexplored facets about the evolution of generalized trust and honesty.

Impersonal exchanges help broaden business exchanges and hinge on trust that others will fulfill promises. If people's attitudes and sources of information are tightly controlled, the mass diffusion of new norms supporting impersonal exchanges may be less feasible, irrespective of the quality of outside opportunities or law and order. In the presence of horizontal communicators that help people mass adopt new norms, ideas, and routines of impersonal honest conduct, change and reform are possible, where old less efficient institutions are dismantled, and new impersonal ones are created. The historical evidence suggests that mere existence of right utility conditions (like lucrative opportunities from trade through the Atlantic coast) is not enough for impersonal exchange to emerge and sustain. There is also a need to support institutions that can act as horizontal communicators of routines/norms/practices that promote impartiality in social conduct. So, media (e.g., fifteenth-sixteenth century printing press (Eisenstein 1980, Dittmar 2011)), religious institutions (e.g., Bishops in twelfth-century North Italy (Guiso, Sapienza & Zingales 2008b)), opinion leaders (e.g., Muhammad in the seventh-century (Montgomery 1961)), etc. all become critical players in building and sustaining impersonal economies built on trust.

Also, the world today is once again undergoing an information revolution. Such a powerful technology that helps in the mass diffusion of new norms, ideas, and routines (like blockchains)

is not enough for the emergence of impersonal exchanges. Just like the adoption of ISO systems, or accounting standards are no panacea for deceptive behaviour in the market, the same way adoption a modern ledger like blockchain may not suppress cheating behaviour in the market unless such adoption of trust enhancing practices is supported by the right economic and institutional (utility) conditions. The central question that we should consider with regards to new internet technologies is, do the economic conditions provide businesspeople the right incentives to have uniform standards of behaviour, or are businesspeople incentivized to devise ways to target and cheat vulnerable partners selectively? When the economy provides little incentives to conduct business with strangers, these technologies may merely become a front for deceptive behaviour. In contrast, when businesses are incentivized to cast a wide net to attract strangers, they are also incentivized to invest in standard practices of ethical conduct.

Chapter 3

How merchant guild networks declined in Northwestern Europe?

> [In the sixteenth century,] the guild-dominated societies of Italy, Iberia, and German-speaking central Europe were unable to adjust to rapid institutional, commercial and demographic changes of the sixteenth century. They lost out to the market-oriented civic culture of the Low Countries and England with their increasingly impersonal markets and impartial states, which encouraged forms of generalized trust that favoured adaptation and growth. The weakening of the particularized trust generated by associative institutions such as merchant guilds created interstices in English and Dutch society within which people could experiment with generalized trust in strangers mediated by impersonal markets and impartial states. This cannot be regarded as an accident. - Ogilvie (2011)

In the previous chapter, I developed a theoretical model, to exhibit how impersonal economies do not emerge spontaneously, and that embedded relationship-based institutions are resistant to change in a wide range of condition. As in tightly knit networks, the sources of information and opportunity may be limited (Burt 1995) it may limit the growth of commerce[1]. In this chapter, I explore how impersonal exchange emerged in the deeply embedded economies of medieval and early modern Europe which were organized around the multiplex ties of the merchant guild networks (Greif, Milgrom & Weingast 1994, Ogilvie 2011).

Understanding why do business organizations and networks look the way they do remains an open puzzle (Stuart & Sorenson 2007). The period of the late fifteenth and sixteenth century is appropriate for answering the puzzle as it was a period of significant changes in the socioeconomic organization of Europe. In the late fifteenth and sixteenth century the Iberian

[1]For example in India, a lack of trust in outsiders may limit the growth of family firms, as they grow proportional to the number of male family members (Bloom et al. 2013). Across Europe, lower regional trust is associated with lower external sourcing, which leads to poorer firm performance (Vanneste & Gulati 2018). In Italian regions that have lower social capital, households may make fewer investments in stocks or receive less credit (Guiso, Sapienza & Zingales 2004), and firms may do less innovation (Laursen, Masciarelli & Prencipe 2012). In New York, embeddedness may increase the survival chances of apparel firms, but only until a threshold, and the highly embedded firms may have a low chance of survival (Uzzi 1996). Similarly, in the Silicon Valley, a loosely linked team structure of firms may have encouraged communication across divisions and organizations, giving them a competitive advantage over the "autarkic" corporations of the Route 128 region, that had rigid boundaries and hierarchies (Saxenian 1996).

Peninsula pioneered the discovery of new sea routes to Asia and the Americas; the Germanic regions developed a printing culture that influenced knowledge production, religion, and culture; and the Renaissance ideas of Italy diffused elsewhere promoting rationalism and scientific inquiry. But of all the places, it was the Northwest European region of the Low Countries and England that underwent transformational change. In the region merchant guild monopolies declined (Ogilvie 2011) and a bourgeois culture emerged (McCloskey (2016), Chapter 54), which respected both profit and innovation (Mokyr (2016), p. 17) and conducted trade with relative strangers (Gelderblom 2004). Cities like Antwerp, Amsterdam, and London became the centers of institutional and business innovation that facilitated impersonal exchange (Gelderblom 2013, Puttevils 2015) and these innovations have gone onto influence the modern world[2]. The sixteenth-century transformation was followed by the seventeenth-century Dutch Golden Age and the eighteenth-century English Industrial Revolution. What made the Northwestern Region of Europe so different? The question remains a central concern in the social sciences, with scholars from diverse fields (Weber 1905, North 1990, Harris 2009, Padgett & Powell 2012, McCloskey 2016, Mokyr 2016, Rubin 2017, Jones 2017) researching the subject.

Several institutional and cultural factors helped in facilitating impersonal exchange in the Northwest European region. The *inclusive political institutions* of the region limited coercive powers of the rulers (North (1990), p. 130; Acemoglu, Johnson & Robinson (2005)), and the *impartial market-based institutions* of its cities began to serve the interests of all businessmen instead of a particular few (Ogilvie (2011), p. 33; Gelderblom (2013)). The *bourgeois intellectual culture* of the region promoted discovery, entrepreneurship and innovation (McCloskey (2016), Chapter 58; Mokyr (2016), p. 17), and due to its cosmopolitan *business friendly popular culture*, profit making ceased to be a taboo (McCloskey (2016), Chapter 58).

Why did the Northwest European region develop these unique institutional and cultural characteristics? Until the end of the fifteenth century, impartial institutions like courts and the police that serve all parties generally - so ubiquitous today in the developed world - were not well developed in Europe (Ogilvie (2011), p. 187). In such a world without impartial institutions, trade often was (is) relationship-based, customary (Kadens 2011), and conducted through networks like guilds (Greif, Milgrom & Weingast 1994). A relationship-based trade through networks of high closure reduced the concerns of information access and reliability (Uzzi & Lancaster 2003, Burt 2001). Merchant guilds were non-kinship-based multiplex networks that dominated the European trade for a significant portion of the second-millennium (Ogilvie 2011). But, in the sixteenth century, the merchant guild system began to lose its significance as more impersonal markets, where traders could directly trade without the need of a guild affiliation, began to emerge and the local authorities stopped granting monopoly privileges to merchant guilds. The traders began to rely less on the networked and collective institutions like merchant guilds, and directly initiated partnerships with traders whom they may not have known well. For example, in Antwerp, the domination of intermediaries (called hostellers), who would connect the foreign traders, declined. Instead, the foreign traders began to conduct such trades directly with each

[2]For example, one of the first permanent commodity bourses was established in Antwerp in 1531, the first modern stock exchange emerged in Amsterdam in 1602, and the joint stock companies that pooled capital from a large base of investors became a promising form of organizing business in London and Amsterdam by the end of the sixteenth century and early seventeenth century.

other in facilities like bourses (Gelderblom (2013), p. 58).

In this chapter, I study the emergence of impersonal markets and the decline of merchant guilds in Europe during the sixteenth century. I survey the fifty largest European cities during the fourteenth, fifteenth and sixteenth century and codify the nature of the sixteenth-century economic institutions in each of the cities. In the survey, I find that the merchant guilds were declining in the Northwestern region of Europe. Elsewhere in Europe, guilds continued to dominate commerce until much later[3], although there were some reforms underway in the Milanese and Venetian regions of Italy. What explains the observed pattern of the emergence of impersonal markets in Europe, in the sixteenth century? Any explanation that I develop should explain three trends.

1. Firstly, why did the decline of merchant guilds only occur in the Northwestern region.
2. Secondly, why did this decline only occur in the sixteenth century.
3. Thirdly, why did other parts of Europe not benefit from the same benefits that were transforming Northwestern Europe.

I focus on the interaction between the *commercial and communication revolutions* of the late fifteenth century Europe and argue that the Northwest European region uniquely benefited from both the revolutions, because of its unique geography. I also collect data on fifteenth and sixteenth-century factors like the level of urban clustering, the prevalence of medieval fairs, and level of population growth.

3.1 The Great Enrichment of Northwestern Europe

3.1.1 The commercial revolution and the value of strangers

What can motivate a trader to seek risky opportunities beyond its close networks? Consider two agents A and B in Figure 2.1, whose networks do not overlap. For agent A to be interested in engaging in a risky impersonal exchange with B, the utility from such a direct exchange should be high relative to the other possible exchanges through relationships[4]. Traders in cities that were geographically suitable for long-distance trade, like entrepôts, would find impersonal trade opportunities attractive, as these cities would attract unfamiliar traders who would come to the city looking for business opportunity. Entrepôts across Eurasia, from Antwerp to Venice to Aden to Malacca, have been cosmopolitan hubs with diasporic communities of foreign merchants (Harris 2009). The Atlantic coast was undergoing a commercial revolution with the discovery of new sea routes to Asia and the Americas. Cities at the Atlantic coast became attractive commercial centers enjoying high population growth, including in Northwestern Europe that also conducted trade with the Baltic region. I find that all the cities where guild privileges

[3]Outside of Europe, business in India, China, and Arabian Peninsula continue to be based on kinship-based networks of *Jati* (caste), *guanxi* (clan relationships) and *qabila* (tribe).

[4]Such an impersonal opportunity can be viewed as a form of structural-hole opportunity (Burt 1995) between two unconnected agents. In long-distance trade, the trusted network of traders resembles the multiplex ties of a merchant guild. If a trader A found a direct partnership with unfamiliar traders (like B) beyond the guild networks to be highly beneficial, that would motivate A to engage in an impersonal exchange with B instead of exchange through guilds.

declined were at the sea and along the Atlantic and the North Sea coast. Moreover, all cities that underwent reform (but did not decline) were within 150km from the sea port, with their average distance being 67km. The average sea port distance of cities that did not undergo reform was almost double at 125km.

While commercial cities close to the sea were natural meeting points for traders with rich impersonal opportunities, not all commercial cities close to the sea developed extensive systems of impersonal exchange[5]. Long-distance trade would involve trade in regions and markets with limited information, and with partners/agents who could not be easily monitored. So, informational asymmetry and moral hazard made such impersonal long-distance trade difficult, and guild system being a socially embedded system of exchange was suitable for mitigating the associated risks[6]. So, for impersonal exchange to emerge, search costs needed to be low and traders needed to feel confident about the reliability of risky impersonal partnerships. Improvements in the availability of information could help in increasing such reliability.

3.1.2 The communication revolution and the value of information

Improvements in the information technology can reduce search costs and increase standardization which can reduce high coordination costs of impersonal market-based (and non-hierarchical) exchange (Yates 1986, Malone, Yates & Benjamin 1987)[7]. In the sixteenth century, the postal system improved across Europe (Figure E.2). The postal system made communication between distant traders easier, helping in long-distance trade across Europe. The Northwest European region did not have a particular advantage over other regions in postal communication. But, it had an advantage in early diffusion of printed books. The Northwest European region was close to Mainz, the city where Johannes Gutenberg invented the movable time printing press in the mid-fifteenth century[8]. Dittmar (2011) showed how the cities closer to Mainz adopted printing sooner than many other regions of Europe in the first few decades of its introduction. Compared to previous methods of information diffusion, printing press was a disruptive technology that enabled horizontal diffusion of information beyond vertical transmission from parents, teachers and other authority figures. So, trade-related books and new (or unknown) business practices like double-entry bookkeeping diffused early and rapidly in the region (Puttevils 2015). Most saliently, printing accelerated the adoption of Hindu-Arabic numeral system, which helped in the rise and popularization of more sophisticated commercial, accounting, and

[5]If merely the opportunity for impersonal exchange was enough for impersonal exchange, then the major ports of the Indian Ocean should have also evolved the kind of trade that the Northwest European region developed in the sixteenth century (Harris 2009). At the Atlantic coast, among eight major Early Modern port cities of Antwerp, Amsterdam, Bordeaux, Hamburg, Lisbon, London, Rouen, and Seville, in only four cities (Antwerp, Amsterdam, Hamburg, and London) merchant guilds declined. To understand the reasons behind the divergence, it is important to appreciate the risky nature of impersonal exchange and the effectiveness of merchant guilds as a solution.

[6]Embedded businesses create more trust and enable rich information transfer between partners (Uzzi 1996, Uzzi & Lancaster 2003).

[7]Such a reduction in coordination costs, like during the *expansion of telegraph and railroads in the nineteenth century, led to the rise of commodities markets in the United States* (Yates 1986). So, if the information technology improved in the sixteenth century, it could increase the confidence traders had in unfamiliar partnerships and could reduce their reliance on the merchant guilds (or intermediaries), by improving their trade-related information and by standardizing their and their potential partners' business practices.

[8]Harlem resident Laurens Janszoon Coster is credited to have simultaneously invented the printing press with Gutenberg, even though it was the printing press invented by Gutenberg in Mainz, that got successfully adopted elsewhere in Europe.

financial techniques (Durham 1992, Chatfield & Vangermeersch 2014).

A high penetration of the printed material in Europe reduced information barriers erected by "guild mysteries" (Ogilvie (2011), p. 378)[9], and improved business practices. I find that all cities where guild monopoly privileges declined or that underwent reform in the sixteenth century, enjoyed high penetration of printed material in the fifteenth century. Among cities within a 150km distance from the sea, the cities where guilds declined or reformed had more than twice the number of diffused books per capita than cities where guilds continued to dominate. A city at the Atlantic coast with about four times more per capita printing penetration in the fifteenth century was one degree higher in the level of market impersonalization[10] in the sixteenth century[11]. Moreover, I find that in the Northwest European cities like London and Antwerp (where guilds declined) there was production of many books related to Economics and related subjects, which were being written in the local language. So, commercial centers of the Northwest European region did not just benefit from the diffusion of commercial, mathematical and financial techniques, but they also created a local demand for it.

The combination of both the commercial revolution along the sea coast, especially the Atlantic coast, and the communication revolution, especially near Mainz, uniquely benefited the Northwestern Europe (see Region 1 of the map in Figure 3.1). It began to attract traders who favored an impersonal market-based exchange over exchange conducted through guild networks. Local authorities began to disfavor privileged monopolies (Ogilvie (2011), p. 187) when impersonal exchange became more feasible and they started to provide facilities for impersonal exchange. In the region, trade democratized (Puttevils 2015), as more people could participate in business, and such participation eventually led to the rise of the first joint stock companies like the Dutch East India Company (Gelderblom 2004).

Regions like Spain and Portugal that benefited only from the commercial revolution of trade through the sea to Asia and the Americas had low levels of printing penetration in the fifteenth century, in part because they were distant from Mainz and thus were late in adopting printing. In contrast, inland regions of Germany and France benefited from the print revolution due to their central location and closeness to Mainz, but they did not enjoy a bustling Atlantic coast that attracted traders. Italy partially enjoyed the benefits of both- a dynamic commercial coast, and high printing penetration. But, the commercial revolution at the Atlantic coast did not disruptively improve prospects of impersonal trade for Italy (see Region 2 of the map in Figure 3.1). One reason for the failure of the Italian cities to tap opportunities through the Atlantic was the region's historical investments in Mediterranean trade and land trade with Asia. For example, Venice enjoyed a monopoly over trade with Asia, especially in the spice trade (Tracy (1993), pp. 26-28). In fact, one of the reasons for the interest of the Portuguese in an alternate

[9]Publication of such guild mysteries was "regarded as a violation of the censorship of guild mysteries' vis-a'-vis non-members... As late as 1558, when the German merchant Lorenz Meder published his merchant manual, he clearly acknowledged in the preface that he was breaking the taboo against making public strictly protected merchant secrets'." (Ogilvie (2011), p. 378)

[10]Market impersonalization: 0=Relationship-based, 1=Undergoing Reform, 2=Impersonal (Footnote 33).

[11]*Atlantic coast and Printing:* There were four impersonal cities- Hamburg, London, Antwerp and Amsterdam, while there were four relationship-based cities- Lisbon, Seville, Rouen and Bordeaux as the major Atlantic ports in the sixteenth century. The fifteenth century per capita printing penetration of these cities would stack as- Lisbon (0.81 books per 1000 population (bptp)), Bordeaux (1.44 bptp), *Hamburg* (3.69 bptp), Seville (4.57 bptp), Rouen (8.11 bptp), *London* (9.33 bptp), *Amsterdam* (14.29 bptp) and *Antwerp* (24.92 bptp).

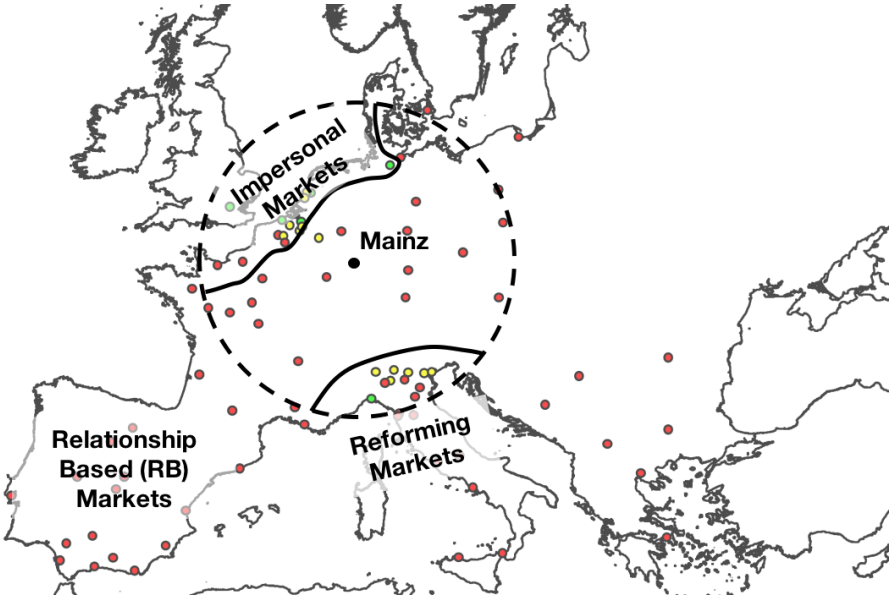

Figure 3.1: Map of 50 largest cities of Europe in the fourteenth, fifteenth and sixteenth-century (Bairoch, Batou & Chevre 1988) based on the type of institutions. The figure describes the different conditions in the different regions in Europe. The Region 1 in Northwestern Europe is the area that is closer to Mainz and also close to the Atlantic Ports. So, it was uniquely situated at the heart of the commercial and communication revolutions. In Region 1, all the cities with emerging impersonal markets were found. No other region won the double lottery of high printing penetration and trade at the Atlantic coast like the Low Countries and England and winning of both the lotteries, I argue, was necessary for the transition to happen. The Region 2 in North Italy is the area that is closer to Mainz and also close to the sea. In Region 2, the cities undergoing reform were found where elites reformed to ensure impersonal markets would not develop. Rest of Europe (Region 3) is made up of Relationship-based cities.The Label- Green: I1 & I2 (impersonal), Yellow: U1, U2 & U3 (undergoing reforming), Red: R1, R2 & R3 (relationship-based). Table E.4 in appendix E.2 details the nature of economic institutions in each city based on historical sources. See section 3.4.2 for more details.

sea route to Asia (that led to the discovery of the Americas) was the domination of Venice over land-trade with Asia. Existing maritime cities of North Italy and Venice in specific were so invested and locked in the more traditional patterns of trade, that they didn't participate in trade through the newly discovered Atlantic routes. The Baltic entrepôt of Lübeck provided a similar case of being locked in the historical trade with the Baltic region (Dollinger 1970), which in the long run resulted in Lübeck losing out to its Atlantic neighbour Hamburg in trade through the Atlantic (see case of the two cities in appendix D).

The above examples of the Iberian Peninsula, Italy, and North Germany show that regions other than Northwestern Europe did not enjoy the unique combination of the full benefits of both the commercial and communication revolutions.

3.2 Guilds and Markets of Europe

The craft and merchant guild system emerged in Europe, beginning around the twelfth century. Guilds were dense non-kinship-based networks of organized and specialized artisans and traders. Craft guilds engaged in secondary activities like manufacturing, while merchant guilds engaged in tertiary retail and wholesale trade[12].

[12]While merchant guilds began to decline in the sixteenth century in Northwestern Europe (Gelderblom & Grafe 2010, Ogilvie 2011), craft guilds began to decline in the later centuries (Epstein 2008). So, the sixteenth century

3.2.1 Why are embedded institutions like merchant guilds so powerful?

Management literature has reported several benefits of embeddedness to businesses (Coleman 1988, Uzzi 1996, Gulati 1998, Fernandez, Castilla & Moore 2000, Kilduff & Brass 2010). Merchant guilds also used to bring several benefits to the merchants in an era that was lacking in formal legal institutions. Merchant guilds helped in the close monitoring of partners through repeated interaction, and by the collective assertion of their rights (Greif, Milgrom & Weingast 1994, Gelderblom & Grafe 2010). A wide range of relationship-based mechanisms (private, third party and community-based) could be applied by guild members to reduce cheating by traders.

- Individual merchants could *privately* constrain a partner from cheating using reputation and repetition-based mechanisms.
- Alternatively, a *third party* (like guild court) could identify and punish cheaters by improving contract enforcement (or law and order) (North 1990, Milgrom, North & Weingast 1990).
- Moreover, entire communities could enable impersonal exchange through the *community mechanism* of community responsibility system (Greif 2006), as long as their community identities were known[13].

Guild networks also were valuable conduits of information for guild members, and merchants would use their "guild mysteries" to maintain competitive advantage (Greif, Milgrom & Weingast (1994); Ogilvie (2011), Chapter 9). Merchant guilds also helped in the development of public goods such as warehousing and protective convoys that helped in trade (Gelderblom & Grafe 2010).

Not surprisingly, because the merchant guild system functioned efficiently (for the members) in the absence of reliable formal institutions, it sustained in Europe for several centuries. In developing countries like India lacking in developed formal institutions, networked institutions like caste still play an important role in business. Before the fourteenth-century merchant guilds were probably non-hierarchical, voluntary and inclusive. But, with time merchant guilds started to become exclusive monopolies, placing high barriers of entry for outsiders (Ogilvie (2011), Chapter 5). There are two reasons for the high barriers to entry.

- Firstly, as repeated committed interaction was the key to the effectiveness of guilds (Greif, Milgrom & Weingast 1994), uncommitted newcomers could behave opportunistically and undermine the system. So, newcomers faced restrictions.
- Secondly, the newcomers also threatened the position of existing businessmen by increasing competition. So, even the genuinely committed newcomers could be restricted to enter the guilds as they threatened the domination of existing members.

factors (like the printing press) that could lead to the decline of merchant guilds in the sixteenth century, were not causing the immediate decline of craft guilds as well.

[13] In the community responsibility system, cheating behaviour by one member of a community was followed by sanctions on the entire community by the members of the community of the cheated partner.

An example of anti-competitive restrictions was reported by Gelderblom & Grafe (2010) who classified merchant guilds (in Amsterdam, Antwerp, Bilbao, and Bruges) on the basis of the control they could extend over individual merchants. Their classification showed that up to, and especially in the fifteenth-century, merchant guilds were no longer voluntary institutions, and they began to resemble almost like cartels[14]. Given the influence of merchant guilds, authors such as Ogilvie (2011) have looked at the redistributive role of merchant guilds, and have provided an account of the close relationship (nexus) that developed between merchant guilds and the local authorities. Gelderblom & Grafe (2010) in their analysis argued that the potential to obtain rents through monopoly was one reason why the individual members were interested in delegating more control to the merchant guilds.

Given the dominance of guilds in the market, merchant guilds were also a favored channel for local authorities to raise taxes (Dessí & Ogilvie 2004). But, such exclusionary guild monopolies meant that large fraction of potential traders was left out. European cities, over time, got divided into the prosperous merchant guild members and the non-members (Ogilvie (2011), Chapter 3, 4).

The merchant guild system started to decline from the sixteenth century in the Low Countries and England. The direct reason for the decline was the change in preferences of the local authorities. Local authorities of cities like Amsterdam or London in the sixteenth century would no longer approve requests for monopolies by local or overseas merchant guilds (Ogilvie (2011), p. 187). As monopolies of guilds were removed, trade opened up for other individuals interested in trading. What motivated local authorities to stop favoring merchant guilds remains a puzzle (Ogilvie (2011), p. 187). Moreover, why did the shift happen only in the sixteenth century and especially in the Low Countries and England?

3.2.2 When is impersonal exchange a viable alternative to guilds?

In the sixteenth century, the Low Countries and England started to move towards impersonal and impartial institutions. Traders conducted trade in impersonal markets, which were supported by institutions like courts that served all parties. So, the period around sixteenth century saw two systems of economic organization. First was the prevailing relationship-based system dominated by merchant guilds where business interactions were limited to dense networks. Even when guilds interacted with outsiders, they did so with the members of other guilds, such that if there was any conflict, it was resolved between guilds collectively.

The second form of economic organization was the emerging impersonal system, where the traders sought partnerships with unfamiliar traders beyond their existing networks. *How did impersonal exchanges occur reliably? Could local authorities, by establishing impersonal and impartial institutions like courts, unilaterally develop impersonal markets?* In other words, were the two systems of economic organization just a matter of choice of institutions? Even if a local authority could provide for impersonal and impartial institutions, it would not eliminate the risks and temptations of cheating in impersonal business transactions, and the costs associated with resolution of conflicts. Moreover, because these impersonal exchanges would be a greater

[14] Analysis of Gelderblom & Grafe (2010) showed a clear reduction in the control of guilds over individual merchants after (not during) the sixteenth-century.

concern among traders of different locations, it is not clear if a local authority could enforce effective contracts between them.

Given the risks in partnership with unfamiliar traders, the partnerships would need to fulfill two conditions- Firstly these impersonal partnerships will need to be highly beneficial to motivate traders to go beyond their familiar and reliable networks. Secondly, regardless of the potential benefits of such a partnership, traders would also need to have confidence in entering in exchange with unfamiliar partners[15].

Utility condition

Social structure and networks are receptive to utility gained from networks (Kilduff & Brass 2010). Considering agents A and B in Figure 2.1, when will A be interested in approaching B for a risky arm's length exchange? Ogilvie (2011) (pp. 319, 320) pointed at the tradeoff between benefits and limitations of guild-based exchange and wrote:

> To be effective in detecting and punishing deviant agents, a merchant guild had to be a non-anonymous closed network with multiplex internal relationships. For a merchant network to solve agency problems through internal social capital, its members would have to refrain from entering into agency relationships with outsiders, against whom social sanctions could not be exercised. But closure imposes costs by excluding commercial ties outside the network. Members of the network would limit their agency relationships, forming links only with other network members, even when profitable relationships could be concluded with outsiders.

If the agent A steadily gains $m = \$10$ from a relationship-based exchange (with familiar ties or with B through brokers or guild networks), and gains $\$c$ from a risky impersonal exchange with B (where there is a risk of losing, say, $h = \$5$ if B cheats A), which of the two options does A choose? If c is less than \$10, A has little motivation to enter in an impersonal exchange with B (ceteris paribus). This will also be true even for a utility c that is only slightly higher than \$10 as there is a risk of being cheated. So, for an alternative impersonal system of exchange to emerge, and for a guild like systems to decline, *the utility from the impersonal exchange (c) should be significantly larger than the utility from the relationship-based exchange (m).*

Reliability condition

A large utility difference between impersonal and relationship-based exchange can motivate agent A to seek impersonal exchanges. But, the sustainability of impersonal exchanges also

[15] *Market-oriented and Bourgeois values:* Ogilvie (2005) pointed at the challenges and argued that markets in regions like the Low Countries were supported by a "market oriented civic culture", where traders could trade with each other reliably. McCloskey (2016) looked at cultural attitudes of early modern English texts (although after the 1600s) and discussed how the meaning of the word "honest" changed from being a noble to being truthful (Chapter 25). She wrote "In English our bourgeois word "honest" once meant not mainly "committed to telling the truth" or "paying one's debts" or even "upright in dealing," but mainly "noble, aristocratic," or sometimes "dignified," in a society in which only the noble were truly dignified.". Similarly, Sahle (2015) looked at merchant advice manuals and argued that ethics were an important part of both Quaker and non-Quaker merchant advice manuals (Quakers are a religious group that were credited to have high standards of ethics in business dealings) and emphasized "virtues of honesty, reliability and risk adversity and warn of the same vices, particularly covetousness."

rests on the reliability of B. If Bs are driven by amoral familist norms (Banfield 1967, Putnam, Leonardi & Nanetti 1994) such that they cheat outsiders, then even motivated merchants like A that seek impersonal relationships may eventually stop approaching Bs after experiencing episodes of cheating. So, a sustainable impersonal exchange requires A to trust in B despite the vulnerability of being cheated, and for B to reciprocate A's trust by acting trustworthily. This principle of trust is succinctly described by Mayer, Davis & Schoorman (1995) and Rousseau, Sitkin, Burt & Camerer (1998) definition of trust- *"willingness to be vulnerable"*, and was discussed in the previous chapter.

What remedies can improve the willingness of A to be vulnerable to B? Firstly, access to credible information about B can help A make more informed decision about whether to enter in an exchange with B. Secondly, adoption of standardized procedures, like the double-entry bookkeeping, can limit the frequency with which B can behave opportunistically, as such standards can increase transparency and compliance. Malone, Yates & Benjamin (1987) showed how improvements in information technology could reduce coordination cost in industries like airline travel, by reducing search costs and increasing standardization across markets. Such an information access and standardization can make activities like making an airline reservation more impersonal and market-based, as they begin to rely less on intermediaries.

Given the challenges associated with the emergence of an impersonal system of trade, the following conditions must hold true[16]. When impersonal exchanges are not highly beneficial compared to exchanges with familiar contacts, impersonal exchanges do not occur because of lack of incentives (utility condition). Similarly, if reliable trade-related information is not available and business practices that make trade more reliable do not get adopted, an impersonal exchange does not occur because of high coordination costs (Malone, Yates & Benjamin 1987), and lack of confidence in impersonal partnerships (reliability condition). This draws me towards the first hypothesis.

Hypothesis 1 *In the absence of utility OR reliability conditions, relationship-based (R) systems like guilds are more efficient, as impersonal exchanges are not sustainable.*

When impersonal exchanges are beneficial compared to exchange with familiar contacts (utility condition) AND information access is improved, which makes trade more reliable through better information and improvements in business procedures (reliability condition), then an impersonal exchange is possible. But, the mere possibility of impersonal exchange does not suffice its emergence. Embedded institutions like merchant guilds are responsive to changing external conditions. When threatened with competition from impersonal exchanges, intermediaries in relationship-based institutions like guilds may reform and make themselves more competitive and efficient. To understand this, let's consider Figure 2.1 again. Consider a case when steady utility from relationship-based exchange through brokers or guild like networks were $m = \$10$, while risky utility for A from direct impersonal exchange with B was $c = \$30$. If such a utility difference motivated rise of impersonal exchange, then in such a case brokers in the relationship-based exchange may start offering a higher steady utility m

[16]In the previous chapter, I undertook a modeling approach to test this argument, exploring the conditions under which impersonal exchange can evolve in the absence of formal legal institutions, and it studies cases where people began to rely on partners beyond tribe, family or guilds.

Table 3.1: The table shows the conditions under which different economic systems sustain and emerge. Low utility condition refers to the case when difference between utility of impersonal and relationship-based exchange is low, while disruptive utility condition is the case when such utility difference is too high for a relationship-based institution to match. Low reliability condition refers to the case when absence of information makes impersonal exchanges costly and unreliable. Under low utility or low reliability conditions, only relationship-based (R) exchanges are suitable, while in presence of high utility and reliability conditions, relationship-based systems undergo reform (U), and impersonal systems (I) emerge if the impersonal opportunities are disruptive (disruptive utility).

Conditions	Low Utility	High Utility	Disruptive Utility
Low Reliability	Relationship-based (R)	Relationship-based (R)	Relationship-based (R)
High Reliability	Relationship-based (R)	Undergoing Reform (U)	Impersonal (I)

(if they could) to make the relationship-based channel of trade competitive. In other words, the relationship-based institutions that benefit from lack of impersonal exchange (like merchant guilds) may reform themselves, if in case the benefits from impersonal exchanges are not disruptive enough to offset the potential utility from relationship-based exchange. One would expect to see such reforms if the existing relationship-based institutions are already efficient and well positioned in trade. This will be true if the steady utility they can potentially offer (m) is not too low compared to that offered by impersonal exchange (c). This draws me towards the second hypothesis.

Hypothesis 2 *If the impersonal opportunities are large but NOT disruptively beneficial AND reliability condition is satisfied, relationship-based institutions like guilds may undergo reform (U) and make themselves more competitive.*

There are limits to how much the relationship-based institutions can reform, and stay competitive as an alternative to impersonal exchange. If relationship-based institutions could reform themselves without limits, then impersonal institutions will never emerge, because relationship-based institutions could always offer a service that was competitive to impersonal exchange. But, under certain conditions impersonal opportunities are disruptive. Under such conditions relationship-based institutions can no longer reform enough to keep themselves competitive in to impersonal exchange. For example, the traditional taxi or travel industry can reduce their prices to compete against crowd sharing based taxi or travel companies, but only till a certain limit . If the utility difference is disruptively large ($c >> m$), relationship-based institutions may make way for more impersonal institutions[17]. This draws me towards the third hypothesis.

Hypothesis 3 *If the impersonal opportunities are disruptively beneficial and there is an availability of trade-related information and adoption of standardized business practices (reliability condition), the relationship-based system becomes less efficient. So, impersonal (I) institutions may emerge as a dominant alternative to relationship-based systems like guilds.*

The different conditions and hypotheses are summarized in Table 3.1. Did the commercial and communication revolutions of Europe create the conditions in the sixteenth century that motivated the local authorities in Northwestern Europe to disband the merchant guilds and establish impersonal institutions?

[17]The above discussion on the emergence of impersonal institutions follows the theory (Proposition 2) developed in the section 2.3 on endogenous institutions of the previous chapter, and illustrated in Figure 2.5.

3.3 Decline of Guilds in Northwestern Europe

The medieval European merchant guilds traded in temporary fairs, of which the Champagne fairs of Northern France were the most prominent. In the thirteenth century, the Champagne fairs declined as a trading place, when the local authorities began to exploit the visiting traders. After the decline, the neighboring regions of the Low Countries began to attract the international merchants who traded at Champagne (Gelderblom (2013), p. 14). The international merchants eventually began to cluster in the emerging cities of Low Countries like Bruges and Antwerp. The visiting traders began to trade with the help of local hostellers and brokers, who acted as intermediaries between unfamiliar traders (Gelderblom (2013), p. 43).

3.3.1 The changing Atlantic coast

In the late fifteenth century, the discovery of the new sea routes to Asia and the Americas during commercial age opened up the Atlantic shores for beneficial long-distance trade (Davis (1973), Chapter 2). As the Low Countries shared a coastline with the Atlantic, the already booming inland trade of the region grew further with long-distance sea trade, bolstering the region's (including England) position as the leading cluster of international trade in Europe. One of the markets that soared in the early sixteenth century in Antwerp was of spice trade with Asia and Africa by Portuguese merchants who would trade through Antwerp with merchants from the Aachen area and South Germany (Tracy (1993), p. 28). As the region began to emerge as a major center for trade of commodities such as spices, the footloose international merchants, who would previously visit the cities for short durations, began to settle in the cities (Gelderblom (2013), p. 58). This led to a large increase in the city populations as reflected in historical population estimates. The new population would like to enjoy greater freedom and not rely on local and increasingly influential brokers and hostellers as intermediaries for networking and trade (Gelderblom (2013), p. 58). Moreover, observing the newly found opportunities in long-distance trade, middle-class non-merchants in the Low Countries and England began to aspire to enter the profession. But, they found the guild monopolies enjoyed by some merchants to be restrictive for their aspirations to build new business networks (Ogilvie (2011), p. 188). If the non-guild members wanted to trade with long-distance traders, they needed alternative impersonal channels other than guilds.

But, the participation in impersonal markets was not only limited by the monopolized guild networks but also by market frictions like information access and moral hazard. The fifteenth century was a period of underdeveloped impartial legal systems and contract enforcement (Ogilvie (2011), p. 33). Economic information was opaquely available to aspirational traders, as it was tightly controlled by merchant guilds (Ogilvie (2011), Chapter 9). Moreover, there was uncertainty regarding the reliability of other traders, which was the prime reason why hostellers and other intermediaries were so dominant[18]. The frictions of information and transparency

[18]Some of the long-distance relationships could be repeated which resolved moral hazard problems by creating ample incentives for partners to trade with each other again. A repeated exchange was the classic way in which long-distance trade has been described to have been organized by "individualist" Genoese merchants in contrast to "collectivist" Maghrib merchants (Greif 1994). But, several transactions did not involve such long sustained interactions. For example, new unestablished traders did not have well developed long-term relationships they could leverage on to enter in long-distance exchange. So, many traders would have liked to build new networks, which was

made making a large shift to impersonal interaction not feasible. So, the economy stayed in a relationship-based equilibrium where traders lacked the confidence in unfamiliar partners, as markets were of unknown reliability.

3.3.2 The development of postal network

During the end of the fifteenth century, just as Europe was undergoing a commercial revolution, it was also undergoing a communication revolution. The postal system was connecting the major trading cities of Europe, including the Northwestern region by the end of the fifteenth and the start of the sixteenth century. The development of postal networks at the end of the fifteenth and sixteenth century was useful in expanding the circle of contacts, as it helped establish lines of long-distance communication. A look at the map of early modern postal network in Europe, based on Giovanni da L'Herba's 1563 travelogue (see appendix Figure E.2), shows that the postal system was already well developed by the middle of the sixteenth century in Western Europe, extending between Italy, Low Countries and the Iberian Peninsula, and criss-crossing France and Germany. The map does not include important centers like London, where there were private arrangements for letter communication[19]. While, individual post offices may not have been established in every major city in Europe, but the already existing postal network ensured that the service could be arranged for traders and cities in need for long-distance communication.

In previous centuries an integrated European postal network did not exist, and it made communication difficult[20]. So, the development of a pan-European (and international) postal network enabled traders across Europe to seek and communicate with other long-distance partners. Such a correspondence through letters was considered to be an important aspect of trade since medieval times. Merchants not only wrote about "mutual business dealings but also ... information about the wider commercial environment: sales successes, market opportunities, prices of major commodities, exchange rates, commercially relevant political events, news of personalities and occurrences in merchant circles" (Ogilvie (2011), p. 372). Written evidence like merchant letters became increasingly acceptable as proof in courts, including in the Low Countries (Gelderblom (2013), p. 83; Puttevils (2015), p. 101).

3.3.3 The horizontal diffusion of printed material

At the end of the fifteenth century, another significant breakthrough in information diffusion was the invention of movable-type printing press. Printing diffused contagiously across Europe. High levels of diffusion of books in London, Antwerp, and other cities in the Low Countries, that got triggered in the fifteenth century, made best practices like the double-entry bookkeeping and new information regarding trade available with the availability of printed books (Puttevils 2015, Chatfield & Vangermeersch 2014). The unique aspect of the diffusion of information in the printing era was that the commercial information would mass-diffuse horizontally amongst

difficult in the absence of confidence in other unknown traders.

[19] In 1496 Henry VII granted the liberty to establish private postal service in London (Daybell (2012), p. 137).

[20] See Schobesberger et al. (2016) for a more detailed look at European postal networks in the early modern period.

peers, and not just vertically by parents, teachers and other authorities.

It is important to note here that Europe was not the first region to invent the printing press. China developed a movable-type printing press at the beginning of the second millennium (the 1040s), yet despite the early start, printing did not become as popular as it did in Europe and its effects on China were marginal. Angeles (2017) argues that this lack of adoption of the movable-type printing technology was due to the differences between the European and the Chinese script, as European script was alphabetic and its limited number of characters was suitable for the movable-type printing press, unlike the logographic script in China. Moreover, printing in Indian languages (which had a non-alphabetic Abugida script) only became popular in the nineteenth century after the invention of lithography (in 1796), which did not require alphabetic woodblocks (Orsini 2016). In fact, when the movable-type printing press was introduced in Goa, India in the late sixteenth century by the Portuguese, Konkani texts (the local language) began to be written in the Roman script as the Roman script was easier to print (Saradesāya 2000). Given these observations, it is important to consider the significance of the Roman Alphabetic Script in helping the spread of Printing in Europe. But, printing did not diffuse equally in all regions of Europe, and it diffused more in some cities than others.

Why did London, Antwerp, and other cities in the Low Countries adopt printing early? These cities were some of the geographically closest non-German cities to Mainz, the city where the movable-type printing press was invented by Gutenberg (Dittmar 2011). Dittmar (2011) argued that the movable-type printing press was heavy to transport and difficult to build without the small group of apprentices who knew how it was built. So, the cities closer to Mainz got the printing press earlier than others in the first few decades of the introduction of printing (the 1450s to 1500s), and such cities were early to print books that appealed to merchants.

Diffusion of "Economics" books and Hindu-Arabic numbers

A look at the Universal Short Title Catalogue (USTC) curated at University of St. Andrews, for books categorized as "Economics", shows the Mainz centric diffusion pattern in the early decades of printing (see Figure 3.2:

- The first book related to Economics according to the USTC catalog, *De contractibus mercatorum* or "On the contracts of Merchants" was printed in 1468 in Cologne by Johannes Nider. For the next ten years, all the economics related books were printed in the German speaking regions of Cologne, Strasbourg (1493), Basel (1475) and Esslingen (1475) (each of which were close to Mainz), with Rome (1473) being the only exception.

- In the subsequent decades, cities in the Low Countries also began to print books related to Economics in Bruges (1477), Brussels (1485), Louvain (1485), Antwerp (1487), Zwolle (1488) and Gouda (1489).

- The French cities of Lyon (1488), Paris (1493) and Provins (1496) also began to print such books, eventually. In the early sixteenth century London (1504) emerged as a printing center for economics-related books.

- The economics-related books began to get printed in Italian cities of Venice (1503), Pavia

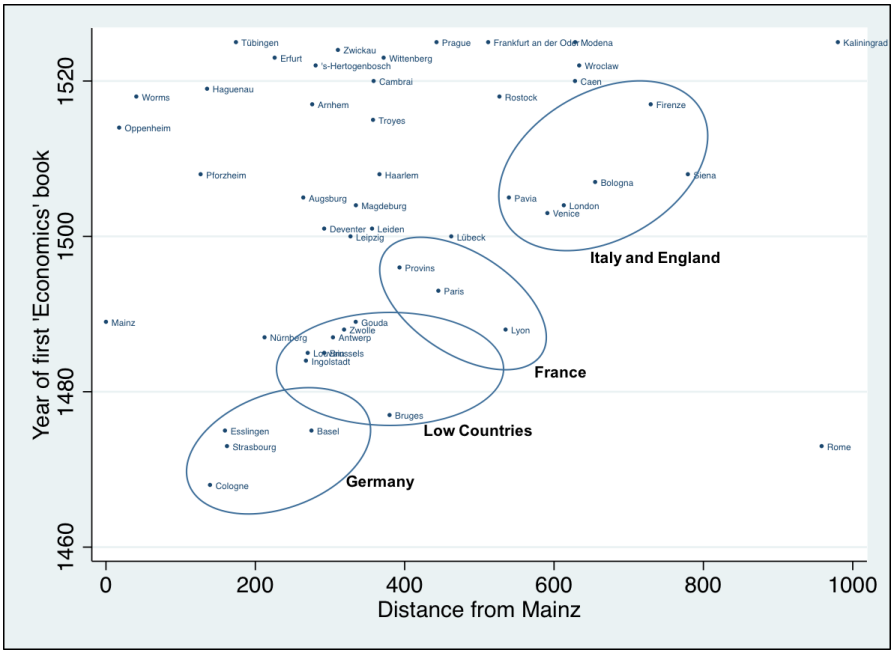

Figure 3.2: The two way plot of cities between variables Year of first Economics book (source: USTC catalogue), and distance from Mainz

(1505) and Bologna (1507) in the early sixteenth century according to the USTC catalog. But, even if economics as a special category emerged only later in Italy, the volume of printing of books in related categories like mathematics was large in the region, with notably the book on double-entry bookkeeping by Luca Pacioli being printed in 1494 in Venice. In fact, many of the commercial techniques being printed elsewhere in Europe, found their origins in Italy, including commercial arithmetic based on the Hindu-Arabic number system.

- While Polish and Czech cities eventually began to print economics-related books from 1522 in Wroclaw and 1525 in Prague, Spanish cities do not find a mention in the USTC catalog.

Of all the European cities, Antwerp (107 books till 1550) and London (72 books till 1550) emerged as the two major cities in the printing of books categorized as "Economics" (Paris being the third with 53 books till 1550). The Northwestern Region had other important centers for printing of economics-related books other than Antwerp and London, including Cologne (50 books till 1550), Ghent (41 books till 1550) and Amsterdam (10 books till 1550). The German, French and Italian cities like Strasbourg, Augsburg, Lyon, and Venice continued to be important cities for production of Economics related books as well. In Antwerp, adapted works of Luca Pacioli by Jan Christoffels, where he illustrated bookkeeping, were printed in Dutch (*Nieuwe instructie*) and French in 1543 (*Nouvelle instructie*), and in English (*A notable and very excellente woorke*) in London in 1547.

Just as the commercial ideas and business techniques were diffusing across Europe through books related to Economics and bookkeeping by writers like Pacioli and Christoffels, another fundamental change was occurring across Europe at the end of the fifteenth and sixteenth century- the adoption of the Hindu-Arabic number system. Once again, printing press played a

central role in the diffusion of the Hindu-Arabic numerals and the standardization of the Arabic notation, especially of numbers 4, 5 and 7 (Smith & Karpinski 1911). Appendix Figure E.3 from Hill (1915) (p. 92) illustrates this standardization in German printing and woodcuts at the end of the fifteenth century and the beginning of the sixteenth century, where numerals 4, 5 and 7 took on the modern forms that were already popular in Italy in the medieval period.

The commercial techniques and the Hindu-Arabic number system were also blending to increase the analytical abilities of merchants. In the bookkeeping records before the fifteenth century, there is little evidence of the use of the Hindu-Arabic number system outside of Pisa (Durham 1992), home to mathematician Fibonacci, who popularized the Hindu-Arabic system in Europe and wrote *Liber Abaci* in 1202 after learning it in Algeria. But, with the advent of the printing press, the Hindu-Arabic number system began to get adopted in account books outside of Pisa and Italy, and across Europe (Durham 1992). Luca Pacioli's 1494 masterpiece, *Summa de arithmetica*, also played a role in the popularization of the Hindu-Arabic numbers (Devlin 2011). Before the usage of Hindu-Arabic numbers, Roman numerals were the standard method of representing numbers, and such a non-positional system restricted the range of arithmetic operations that the merchants could perform. Even a simple operation like division was difficult and involved usage of the abacus. In other words, the usage of the Hindu Arabic numbers and arithmetic, especially in commercial settings, represented a significant (yet underappreciated) leap in commerce[21].

Effect on commercial culture

Trade-related information diffused through merchant manuals (Dittmar (2011)), and the new techniques in commerce and arithmetic diffused through books. Such a horizontal diffusion of information and new techniques through the printing press had a profound effect on European businesses and culture in general (see Eisenstein (1980)). The effect of printing on the business culture of Antwerp was described by Puttevils (2015) as:

> Sixteenth-century Antwerp served as a training ground and test site for commercial techniques and know-how as a result of the concentration of representatives of all European traders in one city. Low Countries merchants showed themselves eager students: they were trained by foreign merchants in Antwerp and abroad (either Low Countries expats or citizens of particular cities), and Antwerp's printing presses - Antwerp was one of the most important book production centres in sixteenth-century Europe - produced so-called *Ars Mercatoria* guides or 'Doing

[21] *Craft vs. Merchant guilds:* Liberti & Petersen (2017) present a range of benefits of "hard" information, which they call "information reduced to numbers", when compared to "soft" information (see Uzzi & Lancaster (2003)). These benefits in contemporary finance include a lower cost in information processing due to automation, and standardization, expansion of the market due to lowering of entry barriers, and increased transparency of durable information. While printing press could help in the diffusion of such commercially useful hard information, its impact could only be indirect on the diffusion of softer and more tacit knowledge. Such transfer of soft or tacit knowledge required person-to-person transmission and was important for the effectiveness of craft guilds (De la Croix, Doepke & Mokyr 2017). So, the disruptive impact of the printing press on craft guilds could only be limited in comparison to its impact on merchant guilds. So, the introduction of the printing press in the late fifteenth century, cannot be expected to cause a decline of craft guilds, unlike the decline of merchant guilds in the sixteenth century. Not surprisingly, craft guilds did not a decline in the same period as the merchant guilds, and they began to decline after the sixteenth century (Epstein 2008).

> business for dummies'. So, merchants from the Low Countries could easily pick up information on Italian-style accounting, letter writing, mathematics, business techniques such as the bill of exchange and languages. This training allowed them to catch up with the most recent techniques, but it did not give them an advantage over their competitors who were familiar with the same techniques.

Similarly Van der Wee (2013) described how double-entry bookkeeping was popularized and taught in Antwerp and elsewhere (p. 331):

> ...the printing press made the considerable distribution of commercial manuals possible. Even amongst ordinary merchants, training in Italy or elsewhere abroad became an established practice. So far double-entry bookkeeping had been limited to Italian firms. During the sixteenth century mainly under the influence of Antwerp, this technique was popularized and conquered commercial circles in Germany, the Low Countries, France, England and even in the Hansa centres. P. Savonne, Fr. Flory, and J. Ympyn were famous Antwerp teachers of the practice of double-entry bookkeeping and published important manuals on the subject there.

Another manner in which printing impacted commerce was by affecting business correspondences. Trivellato (2014) wrote about the role of business correspondence[22] in the rise of business between strangers in a cross-cultural context. She wrote (p. 17) that as "the etiquette of business letters became ever more standardized in the seventeenth and eighteenth centuries, it facilitated fiduciary relations among strangers." Describing the effects of the printing press on the standardization of the etiquette, she wrote (p.185):

> Generally speaking, if the invention of the printing press had a direct impact on business correspondence it was primarily on the codification of the rules according to which merchants ought to compose their letters... In the mid-sixteenth century, instructions for writing business correspondence began to appear among the large body of literature that goes under the rubric of *Ars Mercatoria*, which comprised dictionaries, tables for converting weights and currencies, treatises about bookkeeping and other practical instructions for merchants, as well as more elaborate introductions to the principles of merchants' law, sweeping histories of European commerce since antiquity, and a variety of eclectic pamphlets... The first such manuals were published in Antwerp, at the time when the city was Europe's greatest colonial emporium. Among these were Jean Bourlier's *Lettres communes et familieres*, a compilation of letters in French and Flemish published in octavo format 1576. Multilingualism and the small format were to become standard features of this genre.

A notable aspect of early modern printing was the increasing usage of the local language, especially in the commerce-related books. For a topic such as religion, Latin was the overwhelming language of choice (more than 50 percent of all books categorized as religious by

[22] Trivellato (2004) discussed the role of business correspondence as an important means of private information dissemination in Early Modern Europe.

the USTC catalog were in Latin until 1550). Printing in Latin was starkly not the case in Economics related books, where many books were printed in German, French, Dutch, and English, and only about 13 percent of the books (mostly in Italy and France) were written in Latin. The "Economics" books in England, Germany, and Low Countries were overwhelmingly written in the local language. Luca Pacioli's book *Summa de arithmetica* was written in Italian, while Jan Ympyn Christoffels' books on accounting were written in Dutch, French, and English. Such a vernacular usage in the printed books highlights that the intended audience of these commercially minded books were not just the highly educated elites, but also the common merchants.

Techniques of trade were no longer secrets as guild mysteries, because they were printed in books and manuals on topics related to economics and other related subjects, often in the local languages. There was greater learning on the art of trade, including in the teaching of standard skills like double-entry bookkeeping. With a better availability of information, and with the adoption of practices like bookkeeping, the reliability of traders increased, as courts also began to accept paper proof as evidence. Gelderblom (2013) wrote about the role of double-entry bookkeeping for monitoring of agents, especially in long-distance trade (p. 86):

> Periodic comparison of the accounts of both parties made it very difficult for agents to hide manipulations from their principals. This was obviously true for accounts kept in the Italian style, that is, with double entries, but merchants from Augsburg and Nuremberg, who applied their own, slightly different rules, were also able to keep tabs on their partners and employees. Before 1500 the accounts of merchants in the Northern Germany, France, England, and the Low Countries probably were too rudimentary to keep track of more complex credit transactions and substantiate financial claims in case of disputes. They started collaborating in specific purpose partnerships only in second quarter of the sixteenth century- about the same they adopted double-entry bookkeeping.

Highlighting the significance of specific purpose partnerships in sixteenth-century trade, Gelderblom, de Jong & Jonker (2011) (pp. 32-33) wrote:

> During the second half of the sixteenth century, merchants in Britain and the Habsburg Netherlands began to explore new markets in Russia, the Eastern Mediterranean, and the coast of West Africa. These ventures carried considerable risk because of violence at sea, stark fluctuations in supply and demand, and the difficult monitoring of partners and employees trading in the distant markets. To manage these risks, British and Dutch merchants amended existing partnership contracts with additional clauses about the purpose and duration of the venture, the capital invested by the partners, the division of work between them, and, for those who contributed labor rather than capital, their share in profits and losses... This emendation of the general partnership's rules had become accepted practice in Antwerp as early as 1537, for an accounting manual published in that year [Hoecke, 1537] stated that "there is no difference between the rule of a partnership with specified duration and without specified duration, except that shares are taken for a certain period, and the revenue is calculated according to this share".

The printing press brought more transparency in information and diffusion of business practices and etiquette, refined by the arithmetic based on the Hindu-Arabic system. Such transparency motivated a new class of traders to join trading in an impersonal setting. Increasing pressure from footloose merchants and the rising influence of a middle class that was not part of guilds, bolstered by (i) large benefits of trade at the Atlantic coast, (ii) learning of trade-related information and new business practices, and (iii) more frequent long-distance postal contacts, provided the favorable utility and reliability conditions for impersonal exchange to emerge and be sustained. These favorable conditions motivated the local authorities to disfavour privileged monopolies, and they began to build institutions for supporting impersonal exchange, including development of impartial legal institutions that significantly reduced the risks of such a trade[23]. In London, the livery companies found it increasingly difficult to enforce their economic privileges as the sixteenth century progressed (Ogilvie (2011), pp. 32-33), while alternate forms of partnerships were emerging, like the joint stock company of *Merchant Adventurers to New Lands* was chartered in 1553 with 250 shareholders. Similarly, in Antwerp[24] merchant guilds began to decline right around 1500 (Ogilvie (2011), p. 32), in Bruges only until mid sixteenth century (Ogilvie (2011), p. 12), and in Amsterdam (as it was emerging as a major city only in the sixteenth century) it never had merchant guilds barring in some specific areas of commerce (Ogilvie (2011), p. 32).

3.3.4 Iberian Peninsula: Trade by Elites

Other regions of Europe did not have the conditions as favorable as in Low Countries and England. Spain and Portugal had a long Atlantic coast and commercially advanced cities with postal networks. Cities like Seville grew rapidly in the sixteenth century due to their position on the Atlantic coast. But, printing did not penetrate deep into the region early on. Being distant from Mainz, the region was not an early adopter of printing. So, the diffusion of printed books was low with limited availability of trade-related information (guild mysteries) and business practices. For example, the first book on double-entry bookkeeping in Spanish was published in Madrid in 1590 by Seville based Bartolom Salvador (Edwards (2013), p. 68), when such accounting books in Dutch, French, and English were written much earlier by 1540s by Ympyn, and the original book by Pacioli in Latin in 1494.

Given the traditional information barriers and reliability issues, guilds continued to be an effective system. So, the local authorities continued to favor merchant guilds that held onto their privileged positions in the economy, because impersonal exchange could not emerge as a credible alternative to challenge the guild-based system of the region. In fact, many Spanish cities competed to receive guild privileges in late medieval period (Smith 1940), and cities like Sevilla further tightened their requirements (20 years residence) on who could undertake trade in the city (Ogilvie (2011), p. 54).

[23]Fiscal health in some regions (England and Netherlands) further (through confiscation of church land in newly Protestant areas (Ogilvie (2011), p. 188) ensured that they had enough coffers to risk the shift.

[24]In a relevant book by Padgett & Powell (2012), authors explore the various processes that produce novelty in organization forms, and in Chapter 7 study how migration of Protestant merchants of Antwerp to Amsterdam led to the blending of the commercial and shipping skills of the two cities to create new organization form of the joint stock company.

Figure 3.1 shows the geographic advantage cities in the Northwestern Europe (Region 1) had over other cities because of their proximity to the Atlantic coast, as well as from the city of Mainz. The Spanish and Portuguese cities, on the other hand, were close to the Atlantic coast but were distant from Mainz (Region 3).

3.3.5 Northern Italy: Absence of Incentives

Northern Italy provides another relevant case in contrast to Spain and Portugal. The region had a well developed postal network (Schobesberger et al. 2016) and it also had a high penetration of printed books by the late fifteenth century, and several commercial books were being printed in its cities. Standardized business practices like bookkeeping were well known, and these practices were being popularized elsewhere by the printing of books and sharing by merchant letters. In fact, revolutionary financial innovations like double-entry bookkeeping were developed in the region and diffused in the rest of Europe from there. For example, Luca Pacioli's 1494 book on bookkeeping was written in Venice and was adopted by Jan Ympyn Christoffels later in Antwerp.

But, the opportunity for impersonal exchange in Northern Italy was not disruptive enough (Hypothesis 2), as the region was already at the center of the Mediterranean trade and of land trade to Asia. For the merchants in the Iberian cities, the Atlantic route to Asia (and to the Americas) was an alternative route to undercut the domination of the Venetians in the trade with Asia (Tracy (1993), pp. 26-28). In the region, trade-related information and standardized business practices were widespread, but impersonal trading opportunities were large but not disruptive enough (Region 2 of Figure 3.1). Existing elites could initiate reform that made opportunities from the existing system more beneficial and averted the potential challenge of the emergence of impersonal exchange. In the North Italian region, especially in Venetian and Milanese clusters, reforms were going on during the sixteenth century (Epstein (2004), pp. 301, 308). Existing elites held onto their power in the regions (Ogilvie (2011), p. 53), but traders and producers from the countryside, like the silk producers of village areas of Milanese Lombardy, got more concessions from urban elites (Epstein (2004), p. 308)[25]

3.3.6 Hamburg and Lübeck: A diverging tale

In North Germany[26], the relevance of sea coast on institutional outcomes is starker to observe. Lübeck and Hamburg were the two prominent cities of North Germany. The two cities are situated close to each other around the Jutland peninsula (see Figure 3.3), and both of these cities were part of the Hanseatic league of prominent German trading cities. One key geographic difference between the two closely situated cities was, while Lübeck was on the Baltic coast and a Baltic entrepôt, Hamburg was on the Atlantic coast and would receive ships from the Atlantic. Both the cities had established printing presses, although it was Lübeck which emerged as the larger trading and printing hub. In the sixteenth century, the Atlantic-facing Hamburg (located

[25] *Genoa*, a Mediterranean entrepôt, which benefited from several medieval financial and commercial innovations of Italy, was always a city without guilds with liberal requirements for earning a citizenship, unlike Venice which although did not have guilds but imposed restrictions on who could trade using tough citizenship requirements (Ogilvie (2011), p. 53).

[26] Appendix D provides a case study based on the divergence between Hamburg and Lübeck.

Figure 3.3: Lübeck and Hamburg are located close to each other in the Jutland peninsula, and they acted as important junctions of the traditional trade route of North Europe. In the fifteenth century, a competing trade route around the Jutland peninsula, and through the Sound (Øresund) emerged, which threatened the traditional domination of these two cities.

in Region 1 of Figure 3.1) became more impersonal and dismantled the monopoly privileges merchant guilds enjoyed. In contrast, the Baltic facing Lübeck in the sixteenth century not only continued with the merchant guild privileges, it further made them more entrenched (Dollinger 1970, Lindberg 2009, Ogilvie 2011).

What deterred Lübeck from becoming an Atlantic trading city like Hamburg? Due to the location of Lübeck in the Jutland peninsula, it acted as a gatekeeper of Baltic trade and its dominance came from the staples it extracted through land trade between Lübeck and Hamburg (Bes, Frankot & Brand 2007). Lübeck was locked in Baltic trade like Venice was locked in Mediterranean trade. To undermine Lübeck's domination, Dutch merchants (like Portuguese in the Atlantic) found an alternate route to reach the Baltic Sea by circumnavigating around the Jutland peninsula and through the Sound (Øresund) (Bes, Frankot & Brand 2007). Given the domination Lübeck had enjoyed in the now faltering Hanseatic trade system with the Baltic, Lübeck tried to defend its dominant position in Baltic trade against the emerging Atlantic trading system. In contrast, Hamburg became a participant of the emerging Atlantic trading system, and like other cities of the North Sea (like Amsterdam), it attempted to attract the free moving traders, who in the early sixteenth century were flocking the Low Countries, especially Antwerp (Puttevils (2015), p. 46).

Rest of the German cities, as well as major towns in French inland, also continued to have dominant merchant guilds despite high levels of printing penetration, as they were distant to the beneficial long-distance sea trade, especially the Atlantic coast. Even Hanseatic cities like Cologne that were at the forefront of trade in previous centuries due to the presence of navigable rivers could not compete with cities like Antwerp or Amsterdam (and eventually Hamburg)

which were better located geographically and acted as entrepôts for the whole North European region. In the city of Augsburg (a major center of printing, and commerce), there was some anti-monopoly dissent which was decided in favor of the merchant companies, arguing that private interest and common good were compatible (Häberlein 2012).

Considering Figure 3.1, in the sixteenth century; the Northwestern Europe with a high printing penetration and an Atlantic coast fell in Region 1, which was favorable for emergence of impersonal markets; Northern Italy with high printing penetration but no Atlantic coast fell in Region 2, which was favorable for relationship-based systems undergoing reform; and Spain and Portugal, and most of Europe, because of having either low rates of printing penetration or poorer opportunities for long-distance trade, fell in Region 3, which had persistent relationship-based systems. So, considering different regions of Europe, historical evidence suggests that different intensities of trade opportunity and printing penetration led to different outcomes in different regions of Europe.

Can we observe such patterns in the data collected from medieval and early modern Europe?

3.4 Empirical Study: Why impersonal markets emerged?

3.4.1 Empirical Strategy

In the sixteenth century, different regions of Europe developed different types of economic institutions. The story developed in section 3.2.2 and 3.3 and summarized in Table 3.1 and Figure 3.1 suggests that an area was better placed to have impersonal markets or to undergo reform in the sixteenth century if the area enjoyed both, large benefits from impersonal exchange, and high information access and diffusion of new business practices. Otherwise, the area had relationship-based institutions.

So,the basic hypothesis can be expressed as an OLS model:

$$Inst_a^{1600} = \alpha + \beta_1 Opp_a + \beta_2 Information_a + \beta_3 Opp_a X Information_a + \epsilon_a \tag{3.1}$$

where $Inst_a^{1600}$ is the nature of economic institutions in the sixteenth century in a given area a, Opp_a is the size of benefit from impersonal exchange, $Information_a$ is the size of diffusion of trade-related information and business practices. I expect coefficient β_3 (the interaction) to be the only sizable and significantly positive coefficient (Hypothesis 2). If in an area impersonal exchange was disruptively beneficial (code it as a dummy $dDisrupt_a = 1$), then we expect that area had a higher probability of having impersonal markets. So, we expect to observe a larger effect of $Information_a$ on $Inst_a^{1600}$ in areas with $dDisrupt = 1$ (Hypothesis 3).

To understand the puzzle of transition, we first need to identify the nature of economic institutions in the different parts of Europe in the sixteenth century. Secondly, we need to find good proxies that can measure the size of the benefit of impersonal exchange and the level of information access and adoption of standardized practices. We can then explore the relationship between the proxies and nature of economic institutions.

Table 3.2: The table explains the scheme adopted for classifying cities as impersonal (I), undergoing reform (U), or relationship-based (R) in the 16th century.

Code	Description
I1	A city that never gave monopoly privilege to particular traders/groups.
I2	Sixteenth-century impersonal city where monopolies over trade were removed at least partially.
U1	Sixteenth-century city undergoing reform where while urban elites enjoyed monopoly privileges, but they gave concessions to rural or non-guild producers/traders.
U2	Sixteenth-century auxiliary city close to an impersonal city, where rural or non-guild producers/traders were not coerced to trade through the auxiliary city.
U3	Sixteenth-century city undergoing reform where specific details regarding guild reforms are not documented at the city level, but region has been identified as undergoing reform.
R1	Sixteenth-century relationship-based city where merchant guilds have been documented to enjoy monopoly privileges.
R2	Sixteenth-century relationship-based city where merchant guilds were nonexistent as the social institution in the region was documented to be was feudal.
R3	Sixteenth-century relationship-based city where specific details regarding guild privileges are not documented at the city level, but their monopolistic nature can be inferred based on other descriptions.

3.4.2 Nature of economic institutions, impersonal or not?

Merchant guilds in Europe were associated with particular cities. The German Hansa was an example of a regional association between merchants, but even the Hansa was a federation of different cities. What was the nature of economic institutions in the sixteenth century in the largest cities that dominated Europe between the fourteenth and sixteenth century? Did they allow for impersonal marketplaces?[27] The largest cities of Europe provide a good snapshot of the commercial life of Europe, as the largest cities would also be the business or political centers. Tables E.1, E.2 and E.3 in appendix E.1 list down the largest cities of Europe in the fourteenth, fifteenth and sixteenth century as listed in Bairoch, Batou & Chevre (1988), a historical data source on city level population of European cities.

Looking at secondary historical information on each of the 81 cities, I coded the nature of economic institutions of the cities, with cities coded as relationship-based (R), Undergoing Reform (U) and Impersonal (I) depending upon whether a city restricted trade to particular groups. The coding criteria are detailed in Table 3.2. I code a city as relationship-based (R) if it provided monopoly privileges to merchant guilds, or if guilds did not exist at all, and trade was restricted using other methods. I code a city as undergoing reform if a city was giving concessions to locals and other merchants who were not part of the restrictive group that enjoyed the monopoly privileges. Finally, I code a city as impersonal if it began to remove monopoly privileges in trade in the sixteenth century and opened up trade for a wider demographic.

Genoa (coded as I1) was an exceptional city[28], as throughout history the city allowed all citizens to trade freely, with liberal requirements for citizenship (Ogilvie (2011), p. 53). All

[27] See Gelderblom & Grafe (2010) for a detailed discussion of guild institutions in cities of Amsterdam, Antwerp, Bilbao, and Bruges.

[28] Greif (1994) modeled the impersonal nature of trade in Genoa as a repeated one-sided prisoner's dilemma game.

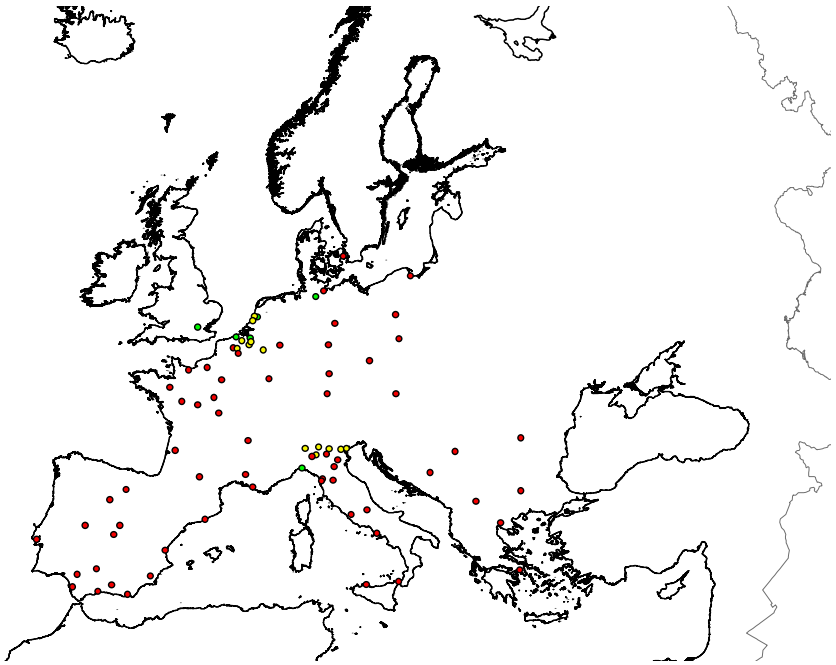

Figure 3.4: Map of 50 largest cities of Europe during the fourteenth, fifteenth and sixteenth century (Bairoch, Batou & Chevre 1988) based on type of institutions. Label- Green: I1 & I2 (impersonal), Yellow: U1, U2 & U3 (undergoing reform), Red: R1, R2 & R3 (relationship-based).

other cities restricted who could trade in the city, and just a handful of cities in Northwestern Europe were impersonal during the sixteenth century period (coded as I2). The cities undergoing reform were found in North Italy and the Low Countries (coded as U1, U2 or U3), where guilds eased restrictions on trade for locals without losing their monopolies. In rest of the cities (coded as R1, R2, and R3) guilds either enjoyed monopoly privileges or else the system of trade was feudal. Some special cases may be mentioned, for example, Nuremberg, which was called a "city without guilds", but was in practice an oligarchy that restricted formation of new guilds (Soly 2008) (thus not being an impersonal but rather a relationship-based city according to the coding in Table 3.2). Another interesting city was Venice, which although did not have guilds, yet imposed several restrictions on who could be a citizen and thus trade (Ogilvie (2011), p. 53). But, as Epstein (2004) (p. 301) recounts, Venice was giving concessions to rural representatives in trade (so, being coded as a reforming city in the sixteenth century). French cities continued to enjoy guild monopolies until the eighteenth century (Horn (2015), p. 224), and the partial abolition of guilds in cities of Rouen, Nancy, Metz, Roussillon, and Paris happened only in 1776, while other cities rebuffed these reforms. Based on the coding criteria in Table 3.2, Table E.4 in appendix E.2 details the nature of economic institutions in each of the 81 cities in the sample, based on secondary sources. Figure 3.4 maps the cities based on their coded type. Genoa was always an impersonal city, and so I do not include it in the dataset.

Variable $Inst_a^{1600}$ measures the level of market impersonalization in the sixteenth century in a given city. I code $Inst_a^{1600} = 2$ for all impersonal cities with coding I2, and cities undergoing reform with code U1, U2 and U3 have $Inst_a^{1600} = 1$. Other cities coded as relationship-based (R1, R2, and R3) have $Inst_a^{1600} = 0$.

3.4.3 Utility Condition: Sea distance as a proxy for impersonal opportunity

Closeness to the sea is a geographical characteristic where cities closer to the sea have natural advantages for long-distance trade. Being a long-distance trader in the relationship-based system required extensive networks, which was especially difficult for immigrants, small and aspirational traders. In such a scenario, if impersonal exchange could be developed, i.e., if unfamiliar partners could be trusted and relied upon, the benefits from such an impersonal exchange was large as immigrants, small and non-traders could enter into partnerships with each other and conduct long-distance trade. So, European cities that were close to the sea were suitable for long-distance trade and can be expected to find a partnership with unfamiliar traders more beneficial (Utility condition).

Acemoglu, Johnson & Robinson (2005) listed down 166 cities that acted as the Atlantic and Mediterranean ports of Europe. I calculate the distance of each city a in the database to the nearest Atlantic or Mediterranean port ($PortDist_a$) (see Figure 3.5). I also calculate the closest distance from the sea ($SeaDist_a$), and for cities which were closer to Baltic Sea (Plovdiv, Poznan, Prague, and Wroclaw) or were Baltic Ports (Lübeck, Gdansk, and Copenhagen) I use the distance from the sea as a measure instead of the closest distance from an Atlantic Port. I square root transform the combined variable $SeaPortDist_a$ and invert to make it an increasing positive variable ($SeaPortCloseness_a = constant - \sqrt{SeaPortDist_a} \geq 0$) representing closeness to Atlantic or Mediterranean port or Baltic sea, to be used as a proxy for Opp_a.

Mediterranean sea acted at the center of European long-distance trade until the discovery of new Atlantic routes to Asia and the Americas, during the fifteenth century. For early modern Europe, the discovery was disruptive as it opened trade in newer lands and of newer commodities (Acemoglu, Johnson & Robinson 2005). Cities that were Atlantic ports can thus be expected to enjoy a disruptive opportunity for long-distance trade and be more likely to impersonalize (Hypothesis 3 and Table 3.1) compared to other cities close to the sea like Venice and Lübeck that were already dominant in their respective Mediterranean and Baltic trades. If a city was an Atlantic port or it was close to an Atlantic Port ($PortDist_a < 50km$), I label such a city as enjoying disruptive opportunity, and code $dNearAtlantic_a = dDisrupt_a = 1$.

3.4.4 Reliability Condition: Printing penetration as a proxy for diffusion of new information

Impersonal exchange was not just limited by lack of beneficial opportunities. Information was opaquely available, and there existed a general doubt about the reliability of unfamiliar traders. Better information diffusion technologies could have helped as discussed extensively in section 3.3.

Printing was a disruptive technology (Dittmar 2011) as it reduced the cost of acquiring information and triggered the rise of horizontal information diffusion. Unlike the postal system that could easily diffuse (thanks to the extensive network) and be served by formal (the Taxis of Europe) and informal (like the arrangement in London) arrangements, the printing technology was not as easily diffusable. Commissioning of a movable-type printing press required experts that were not abundantly available (Dittmar 2011). So, the cities that got an early head start in the

availability of printed books could affect the access traders had to trade-related information and new business practices. So, a city that had higher printing penetration in the fifteenth century can be expected to have been better availability of information regarding trading practices among its residents. To estimate penetration of printed material in the 81 cities in the database, I use a database on early printing in Europe. There are several printing databases, with minor differences. I rely on *Gesamtkatalog der wiegendrucke* (GW) database to build the database of early printing cities and the penetration of printing material.

I would characterize a city in the GW database as a printing city if it printed more than ten books until 1500 (see Figure 3.6). As books were not geographically bounded, books printed in one city b were being read in another. For example, the book called *Summa de arithmetica* by Luca Pacioli that gave the first printed description of double-entry bookkeeping was printed in Venice in 1494, but it quickly became popular across Europe. But, one can assume that more populous cities and closer cities would have better availability of books from a city b than others.

Could it be that some printing hubs were printing books with a target audience in a geographically distant region? Like German books being printed in Venice? Universal Short Title Catalogue curated by the University of St. Andrews lists the language of printed books. I find that a negligent proportion of books printed until 1500 were being printed in a language that was neither local nor Latin. For example out of 3692 books printed in Paris until 1500 (according to USTC) only six were printed in Dutch, the maximum for a non-local language. Similarly out of 3570 books printed in Venice until 1500 (according to USTC), five were printed in German, the maximum for a non-local language.

I use a gravity model to estimate printing penetration in the 81 cities of interest. If one of the 81 cities a with population $Pop1500_a$ and one of the 121 printing cities b printing B_b books, were having a distance d_{ab}, then per capita printing penetration in the city a by books from city b, $PrintIndex_{ab} = \frac{B_b}{d_{ab}^2(\sum_{i=1}^{81}\frac{Pop1500_i}{d_{ib}^2})}$, where $\frac{Pop1500_c}{d_{cb}^2} / \sum_{i=1}^{81}\frac{Pop1500_i}{d_{ib}^2}$ represents the relative influence (mass) of a city a relative to b. When the printing city was the city itself (b=a), I normalise $d_{ab} = 1$. So, the total printing penetration in the city a because of all printing cities (121 in total) was $PrintIndex_a = \sum_{j=1}^{121}\frac{B_j}{d_{aj}^2(\sum_{i=1}^{81}\frac{Pop1500_i}{d_{ij}^2})}$ books per 10000 population (bptp). I log transform the variable to $PrintPentr_a = const + ln(PrintIndex_a) \geq 0$, to use it as a proxy for trade-related information access and diffusion of business practices ($Information_a$) (see Figure 3.7). I also count number of print cities in a 50km radius $PrintCity50km_a$, for a city a, and use it as an alternative measure of printing penetration.

3.4.5 Preliminary Observations

A two-way plot in Figure 3.9 between printing penetration ($PrintPentr_a$) and closeness to the sea ($SeaPortCloseness_a$), supports the basic arguments developed in this chapter.

1. **Cities away from the sea:** In cities that are away from the sea (closer to zero on the X-axis in Figure 3.9) guilds continued to be dominant in the sixteenth century. Cities close to the sea were the ones most likely to attract outsiders, and become cosmopolitan hubs. So, in an inland city like Wroclaw, there was little benefit for traders to seek out impersonal relationships, and the merchant guild system was an effective system already. In other

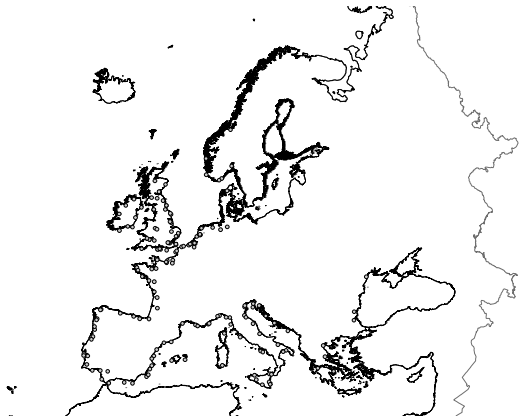

Figure 3.5: Map of the 166 Atlantic and Mediterranean Ports of Early Modern Europe (Acemoglu, Johnson & Robinson 2005).

Figure 3.6: Map of the 121 cities that printed more than 10 books in the 15th century in the GW catalogue. Size of the circle represents the number of books being printed.

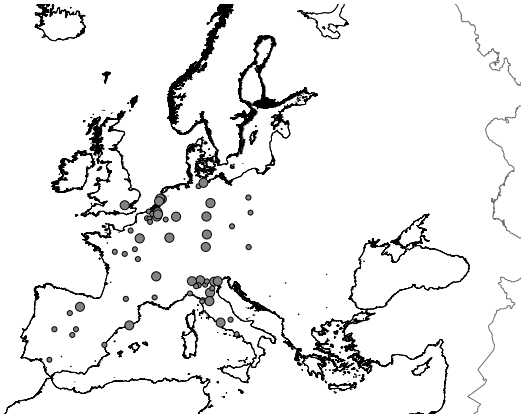

Figure 3.7: The $PrintPentr_a$ measure of the 81 largest cities in the sample. Size of the circle represents the degree of printing penetration.

Figure 3.8: Map of 335 largest cities of Europe during the fourteenth, fifteenth and sixteenth century (Bairoch, Batou & Chevre 1988) with population greater than or equal to 10,000.

Figure 3.9: The two way plot of cities between variables $PrintPentr_a$ and $SeaPortCloseness_a$ for different types of institutions.

words, the only cities where traders could potentially have benefited from impersonal exchange, were those close to the sea, so that they could attract unfamiliar outsiders and motivate traders to seek partnerships with those they did not know too well.

2. **Cities with low printing penetration in the fifteenth century:** Among cities that were closer to the sea (on the right side of the X-axis in Figure 3.9), the cities that were having low levels of printing penetration in the fifteenth century (closer to zero on the Y-axis in Figure 3.9) continued to have dominant guilds in the sixteenth century. These cities primarily included major port cities of the Iberian peninsula like Lisbon.

3. **Cities not on the Atlantic coast:** Among cities that were close to the sea and had high levels of printing penetration in the fifteenth century (on the upper right corner of Figure 3.9), several cities loosened monopoly privileges (in the Low Countries and North Italy), but only a few actually saw removal of guild monopolies. While both the Atlantic and non-Atlantic cities with high levels of printing could be argued to have the conditions that favored the decline of guilds, it was only the Atlantic coast where guilds actually declined as the opportunities found in these cities was disruptive[29]. All impersonal cities (green squares in Figure 3.9), except Bruges, were Atlantic port cities, and Bruges was only 20km away from the Atlantic Port of Vlissingen (in previous centuries Bruges itself

[29] *Growth of Mediterranean and Atlantic ports:* Mediterranean cities especially of North Italy (where printing was high) were more established commercial centers and arguably enjoyed more powerful guilds. This meant that they were more likely to resist opening of markets and instead undergo reforms when they faced the new commercial revolution at the Atlantic coast. One proxy that highlights the differences between the power of the Atlantic and Mediterranean cities is that Mediterranean cities in the sample were having an average population of around 48,000 on an average in 1400, while the Atlantic cities were having an average population of around 38,000 (See appendix Figure E.4). As guilds at the Atlantic coast held a lower clout in previous centuries, they were more likely to decline than Mediterranean cities, and give way to impersonal markets as these cities found the changes due to the commercial revolution more disruptive.

was a major port, but the port silted).

Moreover, I code the cities based on whether they had a postal service as per Giovanni da L'Herba's travelogue in 1563 (the earliest compilation of such a data), and find a small correlation of 0.0099 between the postal dummy and the type of institution, which is expected given that the postal network (formal and informal) had already developed extensively across Europe (not all of them listed in Giovanni's travelogue) by the middle of the sixteenth century and had especially reached most of the large cities (especially Western cities) that have been considered in the database. From the Figure E.2 itself the evidence is clear that the network for postal communication in Europe existed already, so now I would focus on the printing penetration as the city-specific heterogeneous variable for horizontal information diffusion.

3.4.6 Other Factors

Urban Agglomeration and Population

North Italy, Belgium, and the Netherlands enjoyed high levels of urban agglomeration. It was argued by Gelderblom (2013) that the Low Countries were able to evolve institutionally because of competition between close competing cities. Is high level of urban agglomeration a mediating factor that attracted higher rates of printing in regions close to the sea?

To measure the level of urban agglomeration, from Bairoch, Batou & Chevre (1988) population dataset I consider 335 cities that had a population greater than or equal to 10,000 in the fourteenth, fifteenth and sixteenth century (including the 81 largest cities, see Figure 3.8). I cluster the cities using hierarchical clustering method based on average city distance (UPGMA)[30]. The procedure produces a dendrogram (given in Figure E.1 in appendix E.3) which can be cut off at a given threshold. For a given threshold, if a city a is part of a cluster made up of n cities, then the city has clustering index $Clust_a = log(n)$. I cut the dendrogram at different thresholds for robustness and use the clustering in Table E.5 in appendix E.3 for the standard regressions.

I consider the log of population in the fifteenth century ($Log(Pop1500_a)$), and growth in population during the fifteenth century ($Growth1500_a = log(\frac{Pop1500_a}{Pop1400_a})$), as basic controls that may affect nature of economic institutions. Data related to population (total and growth) estimates not only demographic characteristics but also the economic characteristics at the city level, as historically the largest cities were also the most prosperous ones. Large cities with high population being economically dominant could attract more traders. Similarly, growing cities were attracting traders because of greater availability of opportunities. I also consider the square root of elevation ($\sqrt{Elevation_a}$ for each city a), as cities at higher altitude may be more difficult to reach in the medieval and early modern period, and thus they may be unsuitable as commercial towns, and more suitable as political and administrative centers. The three factors could affect impersonal nature of cities.

[30] *Hierarchical clustering:* The UPGMA algorithm constructs a rooted tree (dendrogram) that reflects the structure present in a pairwise similarity matrix (or a dissimilarity matrix). At each step, the nearest two clusters are combined into a higher-level cluster. The distance between any two clusters A and B, each of size (i.e., cardinality) $|\mathcal{A}|$ and $|\mathcal{B}|$, is taken to be the average of all distances d(x,y) between pairs of objects x in $\mathcal{A}$ and y in $\mathcal{B}$, that is, the mean distance between elements of each cluster: $\frac{1}{|\mathcal{A}|\cdot|\mathcal{B}|}\sum_{x\in\mathcal{A}}\sum_{y\in\mathcal{B}} d(x,y)$.

Medieval Fairs

Medieval cities that hosted temporary fairs, where European traders (affiliated to guilds) would gather, could be good precursors for emerging impersonal markets. One of the most prominent medieval fairs in Europe was the Champagne fairs in Northern France. Gelderblom (2013) wrote how the decline of the medieval fairs gave rise to markets in the Low Countries, as footloose merchants trading in Champagne moved away to the North. But, it is noteworthy that cities in the Champagne region did not themselves impersonalize. So, it is possible that fair cities could themselves evolve into impersonal cities. But, it is also alternatively possible that the traditional institutions in established medieval fair cities resisted evolution of impersonal economic institutions.

Regardless of the direction of causality, if any, medieval fairs reflected an important aspect of the medieval commercial system of Europe. I borrow data on the location of fairs during 1450-1500 in Europe, from the University of Iowa Library's Atlas of Printing[31] which builds its data from a variety of sources. I label a city a as a fair city if the city is listed to host fairs in the Atlas of Printing and I code $dFair_a = 1$. The following cities in the sample were fair cities in Europe according to the Atlas of Printing:

- Angers, Bordeaux, Caen, Lille, Lyon, Orleans, Paris, Rennes, Rouen, Toulouse and Tours in France,
- Antwerp, Bruges, and Ghent in Belgium,
- Bologna, Ferrara, Florence, Milan, Rome and Venice in Italy,
- Cordoba, Plasencia, Sevilla, and Valladolid in Spain, and
- London (Westminster) in England.

3.4.7 Empirical Results

Table 3.3 provides descriptive statistics of the data, part of which has already been described in Figure 3.9 where non-relationship-based cities have higher levels of printing and are closer to the sea. But, note that aggregate population growth in 1500 of the impersonal cities is higher than that of the reforming cities whose growth is higher than that of the relationship-based cities. Such high levels of population growth in the medieval and early modern period highlights how impersonal (and relationship-based) cities were attracting people because of the economic opportunities. Moreover, impersonal and reforming cities were more clustered and at a lower altitude from the sea on average than an average relationship-based city. Population level in the fourteenth and fifteen century does not show any distinct pattern- cities undergoing reform were more populated than the general sample in the fifteenth century while impersonal cities were less populated. More impersonal cities on an average were holding medieval fairs than relationship-based cities. But, fewer cities undergoing reform on an average were holding medieval fairs.

[31] The library's data can be accessed at http://atlas.lib.uiowa.edu

Table 3.3: The table presents the summary statistics of key variables for the whole sample of 81 cities, and further decomposes the statistics based on the type of the 16th century economic institution (see section 3.4.2).

	(1)	(2)	(3)	(4)
		Whole Sample		
VARIABLES	N	mean	min	max
# cities in cluster	81	7.852	1	22
Print Index	81	9.250	0.217	63.14
# Print Cities in 50km	81	1.309	0	6
Population 1400	75	36.99	3	275
Population 1500	81	37.37	5	225
Population 1600	81	52.94	7	300
Population 1850	81	122.1	4	2,236
Distance from Port	81	106.4	0	453.4
Growth 1500	75	0.114	-1.273	2.079
Medieval Fair	81	0.309	0	1
Elevation from sea	81	127.1	1	862

	(1)	(2)	(3)	(4)
		Inst = 0		
VARIABLES	N	mean	min	max
# cities in cluster	62	6.258	1	22
Print Index	62	8.381	0.217	63.14
# Print Cities in 50km	62	0.758	0	3
Population 1400	56	35.25	4	275
Population 1500	62	36.48	5	225
Population 1600	62	51.79	7	300
Population 1850	62	97.97	4	1,053
Distance from Port	62	124.5	0	453.4
Growth 1500	56	0.0828	-1.253	2.079
Medieval Fair	62	0.306	0	1
Elevation from sea	62	156.0	5	862

	(1)	(2)	(3)	(4)
		Inst = 1		
VARIABLES	N	mean	min	max
# cities in cluster	13	13.54	4	22
Print Index	13	13.04	3.807	47.20
# Print Cities in 50km	13	3.462	0	6
Population 1400	13	38.46	6	100
Population 1500	13	43.23	12	100
Population 1600	13	49.69	11	151
Population 1850	13	72.62	26	209
Distance from Port	13	67.57	0	143.6
Growth 1500	13	0.202	-0.405	0.847
Medieval Fair	13	0.231	0	1
Elevation from sea	13	43.46	1	150

	(1)	(2)	(3)	(4)
		Inst = 2		
VARIABLES	N	mean	min	max
# cities in cluster	5	12.40	2	22
Print Index	5	11.43	3.692	24.92
# Print Cities in 50km	5	2.800	0	5
Population 1400	5	40	3	125
Population 1500	5	29	15	50
Population 1600	5	73.60	27	200
Population 1850	5	549.6	50	2,236
Distance from Port	5	4.114	0	20.57
Growth 1500	5	0.370	-1.273	1.792
Medieval Fair	5	0.600	0	1
Elevation from sea	5	11	3	21

Table 3.4: The table reports results of regression between the 16th century economic institutions and the 15th century diffusion of printing with additional variables especially closeness to the sea. Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2. Printing Penetration is a measure of level of diffusion of printed books in the 15th century in a given city, from the 121 printing cities in the GW database (see section 3.4.4). Closeness Port is the square root of $constant$ - Distance of the city from a sea port (see section 3.4.3). The independent variables of interest are Print Penetration, Closeness to Port and their interaction. Columns 4 has additional controls described in section 3.4.6. Column 2a additionally reports the results without the interaction effect between Print Penetration and Closeness to Port.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(2a) OLS
Print Penetration	0.127***	0.153***	-0.0762*	-0.0617	0.123***
	(0.0369)	(0.0356)	(0.0407)	(0.0427)	(0.0357)
Closeness Port		0.0336***	-0.0208**	-0.0119	0.0342**
		(0.0102)	(0.00967)	(0.0123)	(0.0130)
Print Pentr X Closeness Port			0.0179***	0.0151***	
			(0.00439)	(0.00469)	
log(# cities in cluster)				0.161*	0.195**
				(0.0825)	(0.0801)
Growth 1500				0.0651	0.0991
				(0.132)	(0.139)
log(Population 1500)				-0.0552	-0.0651
				(0.0838)	(0.0898)
$\sqrt{Elevation}$				0.00140	0.00160
				(0.00898)	(0.00942)
Medieval Fair				-0.0975	-0.0478
				(0.139)	(0.142)
Constant	-0.0799	-0.593***	0.133	0.0127	-0.630*
	(0.0749)	(0.171)	(0.108)	(0.352)	(0.354)
Observations	80	80	80	74	74
R-squared	0.099	0.228	0.312	0.360	0.307

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Interaction between Printing and Closeness to the sea

Table 3.4 reports the OLS regression of the dependent variable $Inst_a^{1600}$ with covariates for the whole sample.

$$Inst_a^{1600} = \alpha + \beta_1 SeaPortCloseness_a + \beta_2 PrintPentr_a + \beta_3 SeaPortCloseness_a X PrintPentr_a + \gamma X_a + \epsilon_a \tag{3.2}$$

where $Inst_a^{1600}$ measures the level of market impersonalization in the sixteenth century of a city and X_a is a vector of control variables.

Columns 3 and 4 of Table 3.4 show that an interaction between the transformed variables Printing Penetration and Closeness to Port had a sizable and significant positive effect on the level of market impersonalization in the sixteenth century ($Inst_a^{1600}$). The individual effects of Printing Penetration and the Closeness to Port (Columns 1-3) lose significance when controls on clustering ($Clust_a$), population growth ($Growth1500_a$), population in 1500 ($log(Population1500)$), medieval fair and elevation are added (Column 4). As printing penetration and opportunities independently do not directly affect market impersonalization, the hypothesis developed in section 3.2.2 passes a key test that it is the interaction of the two factors

that mattered for market impersonalization.

Many cities had little or no printing, so any establishment of printing in and around the city increased the per capita printing penetration of the region several times. For example, London (with $PrintIndex_a = 9.33$ bptp) which with Westminster printed several hundred books in the fifteenth century, had ten times more printing penetration per capita than Lisbon (with $PrintIndex_a = 0.81$ bptp) where according to the GW database only 28 books were printed in the fifteenth century. If a relationship-based city was at the sea ($Opp_a = 21.29\sqrt{km}$), and its per capita printing penetration in the fifteenth century doubled, it was about 0.18 steps closer to undergo reform in the sixteenth century.

The fact that high printing penetration does not independently affect market impersonalization shows that a large shock in the number of printed books alone even in large commercial European cities did not make it easier for traders to trade in an impersonal manner, in the absence of clear incentives to initiate trade beyond familiar networks. If the interaction between printing and closeness to port is not considered (Columns 1, 2 & 2a), then both the factors appear individually significant, which further strengthens the argument that an interaction between the two sizable and significant factors was important to consider to understand the nature of the effect of these two factors. Clustering appears to have a large positive effect (Column 4) but only at a 10 percent significance level, and it would have been attributed a larger and more significant positive effect (Columns 2a) had interaction between printing and closeness to port not been considered.

The regression results strengthen the argument that more cities close to a port and enjoying higher levels of printing activity in the fifteenth century, were reforming and turning impersonal in the sixteenth century, and cities that enjoyed none or only one of the two factors stayed relationship-based.

Instrumental Variable- Distance from Mainz

Is it reasonable to assume that printing was exogenous? Maybe, cities like Antwerp that were already more commercially oriented attracted more printing. Such endogeneity may upward bias the results. But, many cities that were adopting printing, like Paris or Rome were traditional economic and political centers that might have been resistant to institutional change, which may downward bias the results. To allay endogeneity concerns, I use a geographic instrumental variable. One of the geographic reasons that determined the level of early printing adoption in Europe was its distance from Mainz, Germany. The printing press was invented in Mainz by Johannes Gutenberg, and most know-how regarding printing press technology was based in the city (held by Gutenberg and his apprentices) during the fifteenth century Europe. The printing press diffused slowly, with cities closer to Mainz getting the technology sooner (Dittmar 2011), and distance from Mainz has been used an instrument by Dittmar (2011) for the level of print adoption. Mainz was not a distinct city in the fifteenth century, such that the distance from Mainz would affect cities in ways other than the adoption of printing. Figure 3.10 plots the logarithm of distance from Mainz on the X-axis, and $PrintPentr_a$ on the Y-axis.

Columns 2 and 4 of Panel A of Table 3.5 report results of the second stage of the 2SLS regressions and the results of 2SLS regression seem to confirm that the interaction between

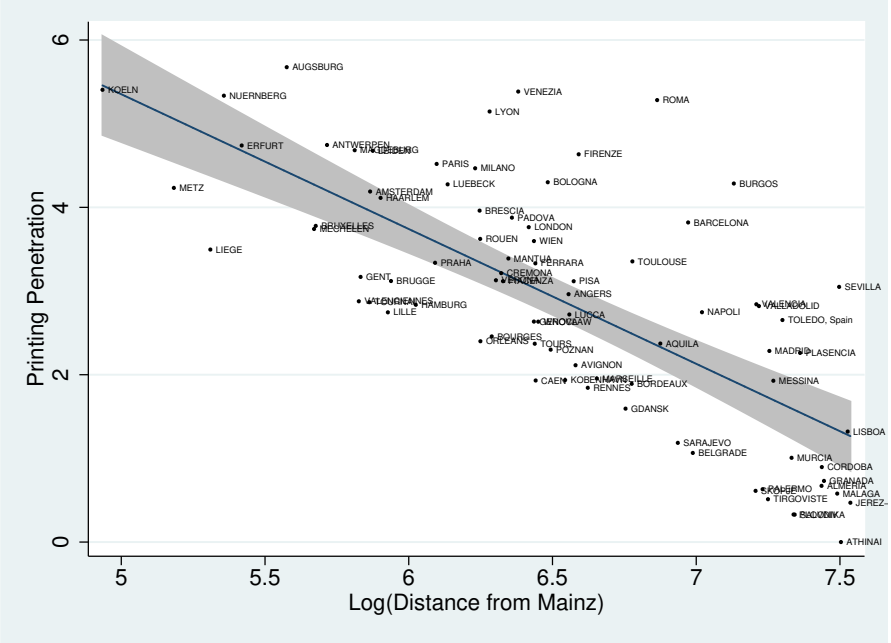

Figure 3.10: The two way plot of cities between variables $PrintPentr_a$ and $log(DistancefromMainz)$ for different types of institutions.

printing and closeness to port were having a sizable and significant positive effect of level of market impersonalization ($Inst_a^{1600}$). Distance from Mainz and its interaction to the distance from a port is a strong instrument with large and highly significant relation to printing as shown in Panel B, with a large F statistic greater than 10. The 2SLS estimate of the effect of interaction between printing and closeness to port (Column 4), is around double the size of the OLS estimate (Column 3). While in the OLS specification, a two-fold increase in the fifteenth century per capita printing penetration led to 0.18 unit increase in the level of market impersonalization in the sixteenth century at cities close to the sea; in the 2SLS specification, a two times increase in the fifteenth century per capita printing penetration led to 0.39 unit increase in the level of market impersonalization in the sixteenth century in cities at the sea. Alternatively, about a six-fold increase in the fifteenth century per capita printing penetration led to a unit increase in the level of market impersonalization in the sixteenth century in the cities at the sea[32].

As the distance from Mainz is a strong instrument, and it is argued that it is reasonable to assume that it followed the exclusivity restriction, the larger estimate can be interpreted as the effect of higher printing and its interaction for cities that adopted printing primarily because of closeness to Mainz. Other factors affected printing adoption, one of which is medieval fair in the city (Column 4a) which had a significant and positive effect on printing adoption, and the 2SLS estimate would exclude the impact of other covariates. Medieval fair in a city seems to negatively affect market impersonalization (Column 4), which hints towards the possibility that medieval traditional trade institutions, that must be more established in medieval fair cities, impeded reform and market impersonalization.

[32] $\Delta Inst_a^{1600} = (-0.111 + 0.0317 * (Opp_a = 21.29)) * ln(\#\ of\ X\ increase\ in\ PrintIndex_a)$ (Column 4 of Table 3.5).

Table 3.5: Columns 2 and 4 of Panel A report results of second stage IV regression between the 16th century economic institutions and the 15th century diffusion of printing of cities at a given distance from the sea. The coefficients are contrasted with results of the OLS regressions reported in Columns 1 and 3. Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2. Printing Penetration is a measure of level of diffusion of printed books in the 15th century in a given city, from the 121 printing cities in the GW database (see section 3.4.4). Closeness Port is the square root of $constant$ - Distance of the city from a sea port (see section 3.4.3). Additional controls are described in section 3.4.6. Panel B reports the coefficients of the first stage regression between endogenous variable Printing Penetration and IV which is log ($constant$ - Distance from Mainz). Columns 2a (4a) and 2b (4b) of Panel B, are the first stage of Column 2 (4) regression of Panel A.

Panel A: Dependent Variable: Institution

	(1)	(2)	(3)	(4)
VARIABLES	OLS	2SLS	OLS	2SLS
Print Penetration	-0.0762*	-0.0933	-0.0617	-0.111
	(0.0407)	(0.0748)	(0.0427)	(0.0770)
Closeness Port	-0.0208**	-0.0539**	-0.0119	-0.0522**
	(0.00967)	(0.0232)	(0.0123)	(0.0261)
Print Pentr X Closeness Port	0.0179***	0.0308***	0.0151***	0.0317***
	(0.00439)	(0.00797)	(0.00469)	(0.00971)
log(# cities in cluster)			0.161*	0.0717
			(0.0825)	(0.0906)
Growth 1500			0.0651	0.00859
			(0.132)	(0.132)
log(Population 1500)			-0.0552	-0.0294
			(0.0838)	(0.0940)
$\sqrt{Elevation}$			0.00140	0.00491
			(0.00898)	(0.0112)
Medieval Fair			-0.0975	-0.251*
			(0.139)	(0.150)
Constant	0.133	0.149	0.0127	0.157
	(0.108)	(0.240)	(0.352)	(0.455)
Observations	80	80	74	74
R-squared	0.312	0.138	0.360	0.175

Panel B: First Stage Regression

	(2a)	(2b)	(4a)	(4b)
Dependent Variable:	Print Penetration	Print Penetration X	Print Penetration	Print Penetration X
VARIABLES		Closeness Port		Closeness Port
Distance from Mainz	1.544***	-2.114	1.971***	1.065
	(0.373)	(5.389)	(0.418)	(6.206)
Mainz Dist X Closeness Port	0.00481	1.814***	-0.0101	1.739***
	(0.0271)	(0.391)	(0.0306)	(0.455)
Closeness Port	-0.00941	1.118**	0.0516	1.597**
	(0.0332)	(0.478)	(0.0436)	(0.648)
log(# cities in cluster)			0.0191	0.298
			(0.152)	(2.261)
Growth 1500			-0.0415	-1.229
			(0.207)	(3.075)
log(Population 1500)			0.241	5.177
			(0.217)	(3.221)
$\sqrt{Elevation}$			0.0583*	0.541
			(0.0292)	(0.434)
Medieval Fair			0.572**	6.331
			(0.281)	(4.176)
Constant	1.402***	1.468	-1.191	-31.87*
	(0.501)	(7.225)	(1.127)	(16.72)
Observations	80	80	74	74
R-squared	0.523	0.693	0.591	0.723
F stat	39.02	39.29	34.51	27.67

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Studying cities nearer to the sea

As Figure 3.9 showed and the OLS and 2SLS regressions in Tables 3.4 and 3.5 confirmed, there is evidence for a strong relationship between the level of market impersonalization ($Inst_a^{1600}$) in the sixteenth century and the interaction between printing and closeness to the sea. I now turn my attention to cities close to the sea. The most distant non-relationship-based city from a port (Brescia, Italy) had a distance of 143km from the port (Table 3.3). I thus now focus on cities that were within a distance of 150km from the sea.

Table 3.6 reports the regression results for the limited sample. Among cities close to the sea, $PrintPentr_a$ continues to have a large and significant positive effect on market impersonalization in the sixteenth century. When the dummy variable $dNearAtlantic_a = dDisrupt_a$ is added to regression, to specifically account for the cities that were closer than 50km from an Atlantic Port, the variable also has a large and significant effect (Column 3). The addition also increases the R square of the regression from 0.268 to 0.444. When an interaction between printing and nearness to an Atlantic Port is considered (Column 4), the interaction has a sizable and significant positive effect, and the R square further increases to 0.503. Among cities close to the sea and on the Atlantic coast a two times increase in the fifteenth century per capita print penetration, increased the level of market impersonalization by 0.3 units in the sixteenth century.

The isolated effect of printing continues to be sizable and significantly positive in the limited sample, but cities near an Atlantic Port did not individually have higher levels of market impersonalization if not enjoying higher print penetration. Also, it is noteworthy that in the restricted sample of cities close to the sea, the effect of urban clustering is insignificant. But, the dummy for medieval fair in a city continues to have a significant negative effect (Column 4), further hinting that established trading institutions in cities with medieval fairs probably impeded the emergence of impersonal economic institutions.

Columns 2 and 4 of Panel A of Table 3.7 report result of the 2SLS regressions with distance from Mainz acting as an instrument for printing penetration. The estimates for the limited sample like in case of whole sample (Table 3.5) show a large and significant effect of printing. The interaction between closeness to Atlantic Port and printing is large but is significant only at 10 percent level. In the 2SLS specification, a two-fold increase in the fifteenth century per capita printing penetration among cities close to the sea and at the Atlantic coast led to a 0.48 unit increase in the level of market impersonalization in the sixteenth century (unlike the 0.3 unit increase in the OLS specification). In other words, about a four-fold increase in the fifteenth century per capita printing penetration among cities at the Atlantic coast led to a unit increase in the level of market impersonalization in the sixteenth century[33].

The instrument is strong (Panel B) and has high F statistic. Once again like in Table 3.5, the higher 2SLS in Table 3.7 estimates can be interpreted as estimating the isolated effect of printing in cities that had high printing penetration primarily because of closeness to Mainz, excluding the effect of covariates like medieval fairs. Like in previous regressions, medieval fairs seem to have a large negative effect on the market impersonalization in the sixteenth century (Column 4).

[33] $\Delta Inst_a^{1600} = (0.289 + 0.407 * (dNearAtlantic_a = 1)) * ln(\#\ of\ X\ increase\ in\ PrintIndex_a)$ (Column 4 of Table 3.7) .

Table 3.6: The table reports results of regression between the 16th century economic institutions and the 15th century diffusion of printing for cities within 150km of sea port. Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2. Printing Penetration is a measure of level of diffusion of printed books in the 15th century in a given city, from the 121 printing cities in the GW database (see section 3.4.4). Additional controls are described in section 3.4.6 and Near Atlantic is a dummy variable denoting whether the closest port to the observed city was an Atlantic Port.

	(1)	(2)	(3)	(4)
VARIABLES	OLS	OLS	OLS	OLS
Print Penetration	0.209***	0.182***	0.181***	0.123***
	(0.0473)	(0.0450)	(0.0438)	(0.0417)
log(# cities in cluster)		0.132	0.118	0.0731
		(0.105)	(0.0870)	(0.0825)
Growth 1500		0.105	0.0153	-0.0873
		(0.213)	(0.174)	(0.165)
log(Population 1500)		-0.104	0.0270	0.110
		(0.109)	(0.113)	(0.110)
$\sqrt{Elevation}$		-0.0139	-0.00187	-0.00267
		(0.0121)	(0.00858)	(0.00845)
Near Atlantic			0.688***	-0.235
			(0.227)	(0.318)
Near Atlantic X Print Penetration				0.318**
				(0.131)
Medieval Fair		-0.106	-0.292*	-0.289**
		(0.178)	(0.150)	(0.139)
Constant	-0.191**	0.121	-0.490	-0.531
	(0.0919)	(0.368)	(0.372)	(0.343)
Observations	59	56	56	56
R-squared	0.201	0.268	0.444	0.503

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Robustness: Alternate variables and specifications

As different factors may drive the rise of impersonal and reforming cities, I also run a multinomial logit regression, and the results are reported in appendix Table E.6. There are two prominent observations to be made. Firstly Columns 2a and 2b show that an interaction between printing penetration and closeness to port is significant in predicting whether a city reforms or turned impersonal when considering the entire sample. If cities close to the sea (within 150km from the sea) are considered (Column 4a), printing increased the chances of a city reforming, while nearness to an Atlantic Port had no significant effect. The observation is in line with the expectation as the North Italian cities reformed, and enjoyed high printing penetration, but they were not close to the Atlantic. In contrast, closeness to the Atlantic exponentially increased the chances that a city impersonalized (Column 4b) (because there are no non-Atlantic cities that impersonal in the sixteenth century), and the effect of printing on market impersonalization of a city was large and positive (which is expected) but not statistically significant at 10 percent level in cities that were close to an Atlantic Port. The loss of significance may be attributed to the small sample of impersonal ($Inst_a^{1600} = 2$) cities. Secondly, urban clustering had a large and significant effect on increasing the chances of reform in a city (Column 2a and 4a), which is not surprising given that all reforming cities were parts of dense urban clusters of Belgium, Nether-

Table 3.7: Columns 2 and 4 of Panel A report results of second stage IV regression between the 16th century economic institutions and the 15th century diffusion of printing for cities within 150 km of sea port. The coefficients are contrasted with results of the OLS regressions reported in Columns 1 and 3. Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2. Printing Penetration is a measure of level of diffusion of printed books in the 15th century in a given city, from the 121 printing cities in the GW database (see section 3.4.4). Additional controls are described in section 3.4.6 and Near Atlantic is a dummy variable denoting whether the closest port to the observed city was an Atlantic Port. Panel B reports the coefficients of the first stage regression between endogenous variable Printing Penetration and IV which is log ($constant$ - Distance from Mainz). Column 2 (4a and 4b) of Panel B, is the first stage of Column 2 (4) regression of Panel A.

Panel A: Dependent Variable: Institutions in cities within 150km of port

	(1)	(2)	(3)	(4)
VARIABLES	OLS	2SLS	OLS	2SLS
Print Penetration	0.209***	0.424***	0.123***	0.289***
	(0.0473)	(0.0848)	(0.0417)	(0.0879)
Near Atlantic			-0.235	-0.494
			(0.318)	(0.648)
Print Penetration X Near Atlantic			0.318**	0.407*
			(0.131)	(0.222)
log(# cities in cluster)			0.0731	-0.00538
			(0.0825)	(0.0972)
Growth 1500			-0.0873	-0.153
			(0.165)	(0.173)
log(Population 1500)			0.110	0.136
			(0.110)	(0.124)
$\sqrt{Elevation}$			-0.00267	-0.00198
			(0.00845)	(0.0119)
Medieval Fair			-0.289**	-0.415**
			(0.139)	(0.167)
Constant	-0.191**	-0.790***	-0.531	-0.905**
	(0.0919)	(0.237)	(0.343)	(0.439)
Observations	59	59	56	56
R-squared	0.201		0.503	0.376

Panel B: First Stage Regression

		(2)	(4a)	(4b)
Dependent Variable:		Print Penetration	Print Penetration	Print Penetration X
VARIABLES				Near Atlantic
Distance from Mainz		1.583***	1.703***	-0.0721
		(0.230)	(0.340)	(0.134)
Distance from Mainz X Near Atlantic			-0.0762	1.344***
			(0.533)	(0.210)
Near Atlantic			0.0353	1.454***
			(0.668)	(0.264)
log(# cities in cluster)			-0.0975	0.0824
			(0.179)	(0.0705)
growth1500			0.00502	0.161
			(0.269)	(0.106)
Growth 1500			0.309	-0.0734
			(0.254)	(0.100)
$\sqrt{Elevation}$			0.0286	0.00197
			(0.0281)	(0.0111)
Medieval Fair			0.474	0.0618
			(0.337)	(0.133)
Constant		1.271***	-0.0389	0.141
		(0.258)	(0.930)	(0.367)
Observations		59	56	56
R-squared		0.453	0.510	0.938
F stat		47.17	16.39	25.39

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

lands, and North Italy. But, noteworthily urban clustering had no significant effect on chances of a city becoming impersonal (Column 2b and 4b). The observation is in line with expectation, as cities like London or Hamburg were not parts of urban clusters and turned impersonal nonetheless. The Medieval fair dummy holds large negative effect (statistically significant in column 2a and 4a, and not significant in column 2b) in all cases (as has been the trend in previous regressions) except Column 2b, where in Column 2b presence of medieval fairs increases the probability of a city became impersonal, which reflects the trends in the summary statistics (Table 3.3) where impersonal cities were on an average more likely to hold medieval fairs. Given the mixed effects of medieval fairs in the results, I conclude it is better not to provide a definitive conjecture on the effect of medieval fairs on economic institutions, and this may be a topic of future research.

Appendix Tables E.7 (considering number of printing cities within 50km from a city ($PrintCity50km_a$) as a proxy for printing), E.8 (considering different thresholds for distance from port and level of clustering), E.9 and E.10 (considering a limited sample of 50 largest cities in the fifteenth century), E.11 and E.12 (considering a limited sample of cities without code R3 or U3) report robustness checks, and the results remain principally unchanged.

When considering the number of printing cities within 50km from a city a as a proxy for printing ($PrintCity50km_a$), its interaction with Closeness to the Sea (Appendix Table E.7 Columns 7 and 8) has no significant effect. The observation can be attributed to the fact that the maximum number of printing cities within 50km of a relationship-based city was 3 (see appendix Figure 3.11), and for a restricted sample of only disruptive cities ($dNearAtlantic = 1$), the maximum number was 2. So, it may be the lack of observation of relationship-based cities that is driving the result. The fact that, all cities enjoying disruptive trade at the Atlantic coast and also having a large number of printing cities around them were non-relationship-based cities, gives strength to the claim that there was indeed an interaction between the two factors.

Robustness: Is aggregate printing a good measure?

Measuring the aggregate number of books being printing around a city provides an estimate of the fifteenth-century printing penetration in the city, which was found to significantly predict institutional reforms in cities closer to the sea, especially on the Atlantic coast. But, printing included all kinds of topics, from religion to art to economy. The soundness of the empirical strategy hinges upon the assumption that printing towns did indeed print books that wrote about new trading practices that helped in reliable impersonal exchange. To test the hypothesis, I collect data on books printed until the fifteenth century from USTC catalog, which classifies early printed books in 37 categories. I consider nine categories of books that can be expected to include books with new ideas and techniques or help individuals in impersonal exchange. The categories are: Academic dissertations, science and mathematics, philosophy and morals, economics (treatises, regulation of guilds) and news books (sensational literature, events, wars) that can be expected to print material with new ideas; and educational books (ABCs, how to write letters, grammars), dictionaries (vocabularies, foreign language instruction), etiquette and courtesy (civil conversation and sumptuary) and travels (topography, maps and navigational manuals) that can be expected to print material teaching about other cultures and traditions. I

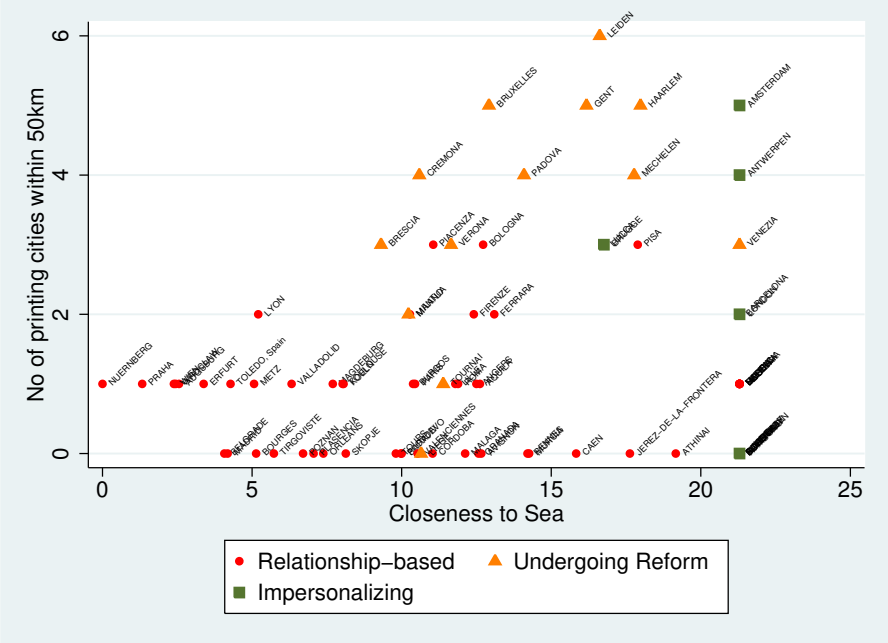

Figure 3.11: The two way plot of cities between variables $PrintCity50km_a$ and $SeaPortCloseness_a$ for different types of institutions.

call the categories- trade friendly categories. Other categories are either related to religion, arts, politics or are industry specific. All the major cities, except Caen, producing books in the trade friendly categories of USTC database were already included in the GW database. I focus on the 121 printing cities of the GW database and develop a new Printing Index for each of the large city a and a printing city b ($PrintIndex_{ab}^{U9}$) considering the number of books printed in the 9 trade-oriented USTC categories (B_b^{U9}), instead of all the books printed in the GW database (B_b).

There is a strong correlation (Figure 3.12) between $PrintPentr_a$, and $PrintPentr_a^{U9}$, i.e., cities which were having a higher printing penetration were also having a high printing penetration for book printing in the nine trade-oriented categories. Considering the ratio ($TradeCatShare_a$) between the $PrintIndex_a^{U9}$ and total number of categorized books in the USTC catalogue $PrintIndex_a^U$, there seems to be no correlation between the original $PrintPentr_a$ variable and $TradeCatShare_a$ (Figure 3.13), showing no systematic bias for or against the 9 USTC categories in minor or major printing towns. These observations suggest that more printing in the fifteenth century was associated with more printing in all categories and that $PrintPentr_a$ is a reasonable variable to measure production of ideas (in general) in a given city, which is exogenously affected by Distance from Mainz.

In appendix Table E.13 I replicate the regressions in Table 3.6, with variable $PrintPentr_a^{U9}$ (which includes only trade friendly books), instead of variable $PrintPentr_a$ (which included all categories of books). The coefficients of trade friendly printing penetration variable $PrintPentr_a^{U9}$ in Table E.13 are larger than $PrintPentr_a$ in Table 3.6, which is expected as printing in trade friendly categories should be more related to market impersonalization in the sixteenth century than printing in all categories. While these coefficients confirm the general argument, the variable $PrintPentr_a$ presents a more unbiased variable to test the hypotheses,

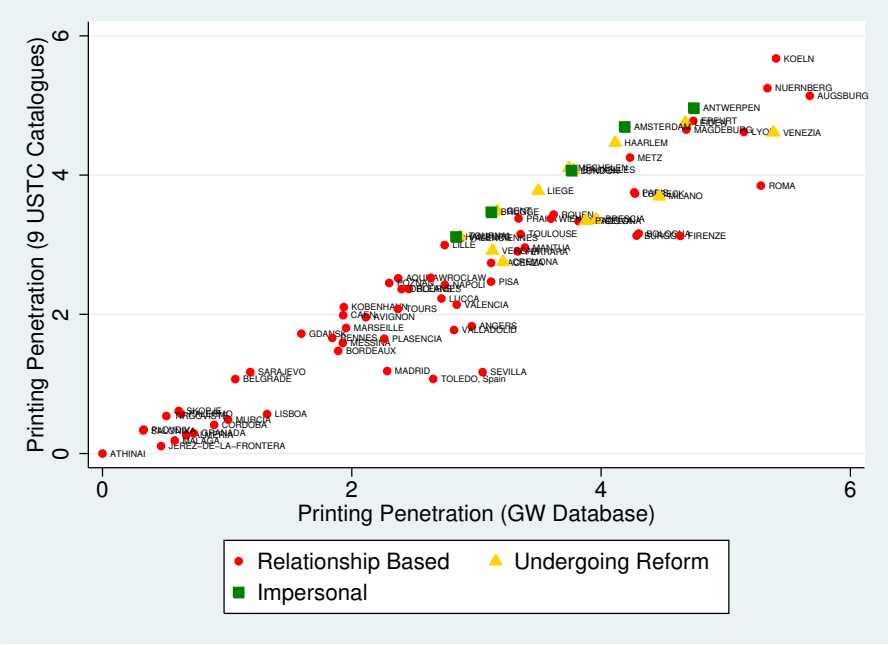

Figure 3.12: The two way plot of cities between variables $PrintPentr_a$ and $PrintPentr_a^{U9}$ for different types of institutions.

and distance from Mainz is a more appropriate instrument for $PrintPentr_a$ because printing only in trade friendly topics may also be demand driven and not just supply driven.

Figure 3.14 shows some additional variation in the content of printing and hints towards the fact that cities with higher rates of printing and close to the Atlantic Coast had a higher share of printing in the 9 USTC categories, which is reasonable as closeness to the Atlantic Coast increased interest in long-distance trade in the fifteenth century Europe and in books that helped in such long-distance trade and travel.

3.5 Conclusion: Emergence of Impersonal Exchange

In this chapter, I explored how impersonal exchange emerges in relationship-based economies, like in the merchant guild dominated economy of medieval Europe. I collected and analyzed city-level data of early modern Europe to answer the following questions. Firstly, why did the relationship-based merchant guild system decline in the Low Countries and England, and why did impersonal market-based exchange emerge? I argued that the opportunity for beneficial long-distance trade with merchants, who were attracted to the growing Atlantic coast, motivated merchants to trade with relatively unfamiliar partners. Moreover, higher availability of printed books in the Low Countries and England, especially of trade-related books, increased information access, accelerated diffusion of Hindu-Arabic numbers, and popularized business practices like double-entry bookkeeping. Also, the integrated postal network of Europe expanded the radius of private communication for all merchants in major European cities. These improvements in trade, information and communication increased the confidence of merchants in impersonal exchange, and under the right utility and reliability conditions, impersonal exchange emerged.

Secondly, why did the transition happen only during the sixteenth century? I argued that

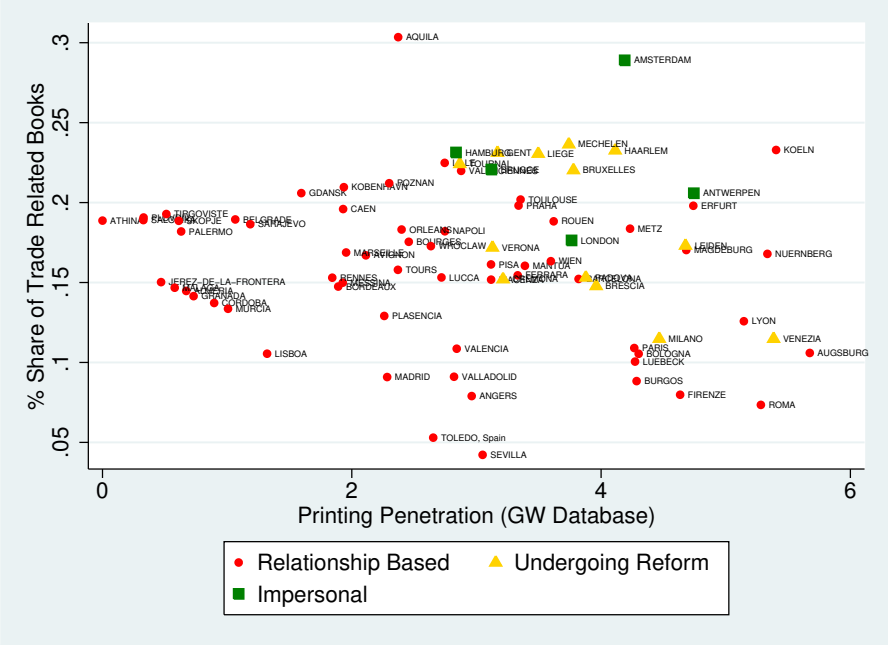

Figure 3.13: The two way plot of cities between variables $PrintPentr_a$ and $TradeCatShare_a$ for different types of institutions.

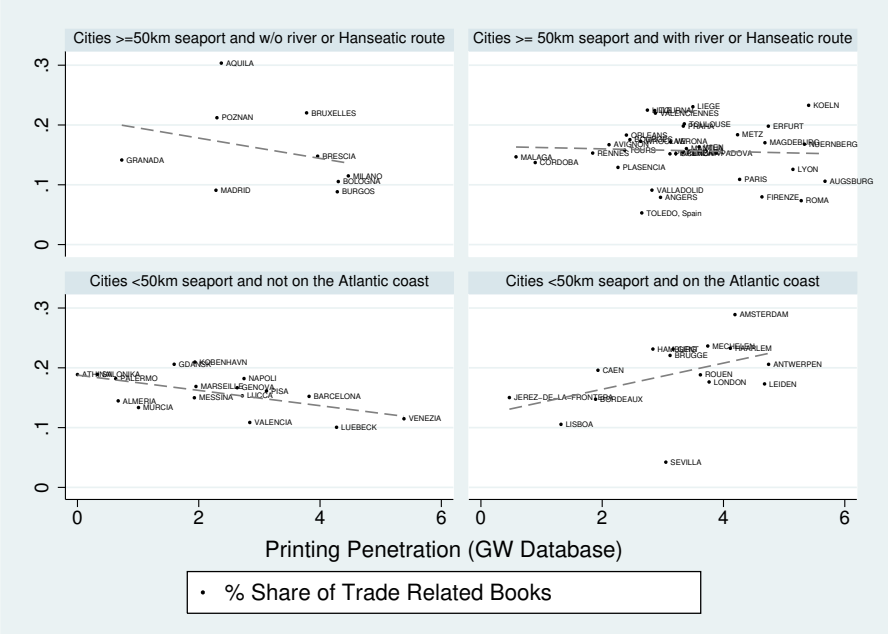

Figure 3.14: The two way plot of cities between variables $PrintPentr_a$ and $TradeCatShare_a$ for different categories of cities.

because (i) trade on the Atlantic coast, (ii) the inter-European Postal Network and (iii) the movable-type printing press, were innovations of the late fifteenth century, the transition only began in the late fifteenth and the early sixteenth century.

Thirdly, why did not other regions also transition to an impersonal system, even if they faced similar initial conditions? I argued that no other region other than Northwestern Europe won the double lottery of high printing penetration and booming trade at the Atlantic coast, and winning of both the lotteries was necessary for the emergence of impersonal exchange.

The chapter looked at early modern Europe, and explored the challenges that are associated with emergence of impersonal exchange in traditional societies where economic exchange is embedded in social relations (Polanyi 1944, Granovetter 1973, Coleman 1988, Greif 1994, Burt 1995, 2001, Padgett & Powell 2012). Societies that are embedded in social relations exhibit considerable institutional (Michalopoulos & Papaioannou 2013, 2014) and cultural persistence (Putnam, Leonardi & Nanetti 1994, Guiso, Sapienza & Zingales 2016, Nunn & Wantchekon 2011), and two notable explanations of persistence have been proposed in the literature. The *cultural approach* (Weber 1905, McCloskey 2016, Mokyr 2016, Putnam, Leonardi & Nanetti 1994, Tabellini 2008b, Greif & Tabellini 2017, Guiso, Sapienza & Zingales 2008a) looks at the differences in people's initial attitudes and shows that these initial attitudes lead to divergent outcomes, where people either stick to or go beyond exchange with familiar contacts. In this chapter, I showed how horizontal diffusion of information and trust enhancing practices can increase reliability in impersonal exchange even in relationship-based economies, and how under right conditions it can eventually lead to the emergence of impersonal exchange. The *institutional approach* (North 1990, Acemoglu, Johnson & Robinson 2005, Greif 2006, Ogilvie 2011, Gelderblom 2013, Rothstein 2011b) has considered the role of quality of institutions on long-term development. The chapter explores how the interaction between information diffusion and economic incentives can increase impersonal exchange, even in the absence of formal legal institutions, which may eventually lead to the development of impartial institutions that serve all parties generally without favoring a particular few.

The chapter showed that network-based institutions like guilds were dominant historically, in a world without formal and impartial legal institutions. The emergence of impersonal markets required a good incentive to go beyond networks and the presence of ample information and reliability-enhancing business practices. While the study focused on the specific setting of sixteenth-century Europe, future studies on similar shocks of information and incentives can provide further evidence whether the hypothesis developed in this chapter about impersonal markets holds true across other geographical and historical settings.

A lack of granular data on guilds and markets of medieval and early modern Europe at a city level means several aspects of trade are not observed, and definitive claims cannot be made about them. If I had better micro-data for the business organizations of the sixteenth century, I could have studied the different effects of information and incentives at a within city level. Future studies with more granular data can enlighten us further about transformations in business organization of early modern Europe. Historically aware studies, that can look at the interaction between incentives and information, and can also help us understand about our globalizing and digitally connected modern world.

Chapter 4

How caste identity became salient in India and influenced networking and governance?

> In India, the real cause of the weakness that cripples our spirit of freedom arises from the impregnable social walls between the different castes. These check the natural flow of fellow feeling among the people who live in our country. The law of love and of mutual respect has been ignored for the sake of retaining an artificial order. This only serves to promote a sense of degeneracy and of defeat. The people of India in this way have built their own cage; but by trying to secure their freedom from one another, they only succeed in keeping themselves eternally captive. - Rabindranath Tagore (1861-1941)

In the previous chapters, I studied how generalized trust evolves, and impersonal markets emerged. We learned that relationship-based social systems like merchant guilds are persistent, and resist change. So, even if the communication and commercial revolutions were individually powerful influences over Europe, it was only under the combined influence of both that merchant guilds declined, and cosmopolitan market-oriented cities like Amsterdam and Hamburg emerged. In the previous chapters, the emphasis was on the importance of trust and information in influencing the social structure of exchange. In this chapter[1] we focus on another critical aspect of exchange - identity. While the previous chapter studied merchant guilds, in this chapter we study another institution like guilds that is salient in India- the caste system. We shall see in this chapter that castes are more than just relationships as they are an intrinsic part of people's social identities. In this chapter, we will examine how social identities of caste and cosmopolitanism get transmitted across generations, and how institutions can influence their transmission in the long run, and influence people's networking patterns, and the quality of political economy and governance.

Individuals have several identity attributes like appearance, gender, skill, lineage or belief.

[1]This chapter is joint work with Paola Sapienza of Kellogg School of Management and Luigi Zingales of University of Chicago Booth School of Business. The co-authors contributed to the research design, while I developed the model, conducted the historical research and data collection, and analysis.

Of these, some attributes become more salient socially than others. These salient attributes can be a source of identity, and social groups may get formed along those attributes that individuals identify with (Akerlof & Kranton 2000, Bénabou & Tirole 2011). For example, if skin color (race) is a salient identity attribute, then individuals may form social groups around and identify with their skin color.

The political engagement of groups formed around the social identity of ethnicity has received considerable attention by social scientists in the past few decades. A region with greater ethnic fractionalization is associated with a poorer quality of government (La Porta, Lopez-de Silanes, Shleifer & Vishny 1999), as fractionalization leads to lower public good generation because ethnic groups may have different preferences (Alesina, Baqir & Easterly 1999). Such differences get further amplified when there exists a greater inequality between different ethnic groups in a country (Alesina, Michalopoulos & Papaioannou 2016). Like ethnicity, when caste fractionalization is high in India, it is associated with a greater acceptance of corrupt candidates in politics (Banerjee & Pande 2007, Banerjee et al. 2014) who promise to pass on specific benefits to voters' caste groups, and with lower production of public goods (Banerjee, Iyer & Somanathan 2005, Munshi & Rosenzweig 2015).

In most of these social identity-related studies, social identity is considered to be exogenously determined. But, ethnic fractionalization or caste categorization can itself be an outcome of politics. Some of the documented ways through which social identity can be altered are through demographic engineering[2] (Glaeser & Shleifer 2005, Friedman & Holden 2008) and passing[3] (Fanon 1967, Jafferlot 2003, Somanathan 2010, Caselli & Coleman 2013, Francis & Tannuri-Pianto 2013, Cassan 2015). Institutions can also influence the salience of particular identity, and even create some new ones (e.g., nationality). Starting with the seminal thesis of Anderson (1991) on "Imagined Communities", there is emerging literature (Glaeser 2005, Blouin & Mukand 2018), which argues that social identity like nationality is endogenous, and is constantly evolving[4]. While certain identity attributes can be made more (Anderson 1991, Glaeser 2005) or less (Blouin & Mukand 2018) salient through deliberate institutional effort, non-deliberate and seemingly "objective" classification of people in specific groups can also socially construct an identity that people then cling on to. The colonial classification of Rwan-

[2]Engineering a change in the demographic makeup of a region is one way in which politics can influence social identity. For example, Glaeser & Shleifer (2005) studied the "curly effect", where politicians engineered the demographic makeup of their constituency, through selectively targeted policies that favored a particular ethnic group and drove others out. Gerrymandering is another method through which politicians can engineer winnable constituencies, by selecting the demographic composition of their constituencies (Friedman & Holden 2008).

[3]Another way in which people's social identity can be altered is through targeted identity - specific benefits that induce people to pass (Caselli & Coleman 2013) to a social identity that receives such benefits. For example, identification of people as black in Brazil (Francis & Tannuri-Pianto 2013), as agrarian castes in colonial India (Cassan 2015) and as disadvantaged castes in India (Somanathan 2010) increased when these groups began receiving specific benefits like reservation in government services. Historically, disadvantaged groups have attempted passing as advantaged groups like whites in the United States (Fanon 1967) or as a "Sanskritized" (more Brahmin like) caste in North India (Jafferlot 2003). But, when people pass from one racial or caste category to another, it does not necessarily alter the overall salience of the racial or caste identity in the society. Passing changes merely the preferred identity sub-group depending on the benefits different groups receive.

[4]Print-capitalism was associated with the promotion of the idea of nationhood, where nations were the "imagined communities" that people shared, leading to a rise of nationalism (Anderson 1991). Recently, Blouin & Mukand (2018) show how radio broadcasts in post-genocide Rwanda helped in reducing the salience of Hutu-Tutsi divisions. Another way in which an identity can become salient is when politicians deliberately promote hate stories that spread xenophobia against particular minority groups (Glaeser 2005).

dans into Hutus and Tutsis (Mamdani 2014), or the measures of "blackness" like the "one drop rule" in the United States (Fanon 1967, Hickman 1997), are some of the examples where the institutions constructed new identities out of arbitrary criteria. One instance where anthropologists like Cohn (1984), Bayly (2001) and Dirks (2001) have elaborately documented the social construction of identity is the colonial evolution of the concept of caste in India about which this chapter will discuss in more significant detail.

In this chapter, we study how social identity can become salient, and look at its effects on the demographically diverse, yet institutionally uniform country of India. We ask if India's colonial institutions could have rigidified the caste identity by making it more salient, and what effects did they have on the general level of cosmopolitanism in the country, and on the nature of networking and governance.

In section 2, we draw from the influential work of Akerlof & Kranton (2000) on identity economics and use the Bisin & Verdier (2001) model of cultural transmission to study how the social identity of caste and cosmopolitanism can get transmitted intergenerationally, and how institutions can shape people's social identities and influence caste identity and cosmopolitanism in the long run. Using tools such as the census, institutions can popularize common prescription of positive or normative attributes that individuals from a given social identity (should) have. Such global prescriptions can increase the conformity to the ideal caste prescriptions and increase the domination of caste identity and marginalization of cosmopolitanism.

In section 3, we discuss the evolution of the concept of caste in India drawing from the work of Dirks (2001), and in sections 4 and 5 we study the effect the 1901 British caste census of India had on the salience of caste in India, and on the nature of out-of-caste networking. We exploit the heterogeneity in the implementation of caste/social precedence tables in the 1901 census of India and find that households in regions where district committees were formed in 1901 to implement detailed caste precedence tables, today (in 2011) have fewer out of community/-caste relationships with political and public officials. These regions also have more households that identify (and pass) as other backward castes, greater presence of caste-based parties, and a poorer quality of government as measured using various proxies of educational and health attainment, and infrastructure provision (section 6).

Our research contributes to several strands of the literature of identity formation, quality of government and public policy for development.

1. **Identity formation:** Primarily, this research contributes to the literature on social construction of identity (Verghese 2016, Lieberman & Singh 2017, Blouin & Mukand 2018). Related papers such as Cassan (2015) and Lee (2015) have looked at the rise of passing and identity politics when colonial bureaucrats categorized some castes undesirably, but neither documents the colonial construction of caste as a category itself. We find that institutions such as colonial bureaucracies can influence people's identities and networks, and these identities can then be persistent, and they impact how people engage with the institutions themselves. The chapter unpacks the long-term impact of an important aspect of Indian colonial history: the colonial construction of the caste system (Dirks 2001) and the impact it has had on contemporary India. A recent paper by Lieberman & Singh (2017) conducted a cross-country analysis where they measured the effect of enumerat-

ing ethnic cleavages (e.g., race, language, tribe and caste) on inter-ethnic violence. They found significant effects at a global level, thus providing even greater evidence in support of this research agenda that studies how categorization of individuals in social groups can influence economic and political outcomes.

2. **Quality of government:** The chapter contributes to the literature that studies the determinants of good governance (La Porta, Lopez-de Silanes, Shleifer & Vishny 1999, Rothstein 2011b). We find that when fractionalized identities like caste are made salient, it negatively affects the quality of government. The novelty of our identification approach is the usage of a historical shock of the 1901 census that made caste more salient in some parts of India than others. In the literature, the salience of identity is assumed to be constant, even though salience of identity can vary between regions. So, two regions which are equally diverse or fractionalized can have different types politics depending on how salient that identity is to the people in the region.

3. **Relevance to policy:** The policy impact of our research is straightforward- policymakers need to be cautious when they develop policies that are specifically targeted for particular identity groups, because such policies may (non-intentionally) make an identity attribute salient and develop newer forms of fractionalization in the society. As partial policies can create a rift within society, impartiality in government services has been prescribed as an important aspect and determinant of good governance (Rothstein 2011b). Our research can offer some caution to policymakers that they should be aware of the impact their policies have on people's sense of self and their identity.

4.1 Theory: Intergenerational transmission of social identity

In this section, we develop a model of intergenerational transmission of social identity drawing from the works of Akerlof & Kranton (2000) and Bisin & Verdier (2001), on identity economics and cultural transmission. In Akerlof and Kranton's framework, individuals are associated with social categories, C, and these social categories are associated with prescriptions P that indicate the appropriate behaviour and attributes for individuals in different social categories. In this framework individuals may have varying tastes, yet due to these prescriptions, they are incentivized to 1) follow the prescriptions associated with their categories even if the prescriptions are incongruent to their tastes and 2) punish those individuals who belong to their categories but violate the prescriptions associated with it.

We combine the Akerlof & Kranton (2000) framework with the cultural transmission model developed by Bisin & Verdier (2001), which considers the effects of vertical and horizontal transmission of values. While vertical transmission of values occurs through parents, the horizontal transmission of values can occur through peers, neighbors or the state.

4.1.1 The "ideal" prescription

We first reinterpret the Akerlof & Kranton (2000) prototype model in the context of the caste system in India.

- **Activities:** There are two possible activities, e.g., having a special meal (Activity one) and having a regular meal (Activity two). There is a population of individuals each of whom has a taste for either special or regular meal. If a person's tastes (i) and actions match, they earn a utility V_i. But, if a person with a taste for regular-meal (special-meal) chooses special-meal (regular meal), they earn a utility $zero$.

- **Categories:** Individuals also belong to social categories e.g. "clean" and "non-clean" castes. For simplicity, we assume all persons think of themselves and others as "clean" castes.

- **Prescriptions:** Moreover, there exists an "ideal" behavioural prescription: *a "clean" caste individual has special-meal* (in contrast to "non-clean" castes who have regular meal). More generally, a behavioural prescription for being a "clean" caste can be adherence to a particular lifestyle which includes a series of dos and do nots, such as dining only with own caste's members, having a particular occupation, and marrying only with members of one's caste.

- **Internalized component:** Individuals follow the prescriptions taught to them in childhood (either vertically through parents or horizontally through socialization), and a follower of the 'ideal" prescription suffers a loss I_s of their "clean" caste identity on having regular-meal (s for self).

- **Externalized component:** In addition, there are identity externalities like in Akerlof & Kranton (2000). If i and j are paired, and i has regular meal, it diminishes j's "clean" caste identity by an amount I_o (o for other), if j follows the prescription of having special-meal.

- **Vigilantism:** If i has regular-meal, j may act as a vigilante and respond, by punishing i at a cost c and causing a loss to i of amount L. Such vigilantism will restore j's identity as j punished the prescription violator i[5].

When a "clean" caste individual j with a taste for special-meal (caste-following) meets a "clean" caste individual i with a taste for regular-meal (cosmopolitan), the following outcomes are possible.

- **Self enforcement:** If i has high self-internalization of the prescription ($V_{cosmo} < I_s$) i self enforces themselves into having the special meal.

- **No enforcement:** If i has low self-internalization of the prescription such that they are willing to be punished for having the regular-meal ($V_{cosmo} \geq I_s + L$) then no enforcement can force them into having special-meal.

[5]If individuals of different categories - "clean" and "non-clean" castes - make the society, and some fraction of both adhere to the ideal prescriptions of special and regular meal respectively, then 1) if i and j are both of the same categories, then j can punish i for violating the respective prescription, and 2) if i and j are of different categories, then j can punish i for following the prescription exclusive to j's caste. In both scenarios i's that violate the ideal prescriptions of their associated categories get punished by those j's who strongly believe that $i's$ should follow i's own caste's prescriptions.

- **Strong and weak social enforcement:** Also, there exists an intermediary region ($I_s < V_{cosmo} < I_s + L$) where vigilantism by j influences i's decision. When j is more likely to feel a large other-identity loss ($I_o > c$) and act as a vigilante by levying a large punishment on i, i chooses to have a special meal to avoid such punishment (strong enforcement). In contrast, if the other-identity loss of j is small when compared to the cost of sanctioning ($I_o < c$), i chooses to have a regular meal in the absence of any punishment (weak enforcement).

4.1.2 A model of cultural transmission

Given the setup, we now construct a society where individuals live for two periods as children (first period) and parents (second period). Parents and children overlap for one period, and parents decide whether to put effort into passing on their tastes to their children who learn their tastes in their first period. Following the model of Bisin & Verdier (2001) on cultural transmission, we assume that parents have imperfect empathy, i.e., they desire the welfare of their children, but consider their child's utility based on their tastes. So, for example, a parent who has special-meal (regular-meal) assumes that if their child consumes regular-meal (special-meal), the child's utility will be $zero$.

Children acquire their tastes by two means. Firstly, these tastes can be taught (vertically) by their parents. Secondly, these tastes can be acquired (horizontally) through socialization with strangers (Boyd & Richerson 1988, Bisin & Verdier 2001, Mokyr 2016). At any time t, if parents do not put any effort in teaching their taste i, then the probability that their child will have the same tastes as them is proportional to the fraction of population with parent's tastes ($q_i(t)$). In contrast if parents put effort into teaching their children their tastes at time t, then it will cost the parents an amount $\frac{\phi d_i(t)^2}{2}$, but children will acquire the taste of their parents with probability $d_i(t)$ through vertical transmission, or with probability $(1-d_i(t))q_i(t)$ through horizontal transmission. If there are two types of parents i and j in the population, the rate of change in the population of type i at any given time t is given as

$$\frac{dq_i(t)}{dt} = q_i(t)q_j(t)(1-d_j(t)) - q_i(t)q_j(t)(1-d_i(t)) = q_i(t)q_j(t)(d_i(t)-d_j(t)) \quad (4.1)$$

Three stationary points for equation 4.1 exist: $q_i = 0, q_i = 1$ and q_i s.t. $d_i = d_j$ (see Bisin & Verdier (2001) for more details on the dynamics). For simplicity we drop the time component t, where all populations (q_i) and efforts (d_i) are at a specific time t.

Parents in a given caste can be of two types: those who have tastes aligned with their prescriptions (caste-following), and those who do not (cosmopolitan). Consider only one caste of "clean" individuals where the prescription is for "clean" members to only have a special meal. Let q_{caste} caste-following-parents have their tastes aligned with this caste prescription at time t, while $q_{cosmo} = 1 - q_{caste}$ cosmopolitan-parents have their tastes deviate from the prescription. Cosmopolitan-parents do not follow prescriptions and have a taste for regular meal and have no internalized disutility $I_s = 0$ from having such a regular meal.

Caste-following-parents' problem

If children of caste-following-parents have tastes aligned with the prescriptions, their utility (from caste-following-parent's perspective) as adults comes from following their taste of having a special-meal (V_{caste}), minus the externalized disutility (I_o) from meeting a cosmopolitan (consuming regular-meal) with probability q_{cosmo}[6]. Alternatively, if the externalized disutility (I_o) is greater than cost c of sanctioning a cosmopolitan (function $\mathbb{F}(I_o > x)$ gives the probability with which $I_o > x$), then the children will act as a vigilante to strongly sanction the cosmopolitan, and so their maximum externalized disutility will be c. In contrast, if children of caste-following-parents have tastes unaligned with the prescriptions (i.e., they have regular meal), their utility from caste-following-parents' perspective is negative, as its a sum of the internalized shame $-I_s$ that the parents perceive the children will have, and the sanction cost $-L$ that the children will have to endure as a cosmopolitan, if they come across a caste-following-individual (with probability q_{caste}) whose disutility (I_o) is greater than the cost of sanctioning (c).

The net utility from making an effort[7] d_{caste} at the cost of $\frac{\phi_{caste} d_{caste}^2}{2}$ for a caste-following-parent is expressed as

$$\Delta^e_{caste} = d_{caste} q_{cosmo}(V_{caste} + I_s + q_{caste} L\mathbb{F}(I_o > c) - q_{cosmo} min(I_o, c)) - \frac{\phi_{caste} d_{caste}^2}{2} \quad (4.2)$$

As caste-following-parents maximize their utility the optimal d^*_{caste} is given to be (where we assume ϕ_{caste} is s.t. $d^*_{caste} < 1$)

$$d^*_{caste} = max(q_{cosmo}(V_{caste} + I_s + q_{caste} L\mathbb{F}(I_o > c) - q_{cosmo} min(I_o, c))/\phi_{caste}, 0) \quad (4.3)$$

Cosmopolitan-parents' problem

If children of cosmopolitan-parents have tastes unaligned with the prescriptions, their utility (from cosmopolitan-parent's perspective) comes from following their taste of having regular meal (V_{cosmo}) minus the sanction cost L that the children will have to endure as cosmopolitans, every time they come across a caste-following-individual (with probability q_{caste}) whose disutility (I_o) is greater than the cost of sanctioning (c). The utility of children who follow the

[6]Because parents have imperfect empathy, they evaluate their child's utility based on current conditions i.e. $q_{cosmo}(t)$ in their period, and not in the period when their children are parents ($q_{cosmo}(t+1)$). While a child as an adult (parent) may have several meals and come across individuals of opposite tastes several times, for simplicity, we condense having meals and meeting individuals of opposite tastes to one act each period.

[7]The utility for a caste-following-parent from putting an effort d_{caste} in teaching their tastes to their child is given as U^e_{caste}.

$$U^e_{caste} = (V_{caste} - q_{cosmo} min(I_o, c)(d_{caste} + (1 - d_{caste}) q_{caste}) - (I_s + q_{caste} L\mathbb{F}(I_o > c))(1 - d_{caste}) q_{cosmo}$$

Similarly, the utility for a caste-following-parent from not putting an effort in teaching their tastes to their child is given as U^{ne}_{caste}.

$$U^{ne}_{caste} = (V_{caste} - q_{cosmo} min(I_o, c)) q_{caste} - (I_s + q_{caste} L\mathbb{F}(I_o > c)) q_{cosmo}$$

prescription is zero from the perspective of cosmopolitan-parents as the utility of the children from having a special meal (from parent's perspective) is $zero$.

The net utility from making an effort[8] d_{cosmo} at the cost of $\frac{\phi_{cosmo}d_{cosmo}^2}{2}$ for a cosmopolitan-parent is expressed as

$$\Delta^e_{cosmo} = d_{cosmo}q_{caste}(V_{cosmo} - q_{caste}L\mathbb{F}(I_o > c)) - \frac{\phi_{cosmo}d_{cosmo}^2}{2} \tag{4.4}$$

As cosmopolitan-parents maximize their utility the optimal d^*_{cosmo} is given to be (where we assume ϕ_{cosmo} is s.t. $d^*_{cosmo} < 1$)

$$d^*_{cosmo} = max(q_{caste}(V_{cosmo} - q_{caste}L\mathbb{F}(I_o > c))/\phi_{cosmo}, 0) \tag{4.5}$$

4.1.3 Weak and strong vigilantism scenarios

The cultural transmission model provides the following proposition.

Proposition 3 *Based on equation 4.1, the populations of caste-following-individuals (q_{caste}) which satisfy the condition $d^*_{caste} = d^*_{cosmo}$, are stationary points, where optimal d^*_{caste} and d^*_{cosmo} are given by equations 4.3 and 4.5. The stability of the stationary points is determined by the functional form of d^*_{caste} and d^*_{cosmo} w.r.t q_{caste}.*

Effort by cosmopolitan-parents d^*_{cosmo} is zero when all in the population are cosmopolitans ($q_{caste} = 0$), and positive when the rest of the population is caste-following ($q_{caste} = 1$) if $V_{cosmo} > L\mathbb{F}(I_o > c)$. Similarly effort by caste-following-parents d^*_{caste} is zero when all in the population are caste-following ($q_{caste} = 1$), and positive when the rest of the population is cosmopolitan ($q_{caste} = 0$) if $min(I_o, c) < V_{caste} + I_s$. We assume $d^*_{caste} > 0$ at $q_{caste} = 0$, i.e. caste-following-parents have a positive utility in teaching their children to have special-meal, in a society made of cosmopolitans ($min(I_o, c) < V_{caste} + I_s$).

Consider a *weak-vigilantism scenario* when the expected loss in a caste-following society ($q_{caste} = 1$) due to strong sanctioning by a vigilante ($L\mathbb{F}(I_o > c)$) is less than the utility gained from having a regular-meal (V_{cosmo}) for cosmopolitans. At $q_{caste} = 1, d^*_{cosmo} > d^*_{caste} = 0$, when $L\mathbb{F}(I_o > c) < V_{cosmo}$. So, the stationary point $q_{caste} = 1$ is not a stable equilibrium. At $q_{caste} = 0$, we assumed $d^*_{caste} > 0$ i.e. $d^*_{caste} > d^*_{cosmo} = 0$, so the stationary point $q_{caste} = 0$ is not a stable equilibrium either. Then, the two curves intersect at a unique value q^*_{caste}, where $d^*_{caste} = d^*_{cosmo}$, and q^*_{caste} is a unique stable heterogeneous equilibrium where both caste-following-individuals and cosmopolitans co-exist. Figure 4.1 plots this scenario.

[8]The utility for a cosmopolitan-parent from putting an effort d_{cosmo} in teaching their tastes to their child is given as U^e_{cosmo}.

$$U^e_{cosmo} = (V_{cosmo} - q_{caste}L\mathbb{F}(I_o > c))(d_{cosmo} + (1 - d_{cosmo})q_{cosmo})$$

Similarly, the utility for a cosmopolitan-parent from not putting an effort in teaching their tastes to their child is given as U^{ne}_{cosmo}.

$$U^{ne}_{cosmo} = (V_{cosmo} - q_{caste}L\mathbb{F}(I_o > c))q_{cosmo}$$

Now, consider a *strong-vigilantism scenario* when the expected loss in a caste-following society ($q_{caste} = 1$) due to strong sanctioning by a vigilante ($L\mathbb{F}(I_o > c)$ is larger than the utility gained from having a special meal (V_{cosmo}) for cosmopolitans. In the neighborhood of $q_{caste} = 1, d^*_{caste} > d^*_{cosmo} = 0$, when $L\mathbb{F}(I_o > c) > V_{cosmo}$. So, the stationary point $q_{caste} = 1$ is a stable equilibrium. In this scenario, the two curves d^*_{cosmo} and d^*_{caste} do not necessarily intersect, and a stable heterogeneous equilibrium may not exist. Figure 4.2 plots this scenario.

The weak and strong enforcement conditions provide three observations:

- Increasing loss due to sanctioning (L) reduces the possibility that the system will be in the weak-vigilantism scenario, reduces the possibility that a stable heterogeneous equilibrium necessarily exists, and increases the possibility of a homogeneous equilibrium made up of caste-following-individuals ($q_{caste} = 1$).

- Increasing cost of vigilantism (c) increases the possibility that the system will be in the weak-vigilantism scenario, increases the possibility that a stable heterogeneous equilibrium necessarily exists, and reduces the possibility of a homogeneous equilibrium made up of caste-following-individuals ($q_{caste} = 1$).

- When a heterogeneous equilibrium exists, the population of cosmopolitans (q_{cosmo}) decreases with increasing L and decreasing c.

As the loss due to sanctioning (L), and the cost of sanctioning (c) depend on the quality of the rule of law in a state, these observations lead us to the following corollary

Corollary 4 *As a state improves in protecting individuals from vigilantes by making it costly for vigilantes to sanction individuals (increasing c), and by protecting individuals from the harmful effects of vigilantism (decreasing L), the society is less likely to be in the homogeneous caste-following equilibrium, and more likely to be the heterogeneous equilibrium, with a higher population of cosmopolitans (q_{cosmo}) who can co-exist with caste-following-individuals.*

4.1.4 Popularization of "ideal" prescription

The census is one of the tools through which such "ideal" prescriptions get popularized by the state, by providing parameters (adherence to "ideal" prescriptions) based on which individuals can be categorized into groups. Note that these prescriptions do not need to be invented, but simply be made more salient. If a state through a tool like the census makes "ideal" caste prescriptions salient, children are more likely to come across the "ideal" prescriptions and become their followers. We consider a simple setting where such promotion of "ideal" prescription i by the state, creates a drift where $d_s q_i$ children of the whole population always become caste-following-individuals (more the followers of the ideal prescription (q_i) the stronger the strength (d_s) of the "ideal" prescription). This drift alters the equation 4.1, as now we have

$$\frac{dq_i}{dt} = q_i q_j(1 - d_j - d_s) - q_i q_j(1 - d_i - d_s) + q_j q_i d_s = q_i q_j(d_i + d_s - d_j) \tag{4.6}$$

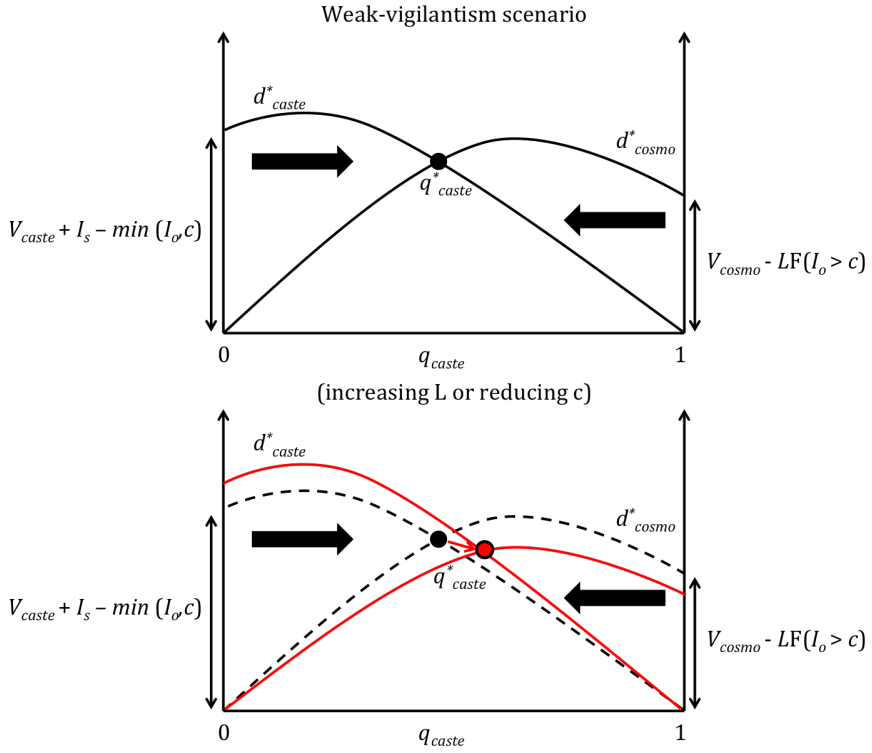

Figure 4.1: The figure plots the optimal effort by caste-following (d^*_{caste}) and cosmopolitan (d^*_{cosmo}) parents as a function of the population of caste-following-individuals (q_{caste}), when $L\mathbb{F}(I_o > c) < V_{cosmo}$ i.e. in weak-vigilantism scenario. The population of caste-based-individuals in the heterogeneous equilibrium (q^*_{caste}) is increasing in increasing L (which reduces d^*_{cosmo}) and decreasing c (which increases d^*_{caste} and decreases d^*_{cosmo}).

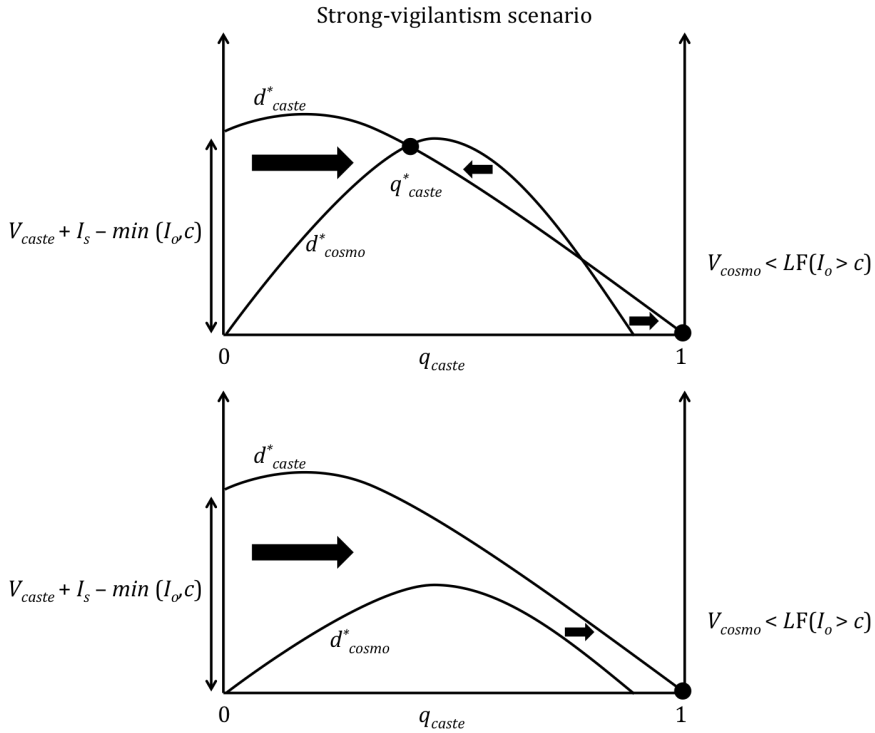

Figure 4.2: The figure plots the optimal effort by caste-following (d^*_{caste}) and cosmopolitan (d^*_{cosmo}) parents as a function of the population of caste-following-individuals (q_{caste}), when $L\mathbb{F}(I_o > c) > V_{cosmo}$ i.e. in strong-vigilantism scenario. The population need not converge to a unique heterogeneous equilibrium q^*_{caste}, even though a homogeneous equilibrium where $q_{caste} = 1$ always exists.

In this scenario, the stationary points are $q_i = 0, q_i = 1$ and q_i s.t. $(d_j - d_i) = d_s$. In the case of caste-following-individuals and cosmopolitans, at the stationary point $d^*_{caste} + d_s = d^*_{cosmo}$. The dynamics provides the following proposition

Proposition 4 *Based on equation 4.6, the populations of caste-following-individuals (q_{caste}) which satisfy the condition $d^*_{caste} + d_s = d^*_{cosmo}$, are stationary points, where optimal d^*_{caste} and d^*_{cosmo} are given by equations 4.3 and 4.5. The stability of the stationary points is determined by the functional form of d^*_{caste} and d^*_{cosmo} w.r.t q_{caste}.*

For a heterogeneous equilibrium to necessarily exist, 1) $V_{cosmo} > L\mathbb{F}(I_o > c) + d_s$ (weak-vigilantism scenario) and 2) $min(I_o, c) > V_{caste} + I_s + d_s$. The new dynamics gives rise to the following observations, which is plotted in Figure 4.3.

The presence of a drift (d_s) towards becoming a caste-following-individual provides the following observations:

- A drift (d_s) due to the popularization of "ideal" prescription reduces the possibility that the system will be in the weak-vigilantism scenario, reduces the possibility that a stable heterogeneous equilibrium necessarily exists, and increases the possibility of a homogeneous equilibrium made up of caste-following-individuals ($q_{caste} = 1$).
- When a heterogeneous equilibrium exists, population of cosmopolitans (q_{cosmo}) decreases in increasing d_s.

Corollary 5 *If a state popularizes "ideal" caste prescriptions ($d_s > 0$), the society is more likely to be in the homogeneous caste-following equilibrium, and less likely to be the heterogeneous equilibrium, with a lower population of cosmopolitans (q_{cosmo}) co-existing with caste-following-individuals.*

4.1.5 Effects of popularization of "ideal" prescriptions

Corollary 5 shows that the popularization of "ideal" prescription decreases the cosmopolitan population and increases the domination by caste-following-individuals in the population. A more homogenous population comprising of caste-following-individuals can lead to 1) an increased within-caste networking, at the cost of out-of-caste networking (i.e., more relative within-caste social capital), and 2) can increase the formation of social identity-based groups (e.g., *caste sabhas*) where individuals can follow their caste prescriptions collectively (e.g., cooking a special meal), and collectively advance their special interests (e.g. by forming caste-based parties).

In such caste-based societies, the willingness and capacity of the state to reduce vigilantism (Corollary 4) may be lower, which further hinders the co-existence of cosmopolitan populations. In section 2.2.5 on identity, parochialism, and honesty of the second chapter, we saw how identity specific benefits to identity-based and partial traders can reduce the scope in which impersonal exchange can emerge (Corollary 3). As caste identity becomes more salient, and cosmopolitanism declines, it may hinder impersonal exchange and the emergence of impersonal institutions (Proposition 2).

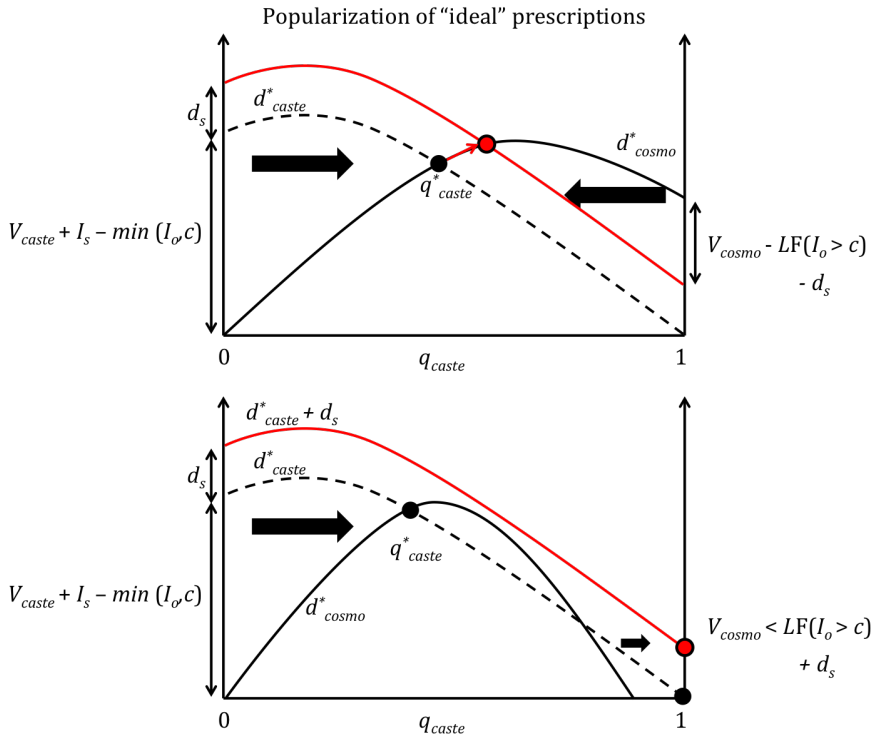

Figure 4.3: The figure plots the optimal effort by caste-following (d^*_{caste}) and cosmopolitan (d^*_{cosmo}) parents as a function of the population of caste-following-individuals (q_{caste}), for weak ($V_{cosmo} > L\mathbb{F}(I_o > c) + d_s$) and strong-vigilantism ($V_{cosmo} < L\mathbb{F}(I_o > c) + d_s$) scenarios, where popularization of "ideal" prescription creates a drift (d_s). The population of caste-based-individuals in the heterogeneous equilibrium (q^*_{caste}) is increasing in increasing d_s, and under strong vigilantism scenario the possibility of a stable heterogeneous equilibrium reduces, even though a homogeneous equilibrium where $q_{caste} = 1$ always exists.

In the next sections, we will see how the notion of caste evolved in India, and how especially British censuses like that of 1901 played a role in increasing the organization of society along caste lines.

1. **Census and caste prescriptions:** The 1901 census of India especially attempted to rank castes on a social precedence table based on criteria determined by prescriptions of the 1901 census director Herbert Risley and on deliberation with district committees of "native gentlemen". The prescriptions for higher social precedence often included an implicit criterion of similarity of castes to *Brahmin* and *Rajput* ideals (the highest priestly and warrior castes). For example, the castes of *Jats* and *Kurmis* were denied a higher Kshatriya (warrior caste) status in the social precedence in the 1901 United Provinces provincial census, on the basis that they allowed widow remarriage, unlike the "twice" born castes where widow remarriage was prohibited.

2. **Emphasis on Sanskritic heritage:** Such formalization of prescriptions gave rise to caste groups whose primary goal was to emphasize and promote their Sanskritic heritage and to raise their social precedence by petitioning to the government officers. Discussing the impact of the 1901 caste ranking on castes, a 1904 author from Allahabad, Kumar Cheda Singh Varma, wrote in the book *"Kshatriyas and would-be Kshatriyas"*:

> About three years ago the Census Commissioner in India (The Hon. Mr. H

> H. Risley, C.S , C.S.I.) directed that, for the Census of India, 1901, a scheme should be drawn up classifying the various Hindu castes under the four groups of Brahmans, Kshatriyas, Vaishyas, and Sudras ; and assigning to each caste in these groups its proper position according to the order of social precedence The members of certain castes which have, hitherto, in an undefined sort of way, been striving to rise in the social scale eagerly seized the opportunity thus afforded them of having their pretensions to a higher status placed on record and, if possible, stamped with the hallmark of official recognition. In the controversy that ensued no subject was more warmly discussed, on both sides, than the claims of certain castes to be considered "twice-born" and, in particular, to belong to the Kshatriya order. Among the numerous aspirants to Kshatriyanic honours, who claim equality with the descendants of the Kshatriyas of old, are found the Khatris, the Kayasthas, the Jats, and the Kurmis.

3. **Caste-based organization:** This trend where intermediary castes such as *Ahirs* (who emphasized a Sanskritic name *Yadavas*) in North India asserted their higher status by forming caste groups (e.g., *Ahir Yadav Kshatriya Mahasabha* in 1910) has been associated with the phenomenon of Sanskritization (Jafferlot 2003), where castes form groups and attempt to enhance their status by accepting Sanskritic ideals (such as vegetarianism)[9]. Such organization along caste lines by intermediary castes like *Yadavas* has ultimately led to the formation of caste-based parties like the Rashtriya Janata Dal and Samajwadi Party.

4.2 History: Evolution of the term caste

Caste in India is a complex classification scheme. The term is significant in modern social, economic and political discourse in India. The term's etymological roots lie in Portuguese, and the term is related to indigenous classification schemes like *Jati*, *Qaum* and *Varna*. Could the colonial state have made this identity salient?

4.2.1 Qaum and Varna: Pre Colonial India

The Mughal administrative book *Ain-I-Akbari* written around 1590 makes reference of ethnic clans called *Qaums* (*Jatis*), when listing the ethnicity of local landlords called *zamindars*[10]. *Qaums* holding these positions of power included Muslim clans like *Sayyids*, traditional warrior clans such as *Rajputs*, priestly groups of *Brahmins*, administrative classes like *Kayasthas*, and agrarian classes such as *Ahirs* and *Kunbis*. Habib et al. (2013) discussed the origins of influence

[9]The 1901 census had a note on the Sanskritization of some *Kurmi* castes of Bihar

> The Awadhiya or Ayodhya Kurmis pride of Bihar and the Kanaujia Kurmis of the United Provinces pride themselves on prohibiting the remarriage of widows and are endeavoring to establish a shadowy title to be recognized as some variety of Kshatriya. But although the Awadhias have achieved complete practical separation from the main body of Kurmis, no one accepts them as Kshatriyas or Rajputs.

[10]The power of *zamindars* has been speculated to be anywhere between tax collectors to influential overlords (see Habib et al. (2013)).

of these different *Qaums*, and presented *zamindari Qaums* to be in competition with each other, vying for domination over land. He also discussed the existence of a market where *zamindari* rights could be sold and bought by potentially wealthy individuals.

In another volume of the same book, *Ain-I-Akbari* also discussed the Hindu philosophy and mentioned the ritualistic social classification of Hindu society called the *Varna* system. In this system the society is divided into four hierarchical *Varnas* (groups), where *Brahmins* (priests) are at the top, followed by *Kshatryas* (warriors), *Vaishyas* (businessmen) and *Shudras* (laborers and artisans), followed by a fifth group of *Malecchas* (out-castes, that included foreigners). Each of these *Varnas* had strict codes of occupation, diet and lifestyle and the intermingling between *Varnas* was regulated. Moreover, while this classification has been claimed to be nonhereditary in earliest Hindu scriptures, this book presents *Varna* as a hereditary social order. So, Mughal era presents two forms of classification, the first was the observed social grouping of *Qaums*/*Jatis*, and the second was the ritualistic hierarchical scheme of *Varna*.

4.2.2 Caste and Colonial Era

Sanskrit theological texts like the *Manusmriti* (which translates as "recollections of Manu[11]" and can be interpreted as the law of *Manu*), written between 200 BC and 300 AD, propounded a Hindu society based on a rigid *Varna* order where member of each *Varna* followed their respective duties. But, the significance of texts like *Manusmriti* in ancient and medieval India is doubted (Buxbaum 1968) as local Hindu laws have mostly been customary (Lariviere 1989). Although the significance of the *Manusmriti* in pre-British India is doubtful, its significance with the beginning of the British rule is clear. It was one of the first Sanskrit books to be translated into English in the late eighteenth century, and it shaped the view of earliest British and other Orientalists.

When the British East India Company gained a foothold in India in the second half of the eighteenth century (beginning with Bengal), the company struggled with administering a diverse and alien country like India. So, the colonial administrators instituted standard personal laws in 1772 for Muslims on the basis of the *Quran*, and in parallel for the Hindus on the basis of "the *Shastra*" (Hastings 1772). The colonial administrators took these Sanskrit texts literally and sought uniformity in the Hindu laws even though the Hindu laws were customary[12].

The British, inspired by these texts, conceptualized the existing Indian society to be divided on the basis of "castes" (a loanword they took from the Portuguese), which was imagined to be

[11] *Manu* is believed to be the progenitor of humanity in Hindu texts, similar to Adam in Abrahamic traditions.

[12] Lariviere (1989) reflecting on the complex system of *Dharmashastras* highlighted its customary nature and the inability of the British to grasp it:

> The *dharmashastras* was not internally consistent, nor was it uniformly applicable to all Hindus. What the British did not understand was that no legal system claiming to be based on a fixed set of texts (in this case the Veda via the interpretations of the *dharmashastras*) could remain viable without a means to adapt the dictates of those texts to the ever-changing needs of society. This adaptation had been done for centuries by the commentators and digest writers. ...This situation was anathema to the British responsible for the administration of justice in India, and it was the motive behind Hastings's decision to commission the composition of *A Code of Gentoo Laws, Ordinations of the Pundits*. It remained a central concern of the British administrators to eliminate the inconsistency and uncertainty they saw in the indigenous legal systems in India.

based on the ritualistic *Varna* hierarchical order, where the different *Jatis*/*Qaums* were the several sub-*Varnas*. With time, the British became more acquainted with Indian society especially by conducting local surveys. Through the surveys they began to discover the inconsistencies in their long-held view that caste was a detailed *Varna* order (Samarendra 2011). Census officer from North-Western Provinces, F.S. Growse, wrote in 1872:

> Impartially judged by either standard, the authority of the Code [Manusmriti] will be found materially shaken. Its theories of origin are as devoid of Vedic confirmation as its pictures of existent society are irreconcilable with the testimony of all independent literature, whatever the age in which it was produced. If such a clearly defined four-fold division ever existed, how happens it that one-half[13] of the division remains in full force to the present day, while the other moiety has sunk into absolute oblivion.

Although a clear mapping of castes into *Varna* was difficult, this did not deter the British from exploring the construct of caste further. The census officers reported the difficulties in collection of this data, as the "ignorant" respondents would often reply their caste as their occupation, or their sub-group, or their clan or some other social identity marker (Risley, 1901). Nonetheless, each of the decadal census operations (from 1872 to 1931) had extensive collections of data on caste. Here's how the classification of castes evolved in British censuses until 1901:

1. The first All-India census of 1871-72 suffered from many inconsistencies with the classification of caste. For example, while the provincial census officer F.S. Growse in North Western Province attempted to classify them occupationally into *Brahmins*, *Rajputs*, *Baniyas* (trading caste) and "Other Castes" (and pointed at the incompatibility of Varna system), in provinces of Bengal, Central Province, and Madras regions, they were classified into 13, 11 and 17 occupational groups (Samarendra 2011).

2. Given these inconsistencies in the 1871-72 census, the 1881 census sought consistency and required castes to be classified into an even broader classification of *Brahmins*, *Rajputs*, and Other Hindus.

3. The 1891 census sought a greater degree of granularity, and the chief census commissioner, J A Baines, sought to classify castes into 49 occupational groups like "Military", "Priests", "Barbers", "Potters" and "Miscellaneous Vagrants" and 11 racial/ethnic groups like "Musalmans Foreign Races", "Eurasians" and "Africans" (Census, 1891, p. 188).

4.2.3 Risley's 1901 Census: A laboratory of identity

Until the 1891 census, the classification of caste was principally occupational. When the initial classification of castes into *Varnas* (in 1871-72 and 1881) and then into broader occupational groups (in 1891) failed, this paved the way for a racial and ethnographic approach in 1901.

[13]The one-half of *Varna* order that Growse believed still existed in India, were the *Brahmins* (priests) and the *Kshatryas* (the ruling class). But, it is worth pondering whether such an exclusive and regulated priestly and ruling class was unique only to India, or a common feature of most pre-modern societies.

The most extensive ethnographic work on caste was undertaken by Herbert Risley, who undertook detailed ethnographic and anthropometric studies of different caste groups in Bengal that resulted in his book *The Tribes and Castes of Bengal* published in 1891. He came up with a racial theory of caste system where he proposed that the upper castes belong to the Aryan races, while the lower castes were indigenous. He was appointed as the chief commissioner of the 1901 census, and hoping to come up with a grand positivist theory of caste, he was not only interested in enumerating and understanding different castes of India, but also in (uniquely[14]) ascribing social precedence to them. One of the motivations of Herbert Risley for ascribing social precedence to castes was his notorious and later discredited hypothesis:

> The social position of a caste varies inversely as its nasal index.

The census operations of 1901, headed by Mr. Herbert Risley, attempted to identify, as well as classify castes in a hierarchy on the basis of native "public opinion". But, the head of census operations in Bengal, E.A. Gait, wrote how the Hindus as a body in Bengal seemed indifferent to such social precedences. He wrote in his provincial report:

> The test laid down by the Census Commissioner for fixing the scale of social precedence is not the rank assigned by the pedantry of pandits, but "Hindu public opinion at the present day." It is very difficult to say precisely what constitutes Hindu public opinion. The Hindus as a body are strangely indifferent to the circumstances of castes that do not clash with their own. Those of good position know very well from whom they can take water and those whose touch defiles, but they neither know nor care much regarding their relative position. The lower castes are even more ignorant of the rank of the higher ones. Where the relative position of two castes is disputed, the persons interested invariably support the claims of their own community.

As the census administration was interested in classifying castes in hierarchies, caste-based associations sprang up especially after 1901. These associations were lobbying and political action groups, that attempted to negotiate their status in these tables and organize their communities. E.A. Gait, census officer of Bengal, writing about social precedence tables wrote:

> The discussion of the relative rank of the different castes aroused an extraordinary amount of ill-feeling and jealousy between some of the castes whose position is disputed, and in more than one instance the committees appointed to report on the subject professed their inability to come to a decision. In some cases, it is very difficult to arrive at any satisfactory conclusion on the evidence available. Moreover, where the relative rank of the castes concerned is very nearly equal, there is nothing to be gained by attempting to adjudicate between them, while to do so would undoubtedly cause much mortification to the community whose claims were over-ruled ; the decision would not be accepted as authoritative, and far from extinguishing, it would add fresh fuel to the fires of acrimonious argumentation.

[14] In censuses from 1911 castes were only enlisted alphabetically, making 1901 censuses in this sense unique.

While Bengal province developed these social precedence based tables, not all of the provinces of British India (including Princely states) adopted these social precedence tables, and the implementation varied considerably. We use this variation as a source of identification in our empirical model about which we discuss in the next section.

4.3 Empirical Strategy

Herbert Risley's 1901 social precedence table was heterogeneously applied across Indian provinces. In this section we shall 1) examine the ex-ante similarities between these heterogeneous regions, 2) discuss a potential instrument variable - distance from Bengal (Calcutta) - that predicts this heterogeneous application, and 3) examine the ex-post effects of this heterogeneous application on caste sabhas in subsequent decades in the different regions. We shall follow this discussion, with studying the differences across these regions in modern day India, on networking characteristics of households, and in the quality of public goods provision today.

Risley's classification was adopted only in some regions of India, where district committees were formed to deliberate a detailed ranking of different castes. In other provinces, either no ranking of castes happened, or else the provincial census officers provided a sketchy social precedence table, that involved little or no deliberation (in other words the ranking was largely internal and did not engage with the general public). So, there were two types of regions in the 1901 census:

1. **Risley's Treatment Group:** In the Census regions of Bengal, Northwestern Provinces and Oudh, Central Provinces, Berar, Madras, Gwalior, Rajputana Agency, Ajmer Merwara and Baroda, social precedence tables with a detailed ranking of castes were formed, with the help of district committees (Table F.1a). These regions are shown in Figure 4.4. These provinces will be called the treated regions.

2. **The Non-Treatment Group:** In other provinces of British India, such detailed caste classification based on district committees was not undertaken. In the Census regions of Assam, Central India Agency, Bombay, Hyderabad, Kashmir, and Cochin a social precedence table was developed, but they were formed without a detailed and explicit consultation with any district committees, or their classification was vague/not detailed (Table F.1b). In the Census regions of Mysore, Travancore, Punjab, and Coorg, no social precedence table was formed (Table F.1c).

4.3.1 Adoption of Risley's classification

What deterred some provinces from undertaking a classification based on social precedence? The historical provinces in the treated region were not different from the other historical provinces on a range of observed characteristics like educational attainment, and agricultural labor share, the data for which was collected from the national and provincial census of 1901. Table 4.1a shows the historical provincial level statistics for treated and untreated regions, while

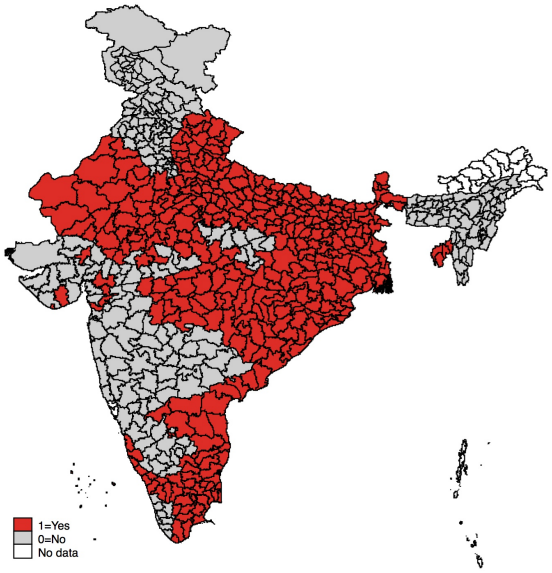

Figure 4.4: Regions where a district committee was formed to ascertain in detail social precedence of caste.

Table 4.1b shows characteristics of the historical districts lying in the treated and untreated regions. As observed, the two regions have no significant difference in observed characteristics that are predictive of backwardness.

Were there pre-colonial differences between the treated and untreated regions in the makeup of dominant castes? The Mughal administrative book *Ain-I-Akbari* written around 1590, contains information on *Qaum* (caste) of Zamindars (landlords) in different subprovinces that were under Akbar's direct or indirect rule. A dummy variable was created to label if a historical subprovince had any "lower" caste (as ascribed by the British in 1901) as a Zamindari *Qaum*. Only the historical location of the capitals of these subprovinces was known (as located by Schwartzberg Atlas, p. 147) and the boundaries were unknown. Hence these historical provinces were grouped in the treated or untreated regions depending on which region their capital belonged in. These capitals were only located in the 13 Northern British provinces of India. Table 4.1c shows the differences between treated and untreated regions, and there existed no statistically significant difference between the two regions.

4.3.2 Distance from Bengal (Calcutta)

What exogenous factors motivated the formation of district committees in 1901? Bengal's caste classification scheme was so detailed because Herbert Risley spent several years studying the Bengali society while working on his book *Castes and Tribes of Bengal, 1891*, before becoming the 1901 census director. So, census collectors in Bengal had a more detailed framework to base their system of social precedence on, in comparison to regions such as Punjab or Mysore.

It is noteworthy that provinces that were close to Bengal like Madras and Northwestern Provinces and Oudh adopted the social precedence tables similar to Risley's. In contrast, provinces like Punjab, Mysore, Bombay, and Travancore - that did not adopt such a caste precedence table - were distant from the Bengal province. One of the reasons why adjacent regions could adopt Risley's classification could be the shared names of castes due to linguistic simi-

Table 4.1a: Summary statistic at province level with differences between treated (which established district committees to rank caste) and untreated regions. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	(1)	(2)	(3)	(4)	(5)	(6)	Treated	Untreated	
	Summary Statistic at Provincial Level								
VARIABLES	mean	median	sd	min	max	N	mean	mean	Diff
Literacy Male 1901 (%)	9.81	8.5	5.45	3.8	22.4	19	8.98	10.55	-1.57 (2.48)
English Male Education 1901 (%)	0.68	0.63	0.43	0.09	1.65	19	0.63	0.72	-0.09 (0.2)
Agricultural Labor share 1901 (%)	62.13	58.6	11.75	46.1	84.2	19	62.59	61.71	0.88 (5.42)
Industrial labor share 1901 (%)	16.24	16.2	5.7	7.8	32.4	19	15.44	16.96	-1.52 (2.55)
Indirect British Rule (Dummy)	0.47	0	0.51	0	1	19	0.33	0.6	-0.267 (0.23)
Density 1901: per sq mile	215.8	165.5	151.9	36	596	18	233.1	201.9	31.23 (71.76)

Table 4.1b: Summary statistic at district level with differences between treated (which established district committees to rank caste) and untreated regions. Robust standard errors clustered at 19 (18 for Industrial labor share) historical province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	(1)	(2)	(3)	(4)	(5)	(6)	Treated	Untreated	
	Summary Statistic at Historical District Level								
VARIABLES	mean	median	sd	min	max	N	mean	mean	Diff
Literacy Male 1901 (%)	8.19	6.6	5.41	1.2	36	275	7.97	8.57	-0.6 (1.98)
English Male Education 1901 (%)	0.61	0.3	1.43	0	14.1	274	0.6	0.61	-0.01 (0.28)
Agricultural Labor share 1901 (%)	65.2	66.85	14.72	3.2	95.6	266	66.59	62.58	4.01 (4.94)
Industrial labor share 1901 (%)	14.98	14.1	6.99	0.2	43.4	251	14.8	15.26	-0.45 (2)

Table 4.1c: Dummy variable for whether a Mughal subprovince during Akbar's reign had zamindars (landlords) that were not considered high caste by the British. If two or more Mughal subprovinces had capital enclosed in the same modern district, they is considered as a single composite subprovince (source Ain-I-Akbari, early 16th c., Schwartzberg Atlas of South Asia). See F.2 for details. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	(1)	(2)	(3)	(4)	(5)	(6)	Treated	Untreated	
	Summary Statistic at Mughal District Level								
VARIABLES	mean	median	sd	min	max	N	mean	mean	Diff
"Low" Caste Mughal Zamindars	0.67	1	0.47	0	1	73	0.66	0.71	-0.05 (0.130)

Table 4.2: Summary statistic of land distance from Kolkata and the city at which the province's census was printed (assumed to be the administrative center) with differences between treated (which established district committees to rank caste) and untreated regions. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	(1)	(2)	(3)	(4)	(5)	(6)	Treated	Untreated	
	Summary Statistic at Provincial Level								
VARIABLES	mean	median	sd	min	max	N	mean	mean	Diff
Land transport dist from Calcutta (in km)	1741	1656	644	0	2453	19	1106	1742	-636** (263)

larities between adjacent regions. For example, the sub-province of Bihar in Bengal shared the Hindi language with the other adjacent Hindi speaking regions, and the sub-province of Orissa in Bengal shared the Oriya language with a handful of districts of the adjacent state of Madras. In the Northern British province of Punjab, census officer H.A. Rose explained his decision not to have a social precedence table by writing:

> We know far too little of the complex organization of the Punjab castes to be able, at present to classify them in any systematic or instructive way. The complications within the castes have their natural counterpart in the chaotic and uncertain relations between the different castes.

To incorporate the spatial closeness of these provinces to Bengal, we measure the land transport distance around 1901 (using highway maps around that period) between the Kolkata and the city where the census report of the respective city was printed, which was in most cases also the administrative center of the linguistic region. Table 4.2 shows that treated region was closer on an average by more than 600 km when compared to the untreated regions. To use this variable as an instrument variable, we normalize this variable by dividing it by the largest distance (between Kolkata and Trivandrum in Travancore).

4.4 Risley's classification and salience of caste identity

4.4.1 Salience of the word "caste"

If the 1901 caste censuses were indeed a historical shock on the salience of caste in India, we should observe that before the 1901 census the number of mentions of the word "caste" was not significantly different in the census documents of the treated and untreated regions. Additionally, for this measure to be relevant, we should observe that the mention of the word "caste" in the 1901 census was higher in the treated regions. Also, we should observe that the mentions of other identity markers like "religion" and "tribe", and other general/related words like "census", "India" and "population" were similar across regions.

The word count analysis in Table 4.3 shows that in 1891, mention of identity-related words (including "caste") and other general related words was similar in the treated and untreated regions. Across time, the pattern for the mention of words for "religion", "tribe" and common words like "census", "population", "India" and " the " did not vary across the treated and untreated regions, in the years 1891, 1901 and 1911.

While the mention of "caste" (see Figure 4.5) per page was similar in the two regions in 1891, but it markedly increased in the treated region in 1901 (about twice) and in 1911 (about

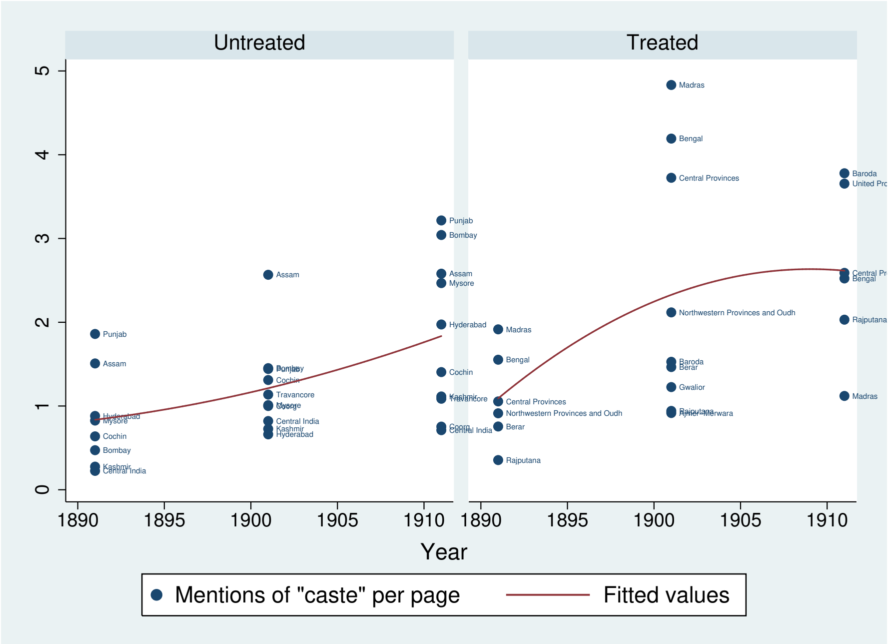

Figure 4.5: Total mentions of word "caste" per page in 1891, 1901 and 1911 provincial censuses across treated and untreated regions.

1.5 times but not statistically significant). The mention of the word "language" per page also increased in in 1901 (about 1.5 times) in the treated region, although it was similar across regions in 1911.

4.4.2 Immediate consequences: Census petitioning

What were the consequences of Risley's caste classifications? In the section 4.1 we emphasized how the state - by making the ideal prescriptions associated with a social category (like castes) common knowledge - can make social adherence to ideal prescriptions more common (see Figure 4.3). Risley's district committees were ranking castes on social precedence in 1901 often in accordance with the prescriptions of Sanskritic ideals. Such Sanskritic prescriptions gave rise to several caste groups in British provinces of Bengal, Berar, Madras and the United Provinces (Lee 2015) that petitioned to level up in the precedence table and especially desired to be recognized as *Brahmins* or *Kshyatriyas*, often touting their Sanskritic credentials.

The provincial census officers provided rich details in their reports describing the disputes and their resolution. What did some of these disputes look like? The 1901 Bengal census report enlists some of the grievances put forth by different caste groups. *Babhans* claimed in Bihar that they were *Brahmins* who were landholders. *Baidyas* claimed in Bengal to be placed next to *Brahmans* above *Rajputs* of Bihar and above *Kayasthas*. *Kayasthas* in Bengal claimed to be *Kshtriyas* above *Vaidyas*. *Chasi Kaibarttas* (class 4 of Bengal) claimed to be different from *Jaliya Kaibarttas* (class 6 of Bengal). They claimed themselves as *Mahisya* which was respectable as *Kshatriyas* and *Vaisyas*. *Jugis* (class 5 of Bengal) claimed they were *Brahmin* ascetics. *Awadhiya Kurmis* (Class 3 Bihar) claimed to be *Kshatriyas* and a separate caste from other *Kurmis*. Some *Pods* (Class 6 of Bengal) claimed to be Kshatriyas and different from other fishing pods. *Rajbansis* claimed to be *Kshatriyas* but they were clubbed as Class 6 of Bengal.

Table 4.3: The tables report the number of times a word is mentioned per page in the OCR scanned files obtained via the Hathi Trust, and the University of Chicago, of provincial census documents of 1891, 1901 and 1911. The word per page is measure is the ratio between the total number of times a word gets mentioned in the document and the total number of pages in the scanned document. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	(1)	(2)	(3)	(4)	(5)	(6)	Treated	Untreated	
	Summary Statistic at 1891 provincial census document level								
WORD	mean	median	sd	min	max	N	mean	mean	Diff
# "caste" per page	0.95	0.85	0.57	0.23	1.92	14	1.09	0.84	0.25
									(0.31)
# "religion" per page	0.36	0.36	0.14	0.17	0.62	14	0.36	0.36	0.01
									(0.07)
# "tribe" per page	0.31	0.20	0.27	0.07	1.08	14	0.31	0.31	0
									(0.14)
# "language" per page	0.20	0.13	0.20	0.02	0.73	14	0.17	0.22	-0.05
									(0.10)
# "census" per page	1.13	0.95	0.76	0.27	3.06	14	1.20	1.08	0.12
									(0.40)
# "India" per page	0.37	0.34	0.16	0.14	0.67	14	0.37	0.37	0
									(0.08)
# "population" per page	0.96	0.83	0.52	0.33	2.39	14	1.21	0.77	0.44
									(0.29)
# " the " per page	17.61	14.83	12.58	4.69	40.72	14	17.37	17.79	-0.42
									(6.59)
# of pages	630.29	589	227.32	345	1111	14	608.33	646.75	-38.42
									(116.13)
	Summary Statistic at 1901 provincial census document level								
WORD	mean	median	sd	min	max	N	mean	mean	Diff
# "caste" per page	1.74	1.31	1.22	0.66	4.83	19	2.33	1.21	1.11*
									(0.53)
# "religion" per page	0.49	0.45	0.24	0.14	1.30	19	0.49	0.50	-0.01
									(0.11)
# "tribe" per page	0.43	0.38	0.23	0.03	0.93	19	0.47	0.39	0.08
									(0.11)
# "language" per page	0.41	0.34	0.18	0.21	0.84	19	0.48	0.34	0.15*
									(0.08)
# "census" per page	0.79	0.68	0.45	0.27	2.27	19	0.70	0.87	-0.18
									(0.2)
# "India" per page	0.44	0.37	0.26	0.18	1.33	19	0.42	0.45	-0.03
									(0.12)
# "population" per page	1.35	1.36	0.52	0.73	2.99	19	1.34	1.36	-0.02
									(0.24)
# " the " per page	22.33	23.75	8.39	7.81	41.71	19	24.61	20.27	4.34
									(3.81)
# of pages	405.42	329	247.20	77	1119	19	381.67	426.80	-45.13
									(114.03)
	Summary Statistic at 1911 provincial census document level								
WORD	mean	median	sd	min	max	N	mean	mean	Diff
# "caste" per page	2.13	2.25	1.01	0.71	3.78	16	2.62	1.83	0.78
									(0.50)
# "religion" per page	0.64	0.65	0.16	0.38	0.94	16	0.65	0.63	0.02
									(0.09)
# "tribe" per page	0.39	0.34	0.21	0.14	0.80	16	0.38	0.40	-0.01
									(0.12)
# "language" per page	0.38	0.37	0.15	0.15	0.74	16	0.40	0.37	0.02
									(0.07)
# "census" per page	1.04	1.06	0.42	0.49	1.73	16	0.95	1.09	-0.14
									(0.19)
# "India" per page	0.81	0.82	0.26	0.31	1.24	16	0.88	0.76	0.12
									(0.14)
# "population" per page	1.77	1.69	0.75	0.82	3.48	16	1.83	1.74	0.09
									(0.38)
# " the " per page	31.05	35.25	11.95	8.28	47.64	16	35.08	28.62	6.45
									(5.12)
# of pages	331.4	301	147.59	65	629	16	393.17	295.20	97.97
									(72.33)

Sadgops claimed to be *Vaishyas* but were clubbed with *Goalas* (Class 4 of Bengal). *Shahas* (Class 5 of Bengal) claimed to be *Kshtriyas* and *Vishya* by profession who got degraded due to drinking. Census officer classed them a rich and progressive caste but not elevated in "Hindu public opinion".

In contrast, census areas that did not have a social precedence table did not mention such caste petitioning. Consequently, caste associations "sprung up to contest their assigned position in the official hierarchy, holding meetings, writing petitions" to contest their assigned positions in the census (Dirks 2001). Dirks (2001) points at the archival records from Madras presidency, where the conflict between "upper and lower castes... rose dramatically in the early decades of the twentieth century... in wake of the politicisation of caste around the census".

Due to the acrimonious nature of the 1901 census's social precedence table, from 1911 the ranking of castes was abandoned. The census officer of Bengal's 1911 census, L.S.S. O'Malley wrote:

> No part of the census aroused so much excitement as the return of the castes... The feeling on the subject [object of the census was to "fix the relative status of different castes"] was very largely the result of castes having been classified in the last census report in order of social precedence. This "warrant of precedence" gave rise to considerable agitation at the time and proved a legacy of trouble. The agitation was renewed when the census operations of 1911 were instituted. Hundreds of petitions were received from different castes- their weight alone amounts to 1.5 mounds- requesting that they might be known by new names, by placed higher in the order of precedence, e recognized as Kshatriyas, Vaisyas, etc. Many castes were aggrieved at the position assigned them and complained that it lowered them in public estimation. The Subarnabaniks, in particular, were offended at being placed low down in the list, whereas in 1891 they were grouped with other Bania castes among Vaisyas. Others thought it a suitable opportunity to advance new claims. It was impossible to comply with these requests, as it was decided from the outset that there should be no classification of castes by status.

Mentioning the politicised nature of caste census since 1901 (even without the social precedence tables), 1931 census commissioner and anthropologist, J.H. Hutton, in his census report, lamented:

> All subsequent census officers in India must have cursed the day when it occurred to Sir Herbert Risley, no doubt in order to test his admirable theory of the relative nasal index, to attempt to draw up a list of castes according to their rank in society. He failed, but the results of his attempt are almost as troublesome as if he had succeeded, for every census gives rise to a pestiferous deluge of representations, accompanied by highly problematical histories, asking for recognition of some alleged fact or hypothesis of which the census as a department is not legally competent to judge and of which its recognition, is accorded, would be socially valueless.

He also pointed that:

> As on the occasion of each successive census since 1901, a certain amount of criticism has been directed at the census for taking any note at all of the fact of caste. It has been alleged that the mere act of labelling persons as belonging to a caste tends to perpetuate the system...

J.H. Hutton's comments from 1931 are backed by petitioning data collected by Lee (2015) see Table 4.4. The level of petitioning in treated provinces was consistently and considerably higher than petitioning in untreated provinces until 1931 (the last British census), and no petitions were made in untreated provinces in 1901. Figure 4.6 highlights this difference.

It is important to note that given the politicization of caste censuses in colonial India, India after independence abandoned conducting caste censuses altogether and hence the 1901 caste census of India was a unique exercise that elaborately placed different castes on social precedence tables in some parts of India.

4.4.3 Modern consequences: Networking patterns

Could there be long-lasting effects of the unique caste classification of 1901? As the caste ranking exercise of 1901 was unique and affected how local populations perceived the census, do people in the treated regions give a greater salience to caste today? To test this, we compare the caste networking characteristics of modern households that belong in the treated region to those that do not.

Out of Caste networks

If caste is more salient in the treated region, we should expect people's social networks to be more inward and caste-based, and to be less cosmopolitan i.e. have fewer out of caste acquaintances (Burt 2001). To gauge local networking characteristics, we look at the Indian Human Development Survey (2011-12) (Desai & Vanneman 2015) of 42,152 households in 372 districts (out of 641) of India. The survey asks households a variety of questions about their socioeconomic status, their caste and - of interest to us - their acquaintances within and outside their caste/community. The survey asks if a household has acquaintances outside of community/caste with 11 different professions (which are elected officers, local politicians, government employees, other officers, inspectors, other police, military personnel, doctors, health workers, teachers and school workers). If a household says yes to x out of the 11 professions, then the average number of noncaste networks it has is x. Table 4.5a shows the summary statistic and differences between the number of out of caste networks.

Table 4.5a shows the significant differences across households in treated and untreated regions, when it comes to having acquaintance with professionals outside of their caste. On an average, a household has acquaintance with approximately 2 out of 11 out-of-caste professionals in the treated region, while this number is approximately 3 out of 11 in the untreated region.

Table 4.5b reports the difference between households in treated and non treated regions on acquaintance with professionals outside of community/caste. As fewer professional acquaintances, in general, may drive acquaintance with professionals outside of community/caste, we

Table 4.4: Number of caste petitions in each of the British provinces as compiled by Lee (2015). Bengal province was split into Bengal, and Bihar and Orissa in 1905. Central Provinces was merged with Berar in 1903.

	1901	1911	1921	1931	Total
Treated	**3**	**4.67**	**5.44**	**8.56**	**21.67**
Baroda	0	3	3	9	15
Bengal	10	15	16	15	56
Bihar and Orissa	3	5	8	8	24
Central Provinces	2	7	0	8	16
Madras	5	3	7	11	26
Rajputana Agency	0	2	4	4	10
United Provinces	7	7	11	22	47
Ajmer Merwara	0	0	0	0	0
Gwalior	0	0	0	0	0
Untreated	**0**	**1.4**	**1.9**	**4**	**7.3**
Bombay	0	3	4	3	10
Central India Agency	0	1	2	9	12
Hyderabad	0	0	0	11	11
Mysore	0	0	5	8	13
Punjab	0	10	6	4	20
Travancore	0	0	2	5	7
Assam	0	0	0	0	0
Coorg	0	0	0	0	0
Kashmir	0	0	0	0	0
Cochin	0	0	0	0	0

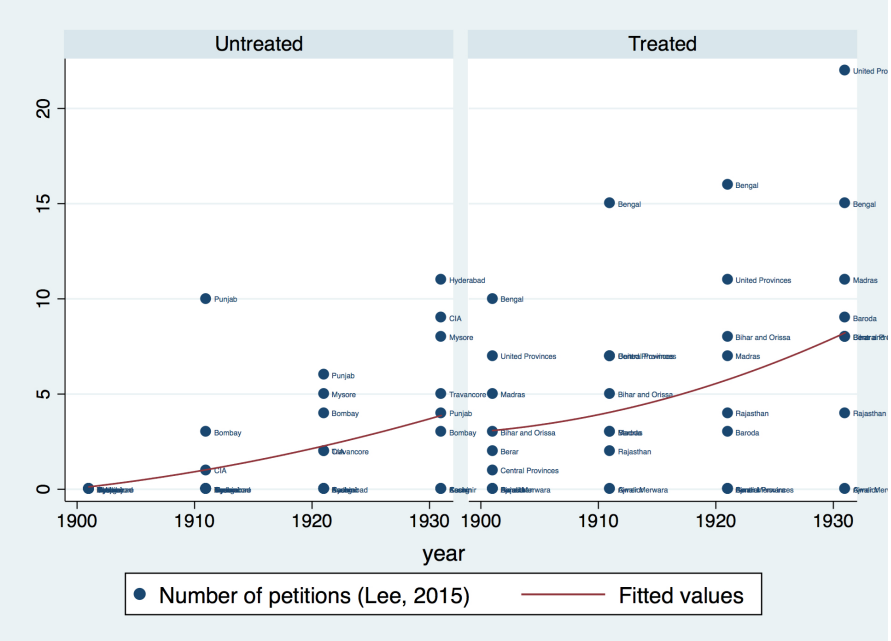

Figure 4.6: Average number of petitions in a treated and untreated province from 1901 to 1931. Source: Lee (2015)

Table 4.5a: Summary statistic at household level with differences between treated (which established district committees to rank caste) and untreated regions. The IHDS survey asks if a household has acquaintances outside of community/caste with 11 different professions. If a household says yes to x out of the 11 professions, then the average size of non caste networks it has is x. Robust standard errors clustered at 19 historical province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	(1)	(2)	(3)	(4)	(5)	(6)	Treated	Untreated	Diff
VARIABLES	mean	median	sd	min	max	N	mean	mean	
# Non Caste Networks	2.47	2	2.68	0	11	41112	2.04	2.97	-0.93***
									(0.26)

control for the number of acquaintances the household has with professionals within family/-caste/community in Columns 2 to 6.

When controlling for within caste household networks, households in districts with 1901 caste committees have 0.68 fewer acquaintances with professionals outside of caste/community. This is a sizable reduction, as the mean acquaintance levels with professionals outside of relatives/caste/community is 2.47.

When distance from Kolkata is used as an instrumental variable, there is a close 1 to 1 correspondence between the distance from Kolkata instrument variable and adoption of 1901 Caste Committees and the estimate in the 2SLS column 3 in Table 4.5b does not change significantly (0.57 percentage point) when compared to the OLS regression in Column 2 (0.68 percentage point). When considering differences within states (Column 4 and 6) and within the same caste (Column 5 and 6), similar differences persist between the treated and untreated regions.

Passing as Other Backward Caste

Table 4.6 describes the demographic characteristics of the IHDS survey households. In modern India castes have been classified into Upper Castes and Scheduled Castes and Tribes. Within the upper castes, Brahmins make a clearly identifiable group, while after the 1979-1983 Mandal Commission, Upper Castes have been further classified into Forward Castes and Other Backward Castes (OBC). While Brahmins (due to their historical ritual dominance), and Scheduled Castes and Tribes (due to their historical backwardness) are clearly identifiable, there exists some ambiguity and dispute among Upper Castes on which castes should get identified as Forward Castes and OBCs.

Table 4.6 shows that the clearly identifiable caste classifications like Brahmins, Scheduled Castes, Scheduled Tribes, Muslims and Hindus (overall) have a similar share in the treated and untreated regions. Also, the combined population of the more ambiguous Forward caste and OBCs classification make half of the sample in both the regions.

The classification of OBCs was conducted to provide targeted benefits to these castes in government jobs and college admissions. So, it creates incentives for the castes that belong to the ambiguous Upper Caste group to pass as OBC (Caselli & Coleman 2013). If caste is more salient in our treated region, we should find more households of a given caste passing as OBCs in the treated region, as they are more likely to agitate for such OBC benefits.

Table 4.6 shows that households in the treated region are less likely to be notified as Forward caste, and more likely to be identified as OBC. When considering a sample of castes that are

Table 4.5b: Networking differences between households in treated and non-treated regions, based on IHDS household survey. The IHDS survey asks if a household has acquaintances outside of community/caste with 11 different professions. If a household says yes to x out of the 11 professions, then the average size of non caste networks it has is x. A corresponding questions asks if the household has acquaintances within family/community/caste in the same 11 professions. The 2SLS regression uses as IV the land transport distance from Calcutta of the city where the provincial census was printed; measured by looking at the traditional train route for these cities to Kolkata around 1901 (using Schwartzberg Atlas of South Asia). The F test value is 18.96. Within caste regression considers castes that have more than or equal to 10 unique observations. Robust standard errors clustered at 19 historical province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	2SLS	OLS	OLS	OLS
VARIABLES	# Non Caste Networks					
Treated region	-0.93***	-0.68***	-0.57**	-0.64***	-0.60***	-0.72***
	(0.26)	(0.23)	(0.23)	(0.21)	(0.19)	(0.2)
# Caste Networks		0.68***	0.68***	0.67***	0.66***	0.65***
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
State FE				Yes		Yes
Caste FE					Yes	Yes
Constant	2.97***	1.74***	1.68***	1.91***	1.73***	2.50***
	(0.24)	(0.21)	(0.21)	(0.07)	(0.19)	(0.17)
Observations	41,112	41,092	41,092	41,092	28,025	28,025
R-squared	0.03	0.33	0.33	0.36	0.4	0.42

Table 4.5c: Networking differences between households in treated and non-treated regions, based on IHDS household survey. The IHDS survey asks if a household has acquaintances outside of community/caste with 11 different professions. A corresponding questions asks if the household has acquaintances within family/community/caste in the same 11 professions. This table reports differences in networking with political officials. Robust standard errors clustered at 19 historical province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	mean	Treated	Untreated	(1)	(2)	(3)
Acq. with out of community/caste X					Difference	
X =						
Elected official	0.13	0.09	0.18	-0.09**	-0.08**	-0.09**
				(0.03)	(0.03)	(0.04)
Local Politicians	0.12	0.10	0.16	-0.06*	-0.06**	-0.07**
				(0.03)	(0.02)	(0.03)
Control: Acq. within community/caste X				No	Yes	Yes
IV				No	No	Yes

Table 4.6: Summary statistic at household level with differences between treated (which established district committees to rank caste) and untreated regions. Forward and Other Backward Caste (OBC) are intermediary castes below Brahmins, and above Schedule castes and tribes, where OBCs receive affirmative action. Hybrid caste refers to castes wfahere > 10 percent of respondents were OBCs and Forward castes i.e. their status varied by region.Robust standard errors clustered at 19 historical province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

	(1)	(2)	(3)	Treated	Untreated	Diff	Within Hybrid
VARIABLES	mean	sd	Obs	mean	mean		caste
Forward Caste	0.17	0.37	6946	0.12	0.22	-0.09**	-0.10*
						(0.04)	(0.05)
Other Backward Class	0.34	0.47	14089	0.38	0.29	0.1	0.22**
						(0.06)	(0.08)
Brahman	.052	0.22	2158	0.05	.05	0.00	
						(0.02)	
Scheduled Caste	0.21	0.41	8921	0.21	0.22	0.00	
						(0.04)	
Scheduled Tribes	0.08	0.28	3445	0.08	0.09	0.01	
						(0.04)	
Hindus	0.86	0.35	35559	0.85	0.86	0.01	
						(0.03)	
Muslims	0.12	0.32	4836	0.10	0.13	0.03	
						(0.04)	

"Hybrid" (i.e. which are classified both as Forward and OBC castes in different regions), and comparing households of the same caste in the treated and untreated region, we find that households (within the same Hybrid caste) are more likely to be classified as OBCs in the treated region, hinting towards a greater instance of passing.

4.5 1901 and today's Quality of Government

In the previous section we found that the 1901 census was associated with an increase in the salience of caste in India, and led to 1) increased the number of petitions between 1901 and 1931, 2) fewer out of caste acquaintances today and 3) more households passing as castes that benefit from affirmative action. These trends highlight an increasing role of caste identity in the government and political sphere. For example, decomposition of acquaintances by different professions shows that the treated regions have significantly lower out of caste acquaintances with elected officials and local politicians (see Table 4.5c).

Rothstein (2011b) developed a theory that argued that more impartial states have a better quality of government. As the caste censuses of 1901 increased the salience of caste in politics, do these treated regions also have a poorer quality of government, as Rothstein (2011b) proposed?

The quality of government literature has attempted to measure the efficiency of governments and their outcomes using a range of proxy variables (La Porta, Lopez-de Silanes, Shleifer & Vishny 1999, Rothstein 2011b). Prior literature such as Alesina, Baqir & Easterly (1999) studying ethnic fractionalization have especially emphasized how fractionalization leads to lower production of public goods. If caste is more salient in the treated region of India, does it lead to more influence of caste-based politics? It is worth noting that explicitly lower caste-based parties (Varshney 2000) of India[15] are predominantly active in the treated region with Uttar Pradesh

[15]Varshney (2000) lists the following parties as explicitly lower caste-based parties: Janata Dal (various versions in *Bihar* and Karnataka), Rashtriya Janata Dal, RJD (*Bihar*), Samajwadi Party, SP (*Uttar Pradesh*), Bahujan Sama-

Table 4.7a: Quality of government variables used by La Porta, Lopez-de Silanes, Shleifer & Vishny (1999) relating to efficiency and output of public goods.

Variable: Output of Public Goods	Related IHDS 2011 variable for India
Log of Infant mortality ($<$ 1 year)	Birth History survey
Log of School Attainment (avg. years of schooling)	Individual school attainment
Illiteracy Rate (15 years or above)	Individual literacy
Infrastructure Quality (Communication and Transport)	Household and village access to Electricity, Water, Sanitation, Road & Phone

(erstwhile British province of United Provinces), Bihar and Orissa (previously part of the British province of Bengal) and Tamil Nadu (previously part of the British province of Madras) being the prominent regions with influential explicitly lower caste-based parties.

Does such caste-based nature of politics in the treated region lead to a lower production of public goods today due to a greater salience of caste? The regional (state and local) governments play a large role in determining the quality of public goods such as electricity, healthcare, and education in India. Table 4.7a shows the different variables used by La Porta, Lopez-de Silanes, Shleifer & Vishny (1999) to measure the quality of government, and the related data on public goods and efficiency available at the sub-national level on India, that can be used to measure within India differences in the quality of government.

Table 4.7b reports the summary statistics on infrastructure quality and the differences between treated and untreated regions. The differences between the regions on water, electricity, sanitation and communication infrastructure appear to be large, and the treated regions have significantly worse infrastructure quality on these parameters. Both regions seem to have similar or statistically insignificant differences in access to pakka (proper) roads and bus facilities.

Table 4.7c reports the summary statistics on educational outcomes and the differences between treated and untreated regions. The differences between the regions on school attainment and literacy rate are large and significant. The same is true for health outcomes where infant mortality rate for children is significantly higher in the treated region. The treated region has significantly worse health and educational outcomes, based on these parameters.

So, the overall quality of public good provision in the treated region is worse.

4.6 Conclusion

Caste has been a persistent feature of Indian society that has evolved over time and is associated with multiple related concepts such as *Quam*, *Jati*, and *Varna*. The colonial rule was an important period in giving rise to the caste system as we know today (Dirks 2001), and one of the ways through which the colonial rule altered the meaning and salience of caste was through the caste-based colonial censuses - especially the census of 1901.

jwadi Party, BSP (*Uttar Pradesh*), Janata Party (disbanded), All India Anna Dravida Munnetra Kazhagam, AIADMK (*Tamil Nadu*), Dravida Munnetra Kazhagam, DMK (*Tamil Nadu*), Marumalarchi Dravida Munnetra Kazhagam, MDMK (*Tamil Nadu*), Paattali Makkall Katchi, PMK (*Tamil Nadu*), Biju Janata Dal, BJD (*Orissa*), and Republican Party of India, RPI (Maharashtra and *Uttar Pradesh*). Treated regions are mentioned in italics.

Table 4.7b: Summary statistic on infrastructure quality at household or village level with differences between treated and untreated regions. Robust standard errors clustered at 19 historical province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)	(4)	(5)	(6)	Treated	Untreated	Diff
					Infrastructure Quality				
VARIABLES	mean	median	sd	min	max	N	mean	mean	
Access to indoor piped drinking water	0.31	0	0.46	0	1	41427	0.22	0.42	-0.21*** (0.07)
Average hours of piped water	1.88	0	4.31	0	24	41427	1.31	2.54	-1.24** (0.53)
Access to electricity	0.87	1	0.33	0	1	41420	0.81	0.94	-0.13** (0.05)
Avg. hours of electricity	13.24	14	8.18	0	24	41422	11.71	15.02	-3.31** (1.56)
Household toilet	0.56	1	0.5	0	1	41830	0.48	0.65	-0.17** (0.07)
Drainage	0.71	1	0.45	0	1	1382 village	0.63	0.81	-0.18** (0.08)
Telephone	0.08	0	0.27	0	1	41555	0.06	0.11	-0.05** (0.03)
Cell Phone	0.80	1	0.43	0	1	41559	0.76	0.86	-0.10** (0.04)
Pakka Road	0.87	1	0.34	0	1	1382 villages	0.87	0.87	-0.00 (0.05)
Bus facility	0.62	1	0.49	0	1	1383 villages	0.55	0.70	-0.15 (0.09)

Table 4.7c: Summary statistic on educational attainment at individual level (for those aged 15 or above) and infant mortality at birth history level, with differences between treated and untreated regions. Robust standard errors clustered at 19 historical province level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)	(4)	(5)	(6)	Treated	Untreated	Diff
					Education Outcomes				
VARIABLES	mean	median	sd	min	max	N	mean	mean	
Avg. years of education	6.40	7	5.07	0	16	145160	5.94	6.90	-0.89*** (0.31)
Literate	0.70	1	0.46	0	1	145197	0.67	0.74	-0.06** (0.28)
					Health Outcomes				
VARIABLES	mean	median	sd	min	max	N	mean	mean	
Infant mortality (<1 year)	0.06	0	0.24	0	1	109964	0.08	0.04	0.03** (0.01)
Infant mortality (<1 month)	0.05	0	0.21	0	1	109964	0.06	0.03	0.03** (0.01)

We developed a model for intergenerational transmission of social identity, where a state's popularization of ideal prescriptions for social categories through enumerating exercises such as the census (also see Lieberman & Singh (2017)), increases conformity to such ideal prescriptions, leading to myriad long-term consequences including a decline of cosmopolitanism, and the rise of caste-based organization of society. The 1901 census classification of India was an exercise that ranked castes on the basis of prescribed Sanskritic ideals of twice-born castes. Our empirical analysis showed that the regions that underwent this unique caste ranking and classification in 1901, were ex-ante similar on observable characteristics such as education, labor share and pre-colonial caste elites (Tables 4.1a, 4.1b and 4.1c), but were geographically closer to Bengal (Table 4.2) - the hub for the 1901 caste-based censuses. These regions immediately saw a rise in caste-based petitioning and agitation in the 1901 census and the subsequent censuses (Figure 4.6).

Today the regions that had such caste classification have a significantly lower level of out of caste acquaintances with professionals (Table 4.5a), especially politicians and elected officials (Table 4.5b), and have similar demographic characteristics, except for the intermediary castes, which today are more likely to pass as Other Backward Castes (Table 4.6), a caste classification which helps castes receive affirmative action. Moreover, today these regions have a greater presence of caste-based parties and also have a poorer infrastructure quality (Table 4.7b) and educational and health outcomes (Table 4.7c), each of which is an important component in measuring the quality of the government.

The evidence presented in the chapter is suggestive of the role caste censuses played in making the social construct of caste salient in modern India. By studying the unique setting of the 1901 caste census, we were able to identify a scenario where such social construction of identity could be causally tested. As caste related data only became common around the 1901 census, it is challenging to develop direct measures of salience of caste in the pre-1901 era. An absence of caste petitioning before 1901 suggests that the political salience of caste indeed was limited before 1901. Also, we collected data about the Mughal period and showed that the regions that were affected by the 1901 caste census were not different historically from other regions of India. In future, using techniques such as textual analysis, we would like to develop better measures of salience of caste identity before and after our treatment period. We would also like to include more variables that can measure the salience of caste identity in modern India.

Chapter 5

Conclusion: Persistence and Change

Research in the past decade has shown that history can cast a long shadow on a society (Michalopoulos & Papaioannou 2017). The positive experience of being a city republic (Guiso, Sapienza & Zingales 2016), or the negative experience of being exposed to slave trade (Nunn & Wantchekon 2011) can have long-term consequences on civic capital and trust. In light of such persistence, it becomes important to ask- when can societies break from the past, and what are the exact mechanisms that are resisting such a change.

In the second chapter, I theoretically scrutinized a classical question that relates to how societies organize: *How honesty and trust evolves among strangers?* I found that under a narrow but identifiable set of conditions trust and honesty between strangers can emerge. These narrow conditions include:

- a high benefit from engaging with strangers (corollary 1),
- a threshold adoption of sticky ($w < \hat{w}$) standardized business routines that reduced partiality in business practices (corollary 2),
- some degree of aversion to risk and making losses ($\lambda > 0$), and
- an absence of strong identities and prescriptions that promote in-group favoritism (corollary 3).

If the above conditions hold, and benefits that relationship-based institutions can provide are not too high (proposition 2), impersonal institutions can emerge. Cycles of trust, dishonesty, distrust, and honesty can regulate the reputation of strangers in the market, and sustain impersonal exchange.

In the third chapter, I focused on one of the most transformative periods of human history- the fifteenth and sixteenth-century Europe. During this period, the European economy underwent significant changes (Mokyr 2016, McCloskey 2016), and entered a path towards modern economic growth, driven by growing markets and cities, especially in Northwestern Europe. In the Northwestern European region, merchant guilds declined even though such guilds persisted in other parts of Europe until much later (eighteenth and even nineteenth century). Drawing from social network literature, I propose an explanation of a central puzzle in history: *How merchant guild networks declined in Northwestern Europe?*

I show that merchant guilds could persist and be dominant for so long because such relationship-based social systems were a good alternative in the absence of credible systems of impersonal exchange. But, when the incentives for entering in impersonal exchanges with unknown traders increased (due to rise of long-distance trade at the Atlantic coast), and there was more information and techniques available to new traders about how to conduct trade (due to the rise of the printing press), merchant guilds declined and impersonal markets emerged. I find that Northwestern European region had geographically favorable conditions, that enabled both - easy access to the Atlantic coast and higher levels of printing penetration due to the region's closeness to Mainz.

In the fourth chapter, I turned the focus on the modern Indian caste system. The caste system in India is not just a network of relationships, but also a source of social identity for people. With Paola Sapienza and Luigi Zingales, I studied: *How caste identity became salient in India and influenced networking and governance?* We study the impact of the 1901 caste censuses on the modern day patterns of networking and quality of government.

We first developed a model of intergenerational transmission of social identity of caste and cosmopolitanism, drawing from the influential research by Akerlof & Kranton (2000) and Bisin & Verdier (2001) on identity economics and cultural transmission. The model predicts that caste-based organization of society will increase, and cosmopolitanism will go down if the state popularized 'ideal" prescriptions of caste. We find that the popularization of "ideal" prescriptions of caste, through the caste censuses in 1901, gave rise to caste organizations that petitioned for a higher caste status in subsequent decades. The regions where these castes censuses occurred in 1901, today have a higher salience of caste identity. This salience of caste identity manifests in the form of lower levels of out-of-caste networks with professionals, and higher rates of passing into castes that benefit from affirmative action. These regions also have a greater presence of caste-based parties and lower levels of public goods generation. While the first two chapters focused on how traditional social systems decline, this chapter showed how social identities could persist through the reinvention of identity and the influence of institutions.

Overall, drawing from the history of the social organization in Europe and India, I make theoretical contributions to topics such as trust (chapter 2), social networks (chapter 3) and social identity (chapter 4). I show how traditional organization of society in the form of guilds and caste can persist (chapters 3 and 4), evolve (chapter 4) and decline (chapters 2 and 3), and how alternative impersonal systems of organizing society can emerge historically (chapters 2 and 3).

Bibliography

Acemoglu, D., Johnson, S. & Robinson, J. (2005), 'The rise of Europe: Atlantic trade, institutional change, and economic growth', The American Economic Review pp. 546–579.

Akerlof, G. A. & Kranton, R. E. (2000), 'Economics and identity', The Quarterly Journal of Economics **115**(3), 715–753.

Alesina, A., Baqir, R. & Easterly, W. (1999), 'Public goods and ethnic divisions', The Quarterly Journal of Economics **114**(4), 1243–1284.

Alesina, A. & Giuliano, P. (2015), 'Culture and institutions', Journal of Economic Literature **53**(4), 898–944.

Alesina, A. & La Ferrara, E. (2002), 'Who trusts others?', Journal of Public Economics **85**(2), 207–234.

Alesina, A., Michalopoulos, S. & Papaioannou, E. (2016), 'Ethnic inequality', Journal of Political Economy **124**(2), 428–488.

Andersen, S. E. (2011), The evolution of Nordic finance, Springer.

Anderson, B. (1991), Imagined communities: Reflections on the origin and spread of nationalism, Verso Books.

Angeles, L. (2017), 'The great divergence and the economics of printing', The Economic History Review **70**(1), 30–51.

Arrow, K. J. (1969), 'The organization of economic activity: Issues pertinent to the choice of market versus nonmarket allocation', The analysis and evaluation of public expenditure: The PPB system **1**, 59–73.

Axelrod, R. (1984), 'The evolution of cooperation: Revised edition'.

Bairoch, P., Batou, J. & Chevre, P. (1988), 'The population of European cities. Data bank and short summary of results: 800-1850'.

Baker, W. E. & Bulkley, N. (2014), 'Paying it forward vs. rewarding reputation: Mechanisms of generalized reciprocity', Organization Science **25**(5), 1493–1510.

Balliet, D., Wu, J. & De Dreu, C. K. (2014), 'Ingroup favoritism in cooperation: A meta-analysis.', Psychological Bulletin **140**(6), 1556.

Banerjee, A., Green, D. P., McManus, J. & Pande, R. (2014), 'Are poor voters indifferent to whether elected leaders are criminal or corrupt? a vignette experiment in rural India', Political Communication **31**(3), 391–407.

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

Banerjee, A. V. & Pande, R. (2007), 'Parochial politics: Ethnic preferences and politician corruption'.

Banfield, E. C. (1967), 'The moral basis of a backward society.'.

Bayly, S. (2001), Caste, society and politics in India from the eighteenth century to the modern age, Vol. 3, Cambridge University Press.

Bénabou, R. & Tirole, J. (2011), 'Identity, morals, and taboos: Beliefs as assets', The Quarterly Journal of Economics **126**(2), 805–855.

Benner, M. J. & Tushman, M. L. (2003), 'Exploitation, exploration, and process management: The productivity dilemma revisited', Academy of Management Review **28**(2), 238–256.

Berg, J., Dickhaut, J. & McCabe, K. (1995), 'Trust, reciprocity, and social history', Games and Economic Behavior **10**(1), 122–142.

Bes, L., Frankot, E. & Brand, H. (2007), Baltic Connections: Archival Guide to the Maritime Relations of the Countries Around the Baltic Sea (including the Netherlands) 1450-1800, Vol. 36, Brill.

Bianchi, F. & Squazzoni, F. (2015), 'Agent-based models in sociology', Wiley Interdisciplinary Reviews: Computational Statistics **7**(4), 284–306.

Bisin, A. & Verdier, T. (2001), 'The economics of cultural transmission and the dynamics of preferences', Journal of Economic theory **97**(2), 298–319.

Bisin, A. & Verdier, T. (2017), On the joint evolution of culture and institutions, Technical report, National Bureau of Economic Research.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D. & Roberts, J. (2013), 'Does management matter? evidence from india', The Quarterly Journal of Economics **1**, 51.

Blouin, A. & Mukand, S. W. (2018), 'Erasing ethnicity?', forthcoming Journal of Political Economy .

Bohnet, I. & Zeckhauser, R. (2004), 'Trust, risk and betrayal', Journal of Economic Behavior & Organization **55**(4), 467–484.

Bowles, S. (2016), The moral economy: Why good incentives are no substitute for good citizens, Yale University Press.

Boyd, R. & Richerson, P. J. (1988), Culture and the evolutionary process, University of Chicago Press.

Bravo, G., Squazzoni, F. & Boero, R. (2012), 'Trust and partner selection in social networks: An experimentally grounded model', Social Networks **34**(4), 481–492.

Buggle, J. C. & Durante, R. (2017), 'Climate risk, cooperation, and the co-evolution of culture and institutions'.

Burt, R. S. (1995), Structural Holes, Harvard University Press.

Burt, R. S. (2001), 'Closure as social capital', Social capital: Theory and research pp. 31–55.

Burt, R. S. & Luo, J.-D. (2017), Angry entrepreneurs: A note on networks prone to character assassination, in 'Academy of Management Proceedings', Vol. 2017, Academy of Management, p. 11201.

Buxbaum, D. C. (1968), Family law and customary law in Asia: A contemporary legal perspective, Martinus Nijhoff Publishers.

Caselli, F. & Coleman, W. J. (2013), 'On the theory of ethnic conflict', Journal of the European Economic Association **11**(suppl_1), 161–192.

Cassan, G. (2015), 'Identity-based policies and identity manipulation: Evidence from colonial Punjab', American Economic Journal: Economic Policy **7**(4), 103–31.

Chatfield, M. & Vangermeersch, R. (2014), The history of accounting (RLE accounting): An international encylopedia, Routledge.

Cohn, B. S. (1984), 'The census, social structure and objectification in south Asia in culture and history of India.', Folk **26**, 25–49.

Coleman, J. S. (1988), 'Social capital in the creation of human capital', American Journal of Sociology pp. S95–S120.

Cook, H. J. (2007), Matters of exchange: Commerce, medicine, and science in the Dutch golden age, Yale University Press.

CORE team, T. (2017), The Economy: Economics for a Changing World, Oxford University Press.

Cox, J. C. (2004), 'How to identify trust and reciprocity', Games and Economic Behavior **46**(2), 260–281.

Dasgupta, P. (2000), 'Trust as a commodity', Trust: Making and breaking cooperative relations **4**, 49–72.

Davis, R. (1973), The rise of the Atlantic economies, Cornell University Press.

Daybell, J. (2012), The material letter in early modern England: Manuscript letters and the culture and practices of letter-writing, 1512-1635, Springer.

De la Croix, D., Doepke, M. & Mokyr, J. (2017), 'Clans, guilds, and markets: Apprenticeship institutions and growth in the preindustrial economy', The Quarterly Journal of Economics **133**(1), 1–70.

Delhey, J., Newton, K. & Welzel, C. (2011), 'How general is trust in "most people"? solving the radius of trust problem', American Sociological Review **76**(5), 786–807.

Desai, S. & Vanneman, R. (2015), 'India human development survey-ii (ihds-ii), 2011-12', ICPSR36151-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor] pp. 07–31.

Dessí, R. & Ogilvie, S. (2004), 'Social capital and collusion: The case of merchant guilds'.

Devlin, K. (2011), The Man of Numbers: Fibonacci's Arithmetic Revolution, A&C Black.

Dirks, N. B. (2001), Castes of mind: Colonialism and the making of modern India, Princeton University Press.

Dittmar, J. E. (2011), 'Information technology and economic change: The impact of the printing press', The Quarterly Journal of Economics **126**(3), 1133–1172.

Dixit, A. (2003), 'Trade expansion and contract enforcement', Journal of Political Economy **111**(6), 1293–1317.

Dohmen, T., Falk, A., Huffman, D. & Sunde, U. (2012), 'The intergenerational transmission of risk and trust attitudes', The Review of Economic Studies **79**(2), 645–677.

Dollinger, P. (1970), 'The german hansa, trans. and ed. ds ault and sh steinberg'.

Durham, J. W. (1992), 'The introduction of arabic numerals in European accounting', Accounting Historians Journal **19**(2), 25–55.

Edwards, J. R. (2013), A History of Financial Accounting (RLE Accounting), Vol. 29, Routledge.

Eisenstein, E. L. (1980), The printing press as an agent of change, Vol. 1, Cambridge University Press.

Ellison, G. (1994), 'Cooperation in the prisoner's dilemma with anonymous random matching', The Review of Economic Studies **61**(3), 567–588.

Epstein, S. R. (2003), An island for itself: Economic development and social change in late medieval Sicily, Cambridge University Press.

Epstein, S. R. (2004), Town and country in Europe, 1300-1800, Vol. 5, Cambridge University Press.

Epstein, S. R. (2008), 'Craft guilds in the pre-modern economy: a discussion', The Economic History Review **61**(1), 155–174.

Fafchamps, M. (2011), 'Development, social norms, and assignment to task', Proceedings of the National Academy of Sciences **108**(Supplement 4), 21308–21315.

Falk, A. & Fischbacher, U. (2006), 'A theory of reciprocity', Games and Economic Behavior **54**(2), 293–315.

Fanon, F. (1967), 'Black skin, white masks [1952]', New York .

Fehr, E. (2009), 'On the economics and biology of trust', Journal of the European Economic Association **7**(2-3), 235–266.

Fehr, E. & Fischbacher, U. (2003), 'The nature of human altruism', Nature **425**(6960), 785–791.

Fehr, E. & Gächter, S. (2000), 'Fairness and retaliation: The economics of reciprocity', The Journal of Economic Perspectives **14**(3), 159–181.

Fehr, E. & Gächter, S. (2002), 'Altruistic punishment in humans', Nature **415**(6868), 137–140.

Fehr, E., Kirchsteiger, G. & Riedl, A. (1993), 'Does fairness prevent market clearing? an experimental investigation', The Quarterly Journal of Economics pp. 437–459.

Fehr, E. & Schmidt, K. M. (1999), 'A theory of fairness, competition, and cooperation', The Quarterly Journal of Economics pp. 817–868.

Fernandez, R. M., Castilla, E. J. & Moore, P. (2000), 'Social capital at work: Networks and employment at a phone center', American Journal of Sociology **105**(5), 1288–1356.

Fowler, J. H. & Christakis, N. A. (2010), 'Cooperative behavior cascades in human social networks', Proceedings of the National Academy of Sciences **107**(12), 5334–5338.

Francis, A. M. & Tannuri-Pianto, M. (2013), 'Endogenous race in Brazil: Affirmative action and the construction of racial identity among young adults', Economic Development and cultural change **61**(4), 731–753.

Friedman, J. N. & Holden, R. T. (2008), 'Optimal gerrymandering: Sometimes pack, but never crack', The American Economic Review **98**(1), 113–44.

Fukuyama, F. (1995), Trust: The social virtues and the creation of prosperity, number D10 301 c. 1/c. 2, JSTOR.

Gambetta, D. et al. (2000), 'Can we trust trust', Trust: Making and breaking cooperative relations **13**, 213–237.

Gargiulo, M. & Benassi, M. (2000), 'Trapped in your own net? network cohesion, structural holes, and the adaptation of social capital', Organization science **11**(2), 183–196.

Gelderblom, O. (2004), 'The decline of fairs and merchant guilds in the low countries'.

Gelderblom, O. (2013), Cities of Commerce: The Institutional Foundations of International Trade in the Low Countries, 1250-1650, Princeton University Press.

Gelderblom, O., de Jong, A. & Jonker, J. (2011), An admiralty for Asia: Business organization and the evolution of corporate governance in the Dutch republic, 1590–1640, in 'Origins of shareholder advocacy', Springer, pp. 29–60.

Gelderblom, O. & Grafe, R. (2010), 'The rise and fall of the merchant guilds: Rethinking the comparative study of commercial institutions in premodern Europe', Journal of interdisciplinary history **40**(4), 477–511.

Ghosh, P. & Ray, D. (1996), 'Cooperation in community interaction without information flows', The Review of Economic Studies **63**(3), 491–519.

Giavazzi, F., Petkov, I. & Schiantarelli, F. (2014), Culture: Persistence and evolution, Technical report, National Bureau of Economic Research.

Glaeser, E. L. (2005), 'The political economy of hatred', The Quarterly Journal of Economics **120**(1), 45–86.

Glaeser, E. L. & Shleifer, A. (2005), 'The curley effect: The economics of shaping the electorate', Journal of Law, Economics, and Organization **21**(1), 1–19.

Gong, J., McAfee, R. P. & Williams, M. A. (2016), 'Fraud cycles', Journal of Institutional and Theoretical Economics JITE **172**(3), 544–172.

Gould, N. E.-S. J. (1972), 'Punctuated equilibria: An alternative to phyletic gradualism'.

Granovetter, M. (1985), 'Economic action and social structure: The problem of embeddedness', American Journal of Sociology pp. 481–510.

Granovetter, M. S. (1973), 'The strength of weak ties', American Journal of Sociology pp. 1360–1380.

Greif, A. (1994), 'Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies', Journal of Political Economy pp. 912–950.

Greif, A. (2000), 'The fundamental problem of exchange: A research agenda in historical institutional analysis', European Review of Economic History **4**(03), 251–284.

Greif, A. (2006), 'History lessons: The birth of impersonal exchange: The community responsibility system and impartial justice', The Journal of Economic Perspectives **20**(2), 221–236.

Greif, A., Milgrom, P. & Weingast, B. R. (1994), 'Coordination, commitment, and enforcement: The case of the merchant guild', Journal of Political Economy pp. 745–776.

Greif, A. & Tabellini, G. (2017), 'The clan and the corporation: Sustaining cooperation in China and Europe', Journal of Comparative Economics **45**(1), 1–35.

Guiso, L., Sapienza, P. & Zingales, L. (2004), 'The role of social capital in financial development', The American Economic Review **94**(3), 526–556.

Guiso, L., Sapienza, P. & Zingales, L. (2006), 'Does culture affect economic outcomes?', The Journal of Economic Perspectives **20**(2), 23–48.

Guiso, L., Sapienza, P. & Zingales, L. (2008a), 'Alfred marshall lecture social capital as good culture', Journal of the European Economic Association **6**(2-3), 295–320.

Guiso, L., Sapienza, P. & Zingales, L. (2008b), Long term persistence, Technical report, National Bureau of Economic Research.

Guiso, L., Sapienza, P. & Zingales, L. (2016), 'Long-term persistence', Journal of the European Economic Association **14**(6), 1401–1436.

Gulati, R. (1998), 'Alliances and networks', Strategic management Journal **19**(4), 293–317.

Häberlein, M. (2012), The Fuggers of Augsburg: Pursuing wealth and honor in Renaissance Germany, University of Virginia Press.

Habib, I. et al. (2013), 'The agrarian system of Mughal India: 1556-1707', OUP Catalogue .

Hail, L., Tahoun, A. & Wang, C. (2018), 'Corporate scandals and regulation', Journal of Accounting Research **56**(2), 617–671.

Harris, R. (2009), 'The institutional dynamics of early modern Eurasian trade: The commenda and the corporation', Journal of Economic Behavior & Organization **71**(3), 606–622.

Hauert, C., De Monte, S., Hofbauer, J. & Sigmund, K. (2002), 'Volunteering as red queen mechanism for cooperation in public goods games', Science **296**(5570), 1129–1132.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. & McElreath, R. (2001), 'In search of homo economicus: Behavioral experiments in 15 small-scale societies', The American Economic Review **91**(2), 73–78.

Herr, R. (2015), The eighteenth-century revolution in Spain, Princeton University Press.

Hickman, C. B. (1997), 'The devil and the one drop rule: Racial categories, African Americans, and the us census', Michigan Law Review **95**(5), 1161–1265.

Hill, G. F. (1915), The development of Arabic numerals in Europe, Clarendon Press Oxford.

Horn, J. (2015), Economic Development in Early Modern France: The Privilege of Liberty, 1650–1820, Cambridge University Press.

Jafferlot, C. (2003), 'India's silent revolution: The rise of low castes in north Indian politics', Delhi: Permanent Block .

James, E. J. (1893), Education of Business Men, American Banker's Association.

Johnson, N. D. & Mislin, A. A. (2011), 'Trust games: A meta-analysis', Journal of Economic Psychology **32**(5), 865–889.

Jones, G. G. (2017), 'Business history, the great divergence and the great convergence'.

Kadens, E. (2011), 'Myth of the customary law merchant', Texas Law Review **90**, 1153.

Kadens, E. (2014), 'Pre-modern credit networks and the limits of reputation', Iowa Law Review **100**, 2429.

Kadens, E. (2019 forthcoming), 'The use and abuse of reputation', Cardozo Law Review **40**.

Kandori, M. (1992), 'Social norms and community enforcement', The Review of Economic Studies **59**(1), 63–80.

Kandori, M. & Obayashi, S. (2014), 'Labor union members play an olg repeated game', Proceedings of the National Academy of Sciences **111**(Supplement 3), 10802–10809.

Kieser, A. (1989), 'Organizational, institutional, and societal evolution: Medieval craft guilds and the genesis of formal organizations', Administrative Science Quarterly pp. 540–564.

Kilduff, M. & Brass, D. J. (2010), 'Organizational social network research: Core ideas and key debates', Academy of management annals **4**(1), 317–357.

Knack, S. & Keefer, P. (1997), 'Does social capital have an economic payoff? a cross-country investigation', The Quarterly Journal of Economics pp. 1251–1288.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U. & Fehr, E. (2005), 'Oxytocin increases trust in humans', Nature **435**(7042), 673–676.

Kranton, R. E. (1996), 'Reciprocal exchange: A self-sustaining system', The American Economic Review pp. 830–851.

Kreps, D. M. (1996), 'Corporate culture and economic theory', Firms, Organizations and Contracts, Oxford University Press, Oxford pp. 221–275.

Kwass, M. (2006), Privilege and the politics of taxation in eighteenth-century France: Liberté, égalité, fiscalité, Cambridge University Press.

La Porta, R., Lopez-de Silanes, F., Shleifer, A. & Vishny, R. (1999), 'The quality of government', The Journal of Law, Economics, and Organization **15**(1), 222–279.

La Porta, R., Lopez-de Silanes, F., Shleifer, A. & Vishny, R. W. (1997), 'Trust in large organizations', The American Economic Review pp. 333–338.

Lampe, J. R. & Jackson, M. R. (1982), Balkan economic history, 1550-1950: From imperial borderlands to developing nations, Vol. 10, Indiana University Press.

Lariviere, R. W. (1989), 'Justices and paitas: Some ironies in contemporary readings of the Hindu legal past', The Journal of Asian Studies **48**(04), 757–769.

Laursen, K., Masciarelli, F. & Prencipe, A. (2012), 'Regions matter: How localized social capital affects innovation and external knowledge acquisition', Organization Science **23**(1), 177–193.

Lee, A. (2015), 'The origins of identity politics: Caste in colonial India', April **9**, 27.

Liberti, J. M. & Petersen, M. A. (2017), 'Information: Hard and soft', Rev. Corporate Finance Stud. .

Lieberman, E. S. & Singh, P. (2017), 'Census enumeration and group conflict: A global analysis of the consequences of counting', World Politics **69**(1), 1–53.

Lindberg, E. (2009), 'Club goods and inefficient institutions: Why danzig and lübeck failed in the early modern period', The Economic History Review **62**(3), 604–628.

Lucas Jr, R. E. (2017), What was the industrial revolution?, Technical report, National Bureau of Economic Research.

Macy, M. W. & Skvoretz, J. (1998), 'The evolution of trust and cooperation between strangers: A computational model', American Sociological Review pp. 638–660.

Macy, M. W. & Willer, R. (2002), 'From factors to factors: Computational sociology and agent-based modeling', Annual Review of Sociology **28**(1), 143–166.

Mair, J., Martí, I. & Ventresca, M. J. (2012), 'Building inclusive markets in rural Bangladesh: How intermediaries work institutional voids', Academy of Management Journal **55**(4), 819–850.

Malone, T. W., Yates, J. & Benjamin, R. I. (1987), 'Electronic markets and electronic hierarchies', Communications of the ACM **30**(6), 484–497.

Mamdani, M. (2014), When victims become killers: Colonialism, nativism, and the genocide in Rwanda, Princeton University Press.

Marquis, C. & Qian, C. (2013), 'Corporate social responsibility reporting in China: Symbol or substance?', Organization Science **25**(1), 127–148.

Mayer, R. C., Davis, J. H. & Schoorman, F. D. (1995), 'An integrative model of organizational trust', Academy of Management Review **20**(3), 709–734.

McCloskey, D. N. (2016), Bourgeois equality: How ideas, not capital or institutions, enriched the world, University of Chicago Press.

Michalopoulos, S. & Papaioannou, E. (2013), 'Pre-colonial ethnic institutions and contemporary african development', Econometrica **81**(1), 113–152.

Michalopoulos, S. & Papaioannou, E. (2014), 'National institutions and subnational development in africa', The Quarterly Journal of Economics **129**(1), 151–213.

Michalopoulos, S. & Papaioannou, E. (2017), 'Series introduction: Historical legacies and contemporary development', The Long Economic and Political Shadow of History **2**.

Milgrom, P. R., North, D. C. & Weingast, B. R. (1990), 'The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs', Economics & Politics **2**(1), 1–23.

Mokyr, J. (1990), 'Punctuated equilibria and technological progress', The American Economic Review **80**(2), 350–354.

Mokyr, J. (2016), A culture of growth: The origins of the modern economy, Princeton University Press.

Molnár, A. (2007), 'Struggle for the chapel of belgrade (1612–1643). trade and catholic church in ottoman hungary', Acta Orientalia **60**(1), 73–134.

Montgomery, W. W. (1961), Islam and the Integration of Society, JSTOR.

Munshi, K. & Rosenzweig, M. (2015), Insiders and outsiders: Local ethnic politics and public goods provision, Technical report, National Bureau of Economic Research.

Nelson, R. R. & Winter, S. G. (1982), An evolutionary theory of economic change, Harvard University Press.

North, D. C. (1990), Institutions, institutional change and economic performance, Cambridge University Press.

Nowak, M. A. & Sigmund, K. (2005), 'Evolution of indirect reciprocity', Nature **437**(7063), 1291–1298.

Nunn, N. & Wantchekon, L. (2011), 'The slave trade and the origins of mistrust in africa', The American Economic Review **101**(7), 3221–3252.

Ogilvie, S. (2005), 'The use and abuse of trust: Social capital and its deployment by early modern guilds', Jahrbuch für Wirtschaftsgeschichte/Economic History Yearbook **46**(1), 15–52.

Ogilvie, S. (2011), Institutions and European trade: Merchant guilds, 1000–1800, Cambridge University Press.

Ogilvie, S. C. (2006), State corporatism and proto-industry: The Württemberg Black Forest, 1580-1797, Vol. 33, Cambridge University Press.

Orsini, F. (2016), The History of the Book in South Asia, Routledge.

Padgett, J. F. & Powell, W. W. (2012), The emergence of organizations and markets, Princeton University Press.

Platteau, J.-P. (2000), Institutions, social norms, and economic development, Vol. 1, Psychology Press.

Polanyi, K. (1944), 'The great transformation', New York: Rinehart pp. 209–227.

Putnam, R. D., Leonardi, R. & Nanetti, R. Y. (1994), Making democracy work: Civic traditions in modern Italy, Princeton University Press.

Puttevils, J. (2015), Merchants and Trading in the Sixteenth Century: The Golden Age of Antwerp, Routledge.

Rădvan, L. (2009), 'On the medieval urban economy in wallachia'.

Rajan, R. G. & Zingales, L. (1998), 'Which capitalism? lessons form the east Asian crisis', Journal of Applied Corporate Finance **11**(3), 40–48.

Rajan, R. G. & Zingales, L. (2003), 'The great reversals: The politics of financial development in the twentieth century', Journal of Financial Economics **69**(1), 5–50.

Reagans, R. & Zuckerman, E. W. (2001), 'Networks, diversity, and productivity: The social capital of corporate r&d teams', Organization science **12**(4), 502–517.

Reuben, E., Sapienza, P. & Zingales, L. (2009), 'Is mistrust self-fulfilling?', Economics Letters **104**(2), 89–91.

Romanelli, E. & Tushman, M. L. (1994), 'Organizational transformation as punctuated equilibrium: An empirical test', Academy of Management Journal **37**(5), 1141–1166.

Rosenthal, R. W. (1981), 'Games of perfect information, predatory pricing and the chain-store paradox', Journal of Economic theory **25**(1), 92–100.

Rothstein, B. (2011a), 'Anti-corruption: The indirect big bangapproach', Review of International Political Economy **18**(2), 228–250.

Rothstein, B. (2011b), The quality of government: Corruption, social trust, and inequality in international perspective, University of Chicago Press.

Rousseau, D. M., Sitkin, S. B., Burt, R. S. & Camerer, C. (1998), 'Not so different after all: A cross-discipline view of trust', Academy of Management Review **23**(3), 393–404.

Rubin, J. (2017), Rulers, Religion, and Riches: Why the West got rich and the Middle East did not, Cambridge University Press.

Sahle, E. (2015), 'An investigation of early modern quakers business ethics'.

Samarendra, P. (2011), 'Census in colonial India and the birth of caste', Economic and Political Weekly pp. 51–58.

Sandholm, W. H. (2015), Population games and deterministic evolutionary dynamics, in 'Handbook of game theory with economic applications', Vol. 4, Elsevier, pp. 703–778.

Sapienza, P. & Zingales, L. (2012), 'A trust crisis', International Review of Finance **12**(2), 123–131.

Saradesāya, M. (2000), A history of Konkani literature: From 1500 to 1992, Sahitya Akademi.

Saxenian, A. (1996), Regional advantage: Culture and competition in Silicon Valley and Route 128, Harvard University Press.

Schobesberger, N., Arblaster, P., Infelise, M., Belo, A., Moxham, N., Espejo, C. & Raymond, J. (2016), European postal networks, in 'News Networks in Early Modern Europe', Brill, pp. 17–63.

Semmann, D., Krambeck, H.-J. & Milinski, M. (2003), 'Volunteering leads to rock–paper–scissors dynamics in a public goods game', Nature **425**(6956), 390–393.

Sethi, R. & Somanathan, E. (1996), 'The evolution of social norms in common property resource use', The American Economic Review pp. 766–788.

Shiller, R. J. (2017), 'Narrative economics', The American Economic Review **107**(4), 967–1004.
URL: *http://www.aeaweb.org/articles?id=10.1257/aer.107.4.967*

Singleton, T. W. & Singleton, A. J. (2010), Fraud auditing and forensic accounting, Vol. 11, John Wiley & Sons.

Smith, D. E. & Karpinski, L. C. (1911), The Hindu-Arabic Numerals, Ginn and Company.

Smith, R. S. (1940), The Spanish guild merchant: A history of the consulado, 1250-1700, Duke University Press.

Solow, R. M. (1995), 'But verify', The New Republic **11**, 36–39.

Soly, H. (2008), 'The political economy of European craft guilds: Power relations and economic strategies of merchants and master artisans in the medieval and early modern textile industries', International Review of Social History **53**(S16), 45–71.

Somanathan, R. (2010), 'The demand for disadvantage', Culture, institutions, and development: New insights into an old debate p. 125.

Stein, B. H. & Stein, S. J. (2014), Crisis in an Atlantic Empire: Spain and New Spain, 1808-1810, number 2, JHU Press.

Stevenson, B. & Wolfers, J. (2011), 'Trust in public institutions over the business cycle', The American Economic Review **101**(3), 281–87.

Stuart, T. E. & Sorenson, O. (2007), 'Strategic networks and entrepreneurial ventures', Strategic Entrepreneurship Journal **1**(3-4), 211–227.

Tabellini, G. (2008a), 'Presidential address institutions and culture', Journal of the European Economic Association **6**(2-3), 255–294.

Tabellini, G. (2008b), 'The scope of cooperation: Values and incentives', The Quarterly Journal of Economics **123**(3), 905–950.

Tocqueville, A. H. C. C. (1862), Democracy in America, tr. by H. Reeve.

Tracy, J. D. (1993), The Rise of Merchant Empires: Long Distance Trade in the Early Modern World 1350-1750, Vol. 1, Cambridge University Press.

Trivellato, F. (2004), Discourse and practice of trust in business correspondence during early modern period, in 'a paper presented during Economic History Workshop-Yale University, Department of Economics'.

Trivellato, F. (2014), The familiarity of strangers: The Sephardic diaspora, Livorno, and cross-cultural trade in the early modern period, Yale University Press.

Tsai, W. & Ghoshal, S. (1998), 'Social capital and value creation: The role of intrafirm networks', Academy of management Journal **41**(4), 464–476.

Tsvetkova, M. & Macy, M. W. (2014), 'The social contagion of generosity', PloS one **9**(2), e87275.

Uslaner, E. M. (2002), The moral foundations of trust, Cambridge University Press.

Uzzi, B. (1996), 'The sources and consequences of embeddedness for the economic performance of organizations: The network effect', American Sociological Review pp. 674–698.

Uzzi, B. & Lancaster, R. (2003), 'Relational embeddedness and learning: The case of bank loan managers and their clients', Management science **49**(4), 383–399.

Van der Wee, H. (2013), The growth of the Antwerp market and the European economy: Fourteenth-sixteenth centuries, Springer Science & Business Media.

Vanneste, B. & Gulati, R. (2018), 'Generalized trust, external sourcing, and firm performance'.

Varshney, A. (2000), 'Is india becoming more democratic?', The Journal of Asian Studies **59**(1), 3–25.

Verghese, A. (2016), The colonial origins of ethnic violence in India, Stanford University Press.

Weber, M. (1905), The Protestant ethic and the" spirit" of capitalism and other writings.

Yamagishi, T., Terai, S., Kiyonari, T., Mifune, N. & Kanazawa, S. (2007), 'The social exchange heuristic: Managing errors in social exchange', Rationality and Society **19**(3), 259–291.

Yamagishi, T. & Yamagishi, M. (1994), 'Trust and commitment in the united states and japan', Motivation and emotion **18**(2), 129–166.

Yates, J. (1986), 'The telegraph's effect on nineteenth century markets and firms', Business and Economic History pp. 149–163.

Zak, P. J. & Knack, S. (2001), 'Trust and growth', The Economic Journal **111**(470), 295–321.

Appendices

Appendix A

Literature Review: Generalized Trust

Trust has become an important area of study in social sciences. Sociologist Edward Banfield wrote his pioneering book Moral Basis of Backward Society in 1958 and described a Hobbesian lack of social trust in southern Italy towards people outside of strict family circles. He attributed the region's present-day underdevelopment to the "amoral familism", that made people cheat non-family members, expecting others to do the same. Contrastingly, Tocqueville in his book Democracy in America (1840) (Tocqueville 1862) gave a flattering description of civic culture in the United States noting, "Each American knows when to sacrifice some of his private interests to save the rest". Emphasizing the importance of civic culture he wrote "science of association is the mother of science; the progress of all the rest depends upon the progress it has made". Among economists, Arrow (1969) was among the first to identify the value of trust, and he wrote "norms of social behaviour, including ethical and moral codes" may compensate for market failure and "in the absence of trust many opportunities for mutually beneficial cooperation would have to be, foregone." Over the last few decades, a considerable macroeconomic literature has emerged that shows trust has economic significance (Knack & Keefer 1997, La Porta, Lopez-de Silanes, Shleifer & Vishny 1997, Zak & Knack 2001) and levels of trust are different across groups and regions, which are persistent (Putnam, Leonardi & Nanetti 1994, Uslaner 2002, Guiso, Sapienza & Zingales 2016, Tabellini 2008a, Buggle & Durante 2017, Nunn & Wantchekon 2011) and only gradually change (Alesina & La Ferrara 2002, Giavazzi, Petkov & Schiantarelli 2014).

A.1 Trust: Theoretical and Empirical literature

Drawing upon influential sociological research being conducted on social capital (Granovetter 1973, 1985, Coleman 1988), in 1990s two influential sociological works, Putnam, Leonardi & Nanetti (1994) and Fukuyama (1995), emphasized the role of trust on economic development. But, some economists (e.g., Solow (1995)) pointed at the problems in measuring trust as a macroeconomic variable. Some empirical papers (Knack & Keefer 1997, La Porta, Lopez-de Silanes, Shleifer & Vishny 1997, Zak & Knack 2001) began to measure trust using the international World Values Survey (WVS), that asked respondents "Generally speaking, would you say that most people can be trusted, or that you can not be too careful in dealing with people?" The greater the proportion of people in a country answering yes to the question, the greater the

general trust in the country was considered to be. A country's level of trust was found to be correlated with its economic growth (Knack & Keefer 1997), and with efficiency in large organizations (La Porta, Lopez-de Silanes, Shleifer & Vishny 1997). Zak & Knack (2001) developed a moral hazard model, where the ability to trust in an investment broker increased investment and income growth and showed that trust and factors that could affect trust (like formal and informal institutions and social homogeneity) were correlated with income growth.

There were also efforts by theorists to define trust. An extensive literature in sociology and management (surveyed by Rousseau, Sitkin, Burt & Camerer (1998)) attempted to understand trust between relationships. Mayer, Davis & Schoorman (1995) focused on the aspect of vulnerability that trusting someone entails in social exchanges. Yamagishi & Yamagishi (1994) made the distinction between impersonal vs. relationship-based trust, and compared Japan and USA to argue that Japanese society relied more on relationship-based trust, while the United States had higher levels of generalized trust. Gambetta et al. (2000) considered trust to be a subjective probability and Dasgupta (2000) theorized reputation for honesty as a sought after but a fragile commodity. Dasgupta also argued that expectations regarding honest and dishonest conduct could be self-fulfilling affecting trust, while Putnam, Leonardi & Nanetti (1994) found that generalized trust was persistent, and linked high present-day generalized trust in North Italy to persistent civic norms that had developed in the region because of the existence of medieval city republics.

Persistence of trust (Putnam, Leonardi & Nanetti 1994) sparked the interest of researchers in its origins. Alesina & La Ferrara (2002) using individual-level General Social Survey (GSS) data of US localities, found that generalized trust was lower in individuals who had a history of traumatic experience, who belonged to historically discriminated groups, who were economically and educationally unsuccessful, or who were living in neighborhoods that were racially mixed or had high-income inequality. Alesina & La Ferrara (2002) found no significant effect of religion or ethnic origin on trust. Uslaner (2002) argued that trusting others was a persistent moral value in individuals that was passed on by parents and was highly correlated with charity and voluntary work. He argued that trusting attitude was unaffected by personal experience or membership to civic groups (unlike Putnam, Leonardi & Nanetti (1994) who proposed that civic groups generated trust). Guiso, Sapienza & Zingales (2006), using GSS data found a strong positive correlation between generalized trust of US immigrants and the level of generalized trust (measured by WVS) in their country of origin, hinting towards persistence. Guiso, Sapienza & Zingales (2016) found strong empirical evidence of historical persistence of trust attitude, as city republic experience during medieval era in North Italy was strongly and positively correlated to present levels of trust and social capital in the region, as proposed by Putnam, Leonardi & Nanetti (1994). Tabellini (2008a) also found evidence of persistence, with higher levels of trust found in second generation US immigrants originating from countries with more democratic institutions over a century ago. Nunn & Wantchekon (2011) similarly found that experience of slavery had a persistent effect on distrust in Africa. Buggle & Durante (2017) similarly showed that annual variability of weather conditions during 1500- 1750 stimulated trust in Europe.

Tabellini (2008b) developed a theoretical model to explain the persistence and found that impersonal values supporting trust between socially distant individuals could persist by creating

a strong incentive for parents to invest in such value education, and by generating institutions that support such values ((Alesina & La Ferrara 2002) found a high social distance to be a barrier to trust). Guiso, Sapienza & Zingales (2008a) developed a model complementary to the value-based model of Tabellini (2008b), showing that beliefs regarding the honesty of others could be persistent and transferred between generations. Giavazzi, Petkov & Schiantarelli (2014) using GSS data studied the persistence of trust attitude (and other cultural traits) among US immigrants in detail, and found that trust attitude brought by first generation US immigrants did not persist (unlike some other cultural traits) among fourth generation US immigrants, but second generation US immigrants showed some persistence.

So, a body of evidence suggests that values and beliefs regarding trust passed down by parents do persist among individuals (Putnam, Leonardi & Nanetti 1994, Uslaner 2002, Guiso, Sapienza & Zingales 2016, Tabellini 2008a, Buggle & Durante 2017, Nunn & Wantchekon 2011), but values and beliefs themselves can gradually change depending on individual experiences (Alesina & La Ferrara 2002, Giavazzi, Petkov & Schiantarelli 2014). Dohmen, Falk, Huffman & Sunde (2012) looking at German Socio-Economic Panel (SOEP) data find similar results where risk and trust attitudes of children was affected by the attitude of parents as well as the prevailing attitude in the region.

A.2 Trust: Experimental literature

During the time a large body of empirical and theoretical literature was emerging, experimental literature was also emerging that highlighted the nature of trust at a micro level. Berg, Dickhaut & McCabe (1995) developed the influential investment game, which has become a standard in literature as a means to understand and model (Zak & Knack 2001, Guiso, Sapienza & Zingales 2008a) trust and honesty (trustworthiness). In the game, an investor decides to invest a particular amount with a trustee which gets multiplied, and the trustee decides how much of money to return to the investor. An investor that invests a larger amount is considered more trusting, while a trustee that returns a larger amount is considered more honest, where the standard Nash equilibrium outcome is for the investor not to invest, and for the trustee to not return. The origins of the investment game can be traced to other stage games, like dictator game, ultimatum game, prisoner's dilemma game, trust game (Kreps 1996), centipede game (Rosenthal 1981) and exchange game (Fehr, Kirchsteiger & Riedl 1993) each of which have been used by researchers to understand behaviour that does not match the standard rational framework of economics.

The experimental literature (starting with Berg, Dickhaut & McCabe (1995)) has found that unlike the standard game theory equilibrium, investors do invest and trustees do return. But, it is not necessarily clear if trusting and honest conduct reflects social preferences such as altruism or a norm of reciprocity (Cox 2004), and if the social preferences are components of trust or separate from it. So, researchers have explored the preferences that underpin individual behaviour to understand trust. Influential set of papers (on inequity aversion by Fehr & Schmidt (1999)); on altruistic punishment by Fehr & Gächter (2000, 2002); on altruism by Fehr & Fischbacher (2003); on betrayal aversion by Bohnet & Zeckhauser (2004); on reciprocity by Falk & Fischbacher (2006)) developed robust theories of fairness, altruism and reciprocity, and

explored the behavioural underpinnings of trust. A classic paper by Fehr & Schmidt (1999) argued that inequity aversion could explain cooperation as individuals get motivated to inflict costly punishment to free riders. Exploring biological underpinnings of trust, Kosfeld, Heinrichs, Zak, Fischbacher & Fehr (2005) found that higher oxytocin increases trusting attitude. Another influential paper Henrich et al. (2001) looked at 15 small-scale societies and showed that in the sampled societies individuals showed other-regarding preferences and deviated from the standard "homo-economicus" model.

The social preference-based framework is important for understanding human behaviour, and it poses a question whether surveys like WVS measure trust or other preferences. But, the behaviour-based theories do not convincingly explain why there exists considerable heterogeneity in trusting behaviour across societies and their determinants as measured in WVS surveys or trust games (Johnson & Mislin 2011). The body of literature (surveyed by Fehr (2009)) has nonetheless convincingly shown that social preferences of individuals (which may or may not be universal) are an important component of what gets observed as trusting or honest conduct, and a narrow focus on economic preferences or beliefs is not enough to understand the topic, but more work needs to be done to understand how heterogeneity emerges in preferences and beliefs across societies and groups.

Appendix B

Simulations of Emergence of Trust

In this section I look at the conditions needed for the evolution to an impersonal equilibrium, using simulation. In the following example, a population of 10,000 agents had 1,000 exchanges, with a randomly paired agent in each exchange. Of all agents, 5,000 were agents with honest attitude ($p_H = 0.5$), of which a tiny fraction of 100 were honest and impersonal ($p_{HI} = 0.01$). Payoffs were set at $c = 0.5, m = 0.05$ and $h = 0.1$. Impersonal traders became relationship-based on being cheated with probability $k_s = 1$, and relationship-based traders became impersonal traders on being cooperated with probability $k_o = 1$. Agents imitate another agent of different attitude with a probability equal to the product of a timescale constant $w = 0.1$, and the difference in fitness of paired agents (calculated in variable $fitness$, and $Mode = 0$). $S = 100$ iterations of the simulation are run, and the average population of trusting (HI-type) agents (variable $meantrust$), and of dishonest (D-type) agents (variable $meandishonesty$) is calculated over time. Figure B.1 presents the results of the simulation of variation of HI-type and D-type agents.

Note that in the theoretical model (section 2.1) agents consider the average relative fitness of the types and not just the relative fitness of the imitated agent. In a simulation model where average fitness is considered ($Mode = 1$), results exactly match theoretical predictions, as is evident in Figure B.2, which notes 10 simulations ($S = 10$). When initial population is $p_H(0) = 0.49$ and $p_{HI}(0) = 0.01$, because the initial population of honest agents is below the necessary threshold, as theoretically expected according to the threshold condition in Corollary 2, the final equilibrium outcome is $p_{HI} = 0$, i.e., no honest and impersonal agents exist in the system.

If the rate of imitation is increased ($w = 0.5$), while other parameters stay equal, then the impersonal equilibrium is reached sooner, and the cyclic pattern of trust and dishonesty as discussed in section 2.2.3 is visible in Figure B.3.

If relationship-based payoff is increased to, for example $m = 0.2$, while other parameters stay equal, then even a population completely dominated by HT-type agents $p_H = p_{HT} = 0.99$ cannot sustain trust as is evident from Figure B.4. This observation is in line with the utility condition in Corollary 1. Similarly, no trust emerged when initial population of agents with honest attitude was 4,000, i.e., $p_H = 0.4$ (all else being equal or at a higher $w = 0.5$), as a critical threshold was not reached.

A simulation exercise incorporating trust anchors (so that $p_A > 0$) shows that impersonal

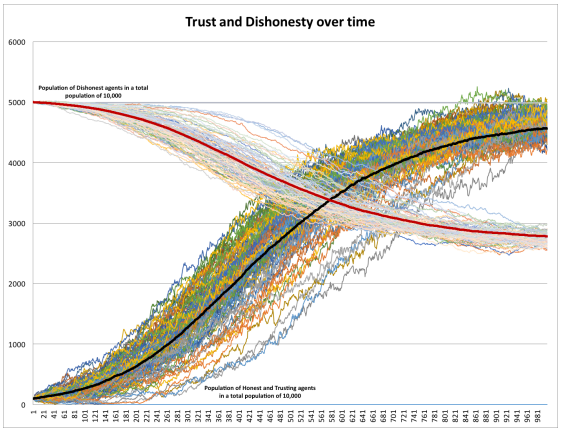

Figure B.1: Y-axis represents number of dishonest (D-type) and honest and trusting (HI-type) agents over time. The individual trends show results of 100 simulations, while thick Black line shows average population of HI-type honest and trusting agents overtime, while thick Red line shows average population of D-type dishonest agents. The simulations match theoretical expectations. $c = 0.5, m = 0.05, h = 0.1, w = 0.1, k_s = 1, k_o = 1$. Total population is 10,000, $p_H(0) = 0.5$ and $p_{HI}(0) = 0.01$. $Mode = 0$.

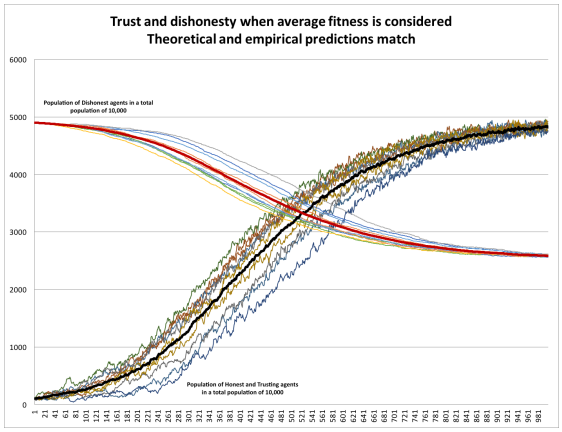

Figure B.2: Y-axis represents number of dishonest (D-type) and honest and trusting (HI-type) agents over time when $Mode = 1$. The individual trends show results of 10 simulations, while thick Black line shows average population of HI-type honest and trusting agents overtime, while thick Red line shows average population of D-type dishonest agents. The simulations match theoretical expectations of equilibrium $p_{HI} = 0.5$ and $p_D = 0.25$. $c = 0.5, m = 0.05, h = 0.1, w = 0.1, k_s = 1, k_o = 1$. Total population is 10,000, $p_H(0) = 0.51$ and $p_{HI}(0) = 0.01$.

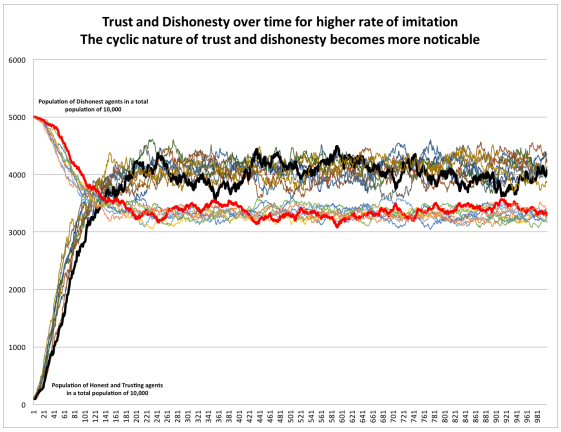

Figure B.3: Y-axis represents number of dishonest (D-type) and honest and trusting (HI-type) agents over time. The individual trends show results of 10 simulations, where thick Black line shows population of HI-type honest and trusting agents overtime for one particular iteration, while thick Red line shows population of D-type dishonest agents of the same iteration. The simulations match theoretical expectations. $c = 0.5, m = 0.05, h = 0.1, w = 0.5, k_s = 1, k_o = 1$. Total population is 10,000, $p_H(0) = 0.5$ and $p_{HI}(0) = 0.01$. $Mode = 0$.

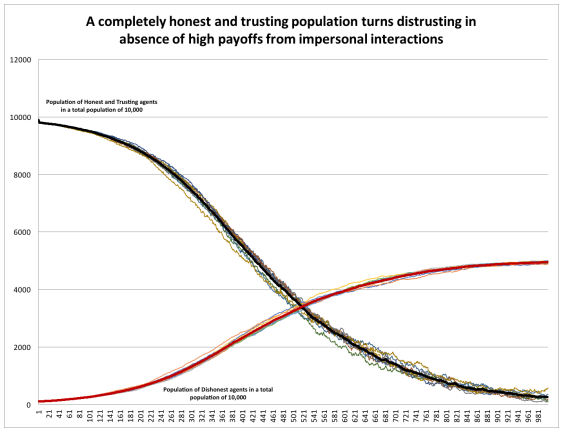

Figure B.4: Y-axis represents number of dishonest (D-type) and honest and trusting (HI-type) agents over time. The individual trends show results of 10 simulations, while thick Black line shows average population of HI-type honest and trusting agents overtime, while thick Red line shows average population of D-type dishonest agents. The simulations match theoretical expectations that $p_{HI} = 0$ as the utility condition in Corollary 1 is not satisfied. $c = 0.5, m = 0.2, h = 0.1, w = 0.1, k_s = 1, k_o = 1$. Total population is 10,000, $p_H(0) = 0.99$ and $p_{HI}(0) = 0.99$. $Mode = 0$.

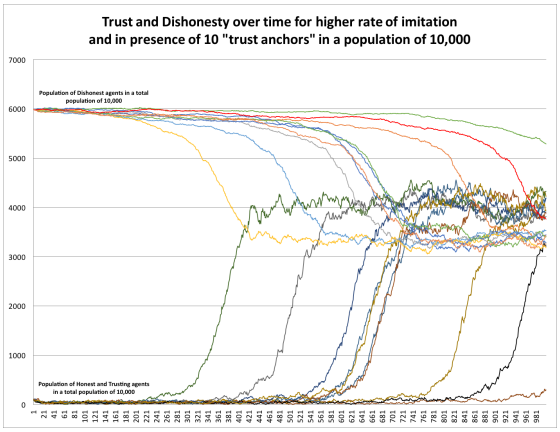

Figure B.5: Y-axis represents number of dishonest (D-type) and honest and trusting (HI-type) agents over time in presence of 10 trust anchors. The individual trends show results of 10 simulations. The simulations match theoretical expectations. $c = 0.5, m = 0.05, h = 0.1, w = 0.5, k_s = 1, k_o = 1$. Total population is 10,000, $p_H(0) = 0.4$ and $p_{HI}(0) = 0.01$, and there are 10 trust anchors. $Mode = 0$

equilibrium can emerge at a lower threshold in the presence of a small fraction of unconditional cooperators (trust anchors). If there were 10 unconditional cooperators in a population of 10,000 in our simulation model, and there were 4,000 honest agents ($p_H = 0.4$) impersonal exchange can still emerge especially when imitation rate ($w = 0.5$) is high, as shown in Figure B.5.

The following MATLAB code was implemented to perform the above simulations.

```
% Cooperation Emergence Case
% Initializing constants

p = 10000; % total population
N = 1000; % number of rounds
S = 100; % number of simulation iterations
w = 0.1; % constant regulating the rate of imitation
c = 0.5;
m = 0.05;
h = 0.1;
k_s = 1;
k_o = 1;
p_H = 5000; % initial population of H-type agents
p_HI = 100; % initial population of HI-type agents
p_A= 0; % number of trust anchors (unconditional cooperators)
trust = zeros(S,N); %Number of HI-type at a given time
dishonesty = zeros(S,N); %Number of D-type at a given time
mode = 0; % mode = 1 for average type fitness, 0 for partner fitness
%simulation

for u = 1:S
% Initializing variables
attitude = 3*ones(p,N); % can take values 1: H, 2: D, 3: A
attitude(1:p_H,1) = 1;
attitude(p_H+1:p-p_A,1) = 2;
```

```
orientation = 4*ones(p,N); % can take values 1: HI, 2: HR, 3: D, 4: A
orientation(1:p_HI,1) = 1;
orientation(p_HI+1:p_H,1) = 2;
orientation(p_H+1:p-p_A,1) = 3;

fitness = zeros(p,N); %Matrix recording the fitness of each agent in
    each interaction
imitation = randi(p,p,N); %A matrix of which agent imitates which
    random agent in a given period
count = zeros(3,N); %Number of H, D, A type attitude at a given time
sumfitness = zeros(3,N); %sum of fitness of H, D, A type attitude at
    a given time
meanfitness = zeros(3,N); %mean fitness of H, D, A type attitude at
    a given time

for t = 1:N-1
   % Interaction
   pairedagent = zeros(p,1); %Matrix recording the paired agents in
       a given interaction
   r = randperm(p); % randomly permute all p agents, first 1:p/2
       agents will be paired to remaining p/2+1:end agents, in that
       order
   pair1 = r(1:floor(p/2));
   pair2 = r(floor(p/2)+1:end);
   pairedagent(pair1) = pair2;
   pairedagent(pair2) = pair1;

   for i = 1:p

      if (orientation(i,t) == 1 && or((attitude(pairedagent(i),t) ==
          1),(attitude(pairedagent(i),t) == 3)))
         fitness(i,t) = c;
         sumfitness(1,t)=sumfitness(1,t)+c;
         orientation(i,t+1) = 1;
      else if (orientation(i,t) == 1 && attitude(pairedagent(i),t)
          == 2)
         fitness(i,t) = -h;
         sumfitness(1,t)=sumfitness(1,t)-h;
         if (rand <= k_s) % with probability k_s
            orientation(i,t+1) = 2;
         else
            orientation(i,t+1) = 1;
         end
      else if (orientation(i,t) == 2 &&
          or((orientation(pairedagent(i),t) ==
          1),(orientation(pairedagent(i),t) == 4)))
         fitness(i,t) = c;
         sumfitness(1,t)=sumfitness(1,t)+c;
```

```
        if (rand <= k_o) % with probability k_o
           orientation(i,t+1) = 1;
        else
           orientation(i,t+1) = 2;
        end
    else if (orientation(i,t) == 2 &&
        or((orientation(pairedagent(i),t) ==
        2),(orientation(pairedagent(i),t) == 3)))
        fitness(i,t) = m;
        sumfitness(1,t)=sumfitness(1,t)+m;
        orientation(i,t+1) = 2;
    else if (orientation(i,t) == 3 &&
        or((orientation(pairedagent(i),t) ==
        1),(orientation(pairedagent(i),t) == 4)))
        fitness(i,t) = c+h;
        sumfitness(2,t)=sumfitness(2,t)+c+h;
        orientation(i,t+1) = 3;
    else if (orientation(i,t) == 3 &&
        or((orientation(pairedagent(i),t) ==
        2),(orientation(pairedagent(i),t) == 3)))
        fitness(i,t) = m;
        sumfitness(2,t)=sumfitness(2,t)+m;
        orientation(i,t+1) = 3;
    else if (orientation(i,t) == 4 &&
        or((attitude(pairedagent(i),t) ==
        1),(attitude(pairedagent(i),t) == 3)))
        fitness(i,t) = c;
        sumfitness(3,t)=sumfitness(3,t)+c;
        orientation(i,t+1) = 4;
    else if (orientation(i,t) == 4 && attitude(pairedagent(i),t)
        == 2)
        fitness(i,t) = -h;
        sumfitness(3,t)=sumfitness(3,t)-h;
        orientation(i,t+1) = 4;
        end
        end
        end
        end
        end
        end
        end
    end
end
count(1,t) = sum(attitude(:,t) == 1);
count(2,t) = sum(attitude(:,t) == 2);
count(3,t) = sum(attitude(:,t) == 3);
meanfitness(1,t)=rdivide(sumfitness(1,t),count(1,t));
meanfitness(2,t)=rdivide(sumfitness(2,t),count(2,t));
```

```
   meanfitness(3,t)=rdivide(sumfitness(3,t),count(3,t));
   % Imitation
   if mode == 0
      for i = 1:p
         if ((fitness(imitation(i,t),t) > fitness(i,t)) &&
            and(attitude(i,t)~=3,attitude(imitation(i,t),t)~=3) &&
            (rand <= w*(fitness(imitation(i,t),t) - fitness(i,t))))
               attitude(i,t+1) = attitude(imitation(i,t),t);
               orientation(i,t+1) = orientation(imitation(i,t),t);
         else
               attitude(i,t+1) = attitude(i,t);
         end
      end
   else
      for i = 1:p
         if ((meanfitness((attitude(imitation(i,t),t)),t) >
            meanfitness((attitude(i,t)),t)) &&
            and(attitude(i,t)~=3,attitude(imitation(i,t),t)~=3) &&
            (rand <= w*(meanfitness((attitude(imitation(i,t),t)),t)
            - meanfitness((attitude(i,t)),t))))
               attitude(i,t+1) = attitude(imitation(i,t),t);
         if attitude(i,t+1) ~= attitude(i,t);
               orientation(i,t+1) = orientation(imitation(i,t),t);
         end
         else
               attitude(i,t+1) = attitude(i,t);
         end
      end
   end

end
trust(u,:) = sum(orientation == 1);
dishonesty(u,:) = sum(orientation == 3);
end
meantrust = mean(trust);
meandishonesty = mean(dishonesty);
```

Appendix C

Proofs on Existence of Genrealized Trust

C.1 Proof of Stationarity and Stability in Proposition 1

Given the dynamics in Equations 2.3 and 2.4 , s_i is a stationary point s.t. $\frac{dp_H(t)}{dt} = 0$ and $\frac{dp_{HI}(t)}{dt} = 0$. So:

$$\begin{aligned} s_1 &= (p_H = 1, p_{HI} = 1); \\ s_2 \in S_2 &= (p_H \in [0,1], p_{HI} = 0); \\ s_3 &= \begin{pmatrix} p_H = 1 - \frac{k_o h}{k_s(c-m)-k_o m}, \\ p_{HI} = 1 - \frac{(k_o+k_s)h}{k_s(c-m)-k_o m} \end{pmatrix} \end{aligned}$$

where $p_{HI} > 0$ if $\frac{k_s}{k_{HI}+k_o} > \frac{h+m}{c}$.

A stationary point s_i is stable if system of Equations 2.3 and 2.4 $F(p_H, p_{HI}) = \begin{bmatrix} \frac{dp_H(t)}{dt} \\ \frac{dp_{HI}(t)}{dt} \end{bmatrix}$ at s_i has Jacobian matrix $J_{F(p_H,p_{HI})}(s_i)$ with Trace $tr(J_{F(p_H,p_{HI})}(s_i)) < 0$ and Determinant $det(J_{F(p_H,p_{HI})}(s_i)) \geq 0$.

It is found:

- At point s_1 for $J_{F(p_H,p_{HI})}(s_1)$[1], $det(J_{F(p_H,p_{HI})}(S_1)) < 0$. So s_1 is unstable equilibrium.
- At points $s_2 \in S_2$ for $J_{F(p_H,p_{HI})}(s_2)$[2], $det(J_{F(p_H,p_{HI})}(s_2)) = 0$. So s_2 is stable but not asymptotically stable if $tr(J_{F(p_H,p_{HI})}(s_2)) < 0$. $tr(J_{F(p_H,p_{HI})}(s_2)) < 0$ if at s_2 $p_H < p_H^{cr} = (\frac{k_s}{k_s+k_o})$. So, points with $p_H > p_H^{cr}$ in S_2 are unstable.
- At point s_3 for $J_{F(p_H,p_{HI})}(s_3)$[3], $det(J_{F(p_H,p_{HI})}(s_3)) > 0$ and $tr(J_{F(p_H,p_{HI})}(s_3)) < 0$ when $w < \frac{k_s c-(k_s+k_o)m}{h(c-f_{HI}(c-m))}$ where $f_{HI} = \frac{p_{HI}}{p_H}$. So, s_3 is a stable equilibrium when w is small.

[1] $J_{F(p_H,p_{HI})}(s_1) = \begin{bmatrix} wh & 0 \\ (k_o+k_s)+wh & -k_o \end{bmatrix}$

[2] $J_{F(p_H,p_{HI})}(s_2) = \begin{bmatrix} 0 & wp_D(cp_H-(m+h)) \\ 0 & k_o p_H - k_s p_D \end{bmatrix}$

[3] $J_{F(p_H,p_{HI})}(s_3) = \begin{bmatrix} wp_D p_{HI} c & -wp_H p_{HI}(c-m) \\ (k_s+k_o)p_{HI} + f_{HI} wp_D p_{HI} c & -k_o p_{HI} - f_{HI} wp_D p_{HI}(c-m) \end{bmatrix}$

C.2 Proof of Proposition 2

Ruler removes monopoly of merchant guilds only if merchant guild elites cannot block such transition by raising relationship-based payoff (m). As merchant guild elites can at most set $m \leq m^{max}$, so when Corollary 1 and 2 are satisfied for $m = m^{max}$, then transition can occur without being blocked by merchant guild elites. Corollary 1 and 2 are necessary conditions for ruler to remove monopoly of merchant guilds, which are true if $p_H(0) > \frac{(1+\lambda_m)(m^{max}+h)}{c+\lambda_m m^{max}+(1+\lambda_m)h}$ and $c > \frac{(1+\lambda_m)}{\lambda_m}h + m^{max}$ (Region 1 of Figure 2.5). In all other conditions, elites can block such a transition by setting $m \in [m^{min}, m^{max})$. When ruler has not removed monopoly, which is true if $p_H(0) \leq \frac{(1+\lambda_m)(m^{max}+h)}{c+\lambda_m m^{max}+(1+\lambda_m)h}$ or $c \leq \frac{(1+\lambda_m)}{\lambda_m}h + m^{max}$, then loss from being cheated $h^i = \beta h$. If merchant guild elites set $m = m^{min}$, traders will not transition if $p_H(0) \leq \frac{(1+\lambda_m)(m^{min}+\beta h)}{c+\lambda_m m^{min}+(1+\lambda_m)\beta h}$ and $c \leq \frac{(1+\lambda_m)\beta}{\lambda_m}h + m^{min}$ (Region 3 of Figure 2.5). If there exists a region which lies outside of Region 1 and 3 (Region 2 of Figure 2.5), then in the region merchant guild elites set $m \in (m^{min}, m^{max})$. The above proves Proposition 2

C.3 Trust Anchors

In presence of trust anchors i.e. impersonal and impartial institutions that exchange honestly and impersonally (e.g. bishops in North Italy (Guiso, Sapienza & Zingales 2016)) regardless of past experience in society such that $k_o p_{HR}(t)\epsilon$ traders are always acting as HI-type traders where $\epsilon \to 0$, evolutionary and behavioural dynamics is given as:

$$\begin{aligned}\frac{dp_H(t)}{dt} = -\frac{dp_D(t)}{dt} = wp_D(t)p_{HI}(t)(p_H(t)c \\ -(c-m)p_{HI}(t) - (m+h))\end{aligned} \tag{C.1}$$

$$\frac{dp_{HI}(t)}{dt} = k_o p_{HR}(t)(p_{HI}(t)+\epsilon) - k_s p_{HI}(t)p_D(t) + \frac{p_{HI(t)}}{p_H(t)}\frac{dp_H(t)}{dt} \tag{C.2}$$

Finding stationary points S_i^o s.t. $\frac{dp_H(t)}{dt} = 0$ and $\frac{dp_{HI}(t)}{dt} = 0$ where S^o is stable if system of equation [5] and [6] $F^o(p_H, p_{HI}) = \begin{bmatrix} \frac{dp_H(t)}{dt} \\ \frac{dp_{HI}(t)}{dt} \end{bmatrix}$ at S_i^o has Jacobian matrix $J_{F^o(p_H,p_{HI})}(S_i^o)$ with Trace $tr(J_{F^o(p_H,p_{HI})}(S_i^o)) < 0$ and Determinant $det(J_{F^o(p_H,p_{HI})}(S_i^o)) \geq 0$.

So there exist two sets of stable equilibria:

1. $p_H = 0$ represents an asymptotic equilibrium.
2. $p_H = 1 - \frac{k_o h}{k_s(c-m)-k_o m}$ and $p_H = 1 - \frac{(k_o+k_s)h}{k_s(c-m)-k_o m}$ represents an asymptotic equilibrium if $w < \hat{w}$ distinct from (i) if $\frac{k_s}{k_o+k_s} > \frac{h+m}{c}$.
3. If $p_H(0) > \frac{h+m}{c}$ and $p_{HI}(0) \to 0$ when $\frac{k_s}{k_o+k_s} > \frac{h+m}{c}$ and $w < \hat{w}$ the system converges to the asymptotic equilibrium (2) and (1) otherwise.

Appendix D

Case Study: A tale of two cities - Hamburg and Lübeck

The cities of Hamburg and Lübeck have an interwoven and eventful history[1]. Whereas Lübeck offers an example of how dominant cities may become unattractive and decline when they end up serving the interests of a privileged few and refuse to change, Hamburg serves as a tale of how cities can reinvent themselves by changing with the times. The cities, in the north of Germany, are just 65 kilometres (40 miles) apart. Yet, given the shape of the Jutland peninsula, Hamburg lies on the Atlantic coast, while Lübeck lies on the Baltic (see Figure 3.3).

Both Hamburg and Lübeck were members of the medieval Hanseatic League. This league was a federation of merchant guilds - an association of wholesale traders that had a privileged regional monopoly over trade - that traded across northern Europe. These guilds were the dominant way of doing trade across medieval Europe.

Within the Hanseatic League, Hamburg and Lübeck were sprawling cities. While Lübeck served as the chief Baltic entrepôt of Europe, Hamburg provided the Hanseatic League with access to the Atlantic. Between the two cities lay the elaborate Elbe river and canal system to facilitate the transport of goods. Lübeck prided itself on being the Queen of the Hanseatic until the fifteenth century, while Hamburg was its smaller, allied partner.

Times became rocky for the Hanseatic system in the fifteenth century. This was in part due to the rise of the Dutch, who were once beneficiaries of trade with the Hanseatic but were now the league's seafaring competitors. Before the arrival of the Dutch, almost all trade to and from the Baltic passed through Lübeck. Likewise, Hamburg benefited from being the sole major Atlantic port of the Hanseatic. The link between Lübeck and Hamburg was a crucial route for trade in the north. But, the Dutch began to trade with the Baltic by navigating around the Jutland peninsula and through the Sound (Øresund). Thus, the Dutch soon began to reach the Baltic shores without the need to visit Hamburg or Lübeck. This competition from the Dutch disrupted the two cities centuries-old domination over trade between the Atlantic and the Baltic.

How did the two cities respond? Differently. Lübeck responded to this competition with the Dutch by giving more privileges to its own merchants and by leading a persistent attempt to disrupt the Dutch trade through the Sound (which included taking part in the Dano Hanseatic

[1]Originally published in *Promarket* blog of the Stigler Center at the University of Chicago Booth School of Business

War of 1426-35 and the Dutch Hanseatic War of 1438-41). In contrast, while Hamburg initially was an ally to Lübeck in its resistance to the Dutch (including in the two wars), it eventually began to diverge from its partner in the sixteenth century. Hamburg opened trade to all locals and non-locals, and instead of resisting this rising Dutch trade, it adapted itself perfectly to the changing situation and moved toward an open system of trade that welcomed diverse merchants (Dollinger (1970), p. 355). Thus, Hamburg internally reformed, and the centuries-old privileges that a few of its merchants enjoyed declined, especially in the sixteenth century. This made a difference.

Hamburg over time became integrated with the Atlantic trading system to its west, and expanded as a major Atlantic entrepôt of northern Europe. Traders from around Europe could trade in Hamburg, and this attracted more merchants and more trade. And what about Lübeck? While the traditional traders held onto their privileges in the city, Lübeck as a whole declined slowly but persistently, especially after the sixteenth century.

Dollinger (1970) (p. 372), recounting the decline of Lübeck in his classic book The German Hansa, wrote:

> The spirit of conservatism, which seems to have been particularly strong in Lübeck, was one of the main causes of Hanseatic weakness in and after the fifteenth century. As foreign competition intensified, the only remedy proposed was an even stricter regulation of trade. In striving to limit the business operations of foreigners in her domain, the Hansa hindered the activities of her own merchants, forbidding them to go into partnership with foreigners or to trade on credit.

This contrasts with Hamburg, about which Dollinger (1970) (p. 356) wrote:

> To a large extent the greatness of Hamburg was the work of foreign merchants. They came in large numbers and they stayed. The town did all it could to facilitate their business. Foreigners were authorized to trade freely among themselves and to enter into partnerships with Hamburg merchants.

One can make a simpler argument that Lübeck could not compete with Hamburg in the Atlantic trade era, simply because it was not an Atlantic city. Yet, the story seems to be more nuanced. It is indeed true that not being directly on the Atlantic coast was a hurdle for Lübeck that Hamburg did not have. Geographical differences must have played a role in the institutional divergence between Lübeck and Hamburg. Conditions were more favourable for Hamburg to develop open markets of the kind that had developed in its Atlantic neighbourhood, like in Amsterdam and Antwerp.

Yet, it was the closed and protective nature of trade in Lübeck that was key to its decline[2]. Lübeck ceased to be an attractive destination for new merchants, as it protected its existing merchants from foreigners and as competition increased from trade through the Sound. It is not as if Lübeck could not have competed or cooperated with Hamburg (and with the Danes who controlled the Sound) and remained a key player in the emerging Atlantic trading system.

[2] see Lindberg (2009) for a look at the political economy of Lübeck.

After all, Lübeck was the larger and the richer of the two cities, it was the chief Baltic entrepôt that had elaborate waterways to the Atlantic, and in the fifteenth century it was still the Queen of the Hansa with significant political clout. It could have welcomed traders from the Atlantic and the Baltic, and remained an international trading hub, instead of ceding influence to its close and smaller neighbour, Hamburg. Yet, Lübeck chose to use its clout to protect its own privileged merchants in a vain attempt to stifle unguilded Dutch and English competition (Ogilvie (2011) p. 87).

It also continued using its clout to restrict trade of Dutch merchants through the Sound (e.g., in 1470), but failed in its attempts, as other Hanseatic towns (especially on the eastern and western flank) refused to cooperate (Bes, Frankot & Brand (2007), p. 5). Hence, Lübeck attempted to conserve the old ways of Baltic trade via privileged monopolies and restrictions, while Hamburg repealed the monopolies of its local merchant guild, and permitted non-locals to trade freely among themselves and form partnerships with locals (Ogilvie (2011) p. 87).

While historical anecdotes should not be used as conclusive evidence, they can serve as cautionary tales. In this case, Lübeck serves as an example of how dominant cities may become unattractive and decline when they end up serving the interests of a privileged few and refuse to change. In contrast, Hamburg serves as a tale of how cities can reinvent themselves by changing with the times and by investing in pro-market institutions that make them attractive centres for trade and commerce for a broad and diverse base of potential traders.

Appendix E

Data on Cities of Europe

E.1 50 largest European (excluding Russian) cities in the fourteenth, fifteenth and sixteenth century

Table E.1: Largest European cities in 1400. Source: Bairoch, Batou & Chevre (1988).

City	Population (,000) 1400	Rank 1400
$PARIS$	275	1
$BRUGGE$	125	2
$GENOVA$	100	3
$GRANADA$	100	3
$VENEZIA$	100	3
$PRAHA$	95	6
$MILANO$	90	7
$SEVILLA$	70	8
$GENT$	56	9
$FIRENZE$	55	10
$LISBOA$	55	10
$BOLOGNA$	45	12
$LONDON$	45	12
$NAPOLI$	45	12
$TOLEDO$	45	12
$SALONIKA$	42	16
$CORDOBA$	40	17
$FERRARA$	40	17
$KOELN$	40	17
$MALAGA$	40	17
$TOURNAI$	40	17
$BARCELONA$	38	22

VALENCIA	36	23
ATHINAI	35	24
CREMONA	35	24
ROUEN	35	24
VERONA	35	24
PADOVA	34	28
LYON	33	29
ROMA	33	29
AQUILA	30	31
AVIGNON	30	31
BELGRADE	30	31
BORDEAUX	30	31
LIEGE	30	31
ORLEANS	30	31
BRESCIA	27	37
BURGOS	27	37
PALERMO	27	37
BRUXELLES	26	40
ALMERIA	25	41
ANGERS	25	41
LUEBECK	25	41
MANTUA	25	41
METZ	25	41
PIACENZA	25	41
TIRGOVISTE	25	41
ERFURT	24	48
LUCCA	23	49
PISA	23	49
VALENCIENNES	23	49

Table E.2: Largest European cities in 1500. Source: Bairoch, Batou & Chevre (1988).

City	Population (,000) 1500	Rank 1500
PARIS	225	1
NAPOLI	125	2
MILANO	100	3
VENEZIA	100	3
GRANADA	70	5
PRAHA	70	5
LISBOA	65	7

TOURS	60	8
GENOVA	58	9
FIRENZE	55	10
GENT	55	10
PALERMO	55	10
ROMA	55	10
BOLOGNA	50	14
BORDEAUX	50	14
LONDON	50	14
LYON	50	14
ORLEANS	50	14
SKOPJE	50	14
TIRGOVISTE	50	14
VERONA	50	14
BRESCIA	49	22
KOELN	45	23
MARSEILLE	45	23
SEVILLA	45	23
FERRARA	42	26
MALAGA	42	26
VALENCIA	42	26
CREMONA	40	29
ROUEN	40	29
NUERNBERG	38	31
BRUGGE	35	32
CORDOBA	35	32
TOURNAI	35	32
BRUXELLES	33	35
BOURGES	32	36
POZNAN	32	36
TOLEDO	32	36
ALMERIA	30	39
ANTWERPEN	30	39
AUGSBURG	30	39
BELGRADE	30	39
GDANSK	30	39
LUCCA	30	39
PLOVDIV	30	39
TOULOUSE	30	39
VALLADOLID	30	39

$PADOVA$	29	48
$LILLE$	26	49
$CAEN$	25	50
$LUEBECK$	25	50
$MANTUA$	25	50
$MECHELEN$	25	50
$MESSINA$	25	50
$MURCIA$	25	50
$PIACENZA$	25	50
$PLASENCIA$	25	50
$WROCLAW$	25	50

Table E.3: Largest European cities in 1600. Source: Bairoch, Batou & Chevre (1988).

City	Population (,000) 1600	Rank 1600
$PARIS$	300	1
$NAPOLI$	275	2
$LONDON$	200	3
$VENEZIA$	151	4
$SEVILLA$	135	5
$LISBOA$	130	6
$MILANO$	120	7
$PALERMO$	105	8
$PRAHA$	100	9
$ROMA$	100	9
$GDANSK$	80	11
$TOLEDO$	80	11
$FIRENZE$	76	13
$ROUEN$	70	14
$GRANADA$	69	15
$MADRID$	65	16
$TOURS$	65	16
$VALENCIA$	65	16
$BOLOGNA$	63	19
$GENOVA$	63	19
$BELGRADE$	55	21
$SKOPJE$	55	21
$AMSTERDAM$	54	23
$BRUXELLES$	50	24
$MESSINA$	50	24

$SALONIKA$	50	24
$WIEN$	50	24
$ANTWERPEN$	47	28
$AUGSBURG$	45	29
$MARSEILLE$	45	29
$SARAJEVO$	45	29
$VERONA$	45	29
$LEIDEN$	44	33
$VALLADOLID$	41	34
$BORDEAUX$	40	35
$HAMBURG$	40	35
$KOBENHAVN$	40	35
$KOELN$	40	35
$MAGDEBURG$	40	35
$NUERNBERG$	40	35
$ORLEANS$	40	35
$TOULOUSE$	40	35
$WROCLAW$	40	35
$HAARLEM$	39	44
$BRESCIA$	36	45
$CREMONA$	36	45
$LYON$	35	47
$ATHINAI$	33	48
$FERRARA$	33	48
$JEREZ - DE - LA - FRONTERA$	33	48
$LILLE$	33	48
$PADOVA$	33	48
$PIACENZA$	33	48
$RENNES$	33	48

E.2 Type of Institution by city

Table E.4: Economic Institutions in European cities

City	Note	Code
$ALMERIA$	Spanish cities were feudal in character and cities began receiving merchant guild permissions only in late medieval period. Alemeria was not granted a Consulado (merchant guild) and thus lost out in patronage to other Spanish cities (Smith (1940), p. 15).	R2

$AMSTERDAM$	Amsterdam rose during the sixteenth century and did not have a major local merchant guild barring some areas of commerce, and prohibited alien merchant guilds from gaining monopoly privileges (Ogilvie (2011), p. 32).	I2
$ANGERS$	No specific information on the city. French cities resisted any guild reform until at least the eighteenth century. Rouen, Nancy, Metz, Roussillon and Paris had some partial reform during the eighteenth century. (Horn (2015), p. 224).	R3
$ANTWERPEN$	In Antwerp, merchant guilds began to decline around 1500, with a shift away from dominance by the merchant guild towards one in which individual merchants could conduct trade without the restrictions often imposed by guilds. (Ogilvie (2011), p. 32)	I2
$AQUILA$	Feudal rights were were being enjoyed by elites at least until the eighteenth century in Aquila (Epstein (2004), p. 328).	R2
$ATHINAI$	No specific information on nature of institutions in Athens, but given that ruling Ottoman empire had a system of granting privileges (Ogilvie 2011), it is assumed economic organization in Ottoman influenced Athens was similar.	R3
$AUGSBURG$	Augsburg was dominated by Fuggers who continued their influence throughout the sixteenth century albeit some anti monopoly dissent caused by Martin Luther inspired writings. In an edict in 1525 Charles V weighed in on monopoly issue in favor of the merchant companies, arguing that private interest and the common good were compatible (Häberlein 2012).	R1
$AVIGNON$	No specific information on the city. French cities resisted any guild reform until at least the eighteenth century. Rouen, Nancy, Metz, Roussillon and Paris had some partial reform during the eighteenth century (Horn (2015), p. 224).	R3
$BARCELONA$	In Catalonia (Barcelona) merchant guilds enjoyed privileges in early modern period (Ogilvie (2011), pp. 33-34).	R1
$BELGRADE$	Belgrade had merchant guilds during the sixteenth-seventeenth century. Merchant colonies of Bosnian merchants blocked the settlement of Ragusan merchant colonies in Belgrade and other Hungarian and Slavonian towns (Ogilvie (2011), p. 124).	R1

$BOLOGNA$	Silk and hemp merchant guilds continued to enjoy merchant guild monopolies in the eighteenth century Bologna (Ogilvie (2011), p. 113).	R1
$BORDEAUX$	No specific information on the city. French cities resisted any guild reform until at least the eighteenth century. Rouen, Nancy, Metz, Roussillon and Paris had some partial reform during the eighteenth century (Horn (2015), p. 224).	R3
$BOURGES$	No specific information on the city. French cities resisted any guild reform until at least the eighteenth century. Rouen, Nancy, Metz, Roussillon and Paris had some partial reform during the eighteenth century (Horn (2015), p. 224).	R3
$BRESCIA$	In regions of Bergamo and Brescia, in the sixteenth century, strategic considerations led Venice to make significant concessions of autonomy that freed such communities from direct dominion by the cities (Epstein (2004), p. 301).	U1
$BRUGGE$	In Bruges, merchant guilds continued their privileges only until the mid sixteenth century (Ogilvie (2011), p. 12).	I2
$BRUXELLES$	Antwerp emerged as a major trading market for traders of the region Ogilvie (2011). No evidence exists such that individual merchant guilds successfully blocked locals from doing trade through Antwerp. The closeness of Belgian cities hindered them from enforcing exclusive privilege (Ogilvie (2011), pp. 31-32).	U2
$BURGOS$	In Spain, local merchant guilds such as those of Burgos and Bilbao used staple privileges to impose costs on foreign traders throughout the early modern period (Ogilvie (2011), p. 70-71).	R1
$CAEN$	Along with records of persistent guild system in Caen until at least the eighteenth century Europe (Horn (2015), p. 224), there are accounts about Caen Butcher guilds, where workers complained that guild masters taxed them unfairly (Kwass (2006), pp. 136-137).	R1
$CORDOBA$	No specific information on the city. But, Andalusian merchant guilds used their privileges to guard their interests in the eighteenth century (Stein & Stein (2014), p. 95-96).	R3

$CREMONA$	No specific account of the city. In 1593 Milanese Lombardy, representatives of the village areas managed to have the privileged allocation of raw silk suspended, even though cities maintained their monopoly over most of the activities (Epstein (2004), p. 308).	U3
$ERFURT$	No specific information on the city, but Erfurt was acquiring landed territories (Epstein (2004), p. 209) pointing towards a possibility of extending powers of urban elites which in line with the general trend of cities extending their commercial privileges (see Chapter 9 on Germany, Epstein (2004)).	R3
$FERRARA$	No specific information on Ferrara, but Papal states were unable or unwilling to grant rural communities autonomy at cost of weakening the Bolognese oligarchy (Epstein (2004), p. 315).	R3
$FIRENZE$	In the sixteenth century, Florence had a high level of political centralism, which led to hollowing out of region's urban network (Epstein (2004), pp. 295-296) (Lucca and Pisa), and merchant guild privileges persisted (Ogilvie (2011), pp. 33-34).	R1
$GDANSK$	In the sixteenth century, Gdansk excluded foreign merchants from trading without the agreement of existing burghers. (Ogilvie (2011), p. 55)	R1
$GENOVA$	Genoa allowed all citizens to trade freely, and had liberal requirements for earning citizenship (Ogilvie (2011), p. 53)	I1
$GENT$	Antwerp emerged as a major trading market for traders of the region Ogilvie (2011). No evidence exists such that individual merchant guilds successfully blocked locals from doing trade through Antwerp. The closeness of Belgian cities hindered them from enforcing exclusive privilege (Ogilvie (2011), pp. 31-32).	U2
$GRANADA$	No specific information on the city. But, Andalusian merchant guilds used their privileges to guard their interests in the eighteenth century. (Stein & Stein (2014), pp. 95-96)	R3
$HAARLEM$	After 1581, new Dutch republic refused to grant privileges to merchants, and literature does not mention Haarlem and Leiden (unlike Dordrich and Middleberg) relying on staple rights (Ogilvie (2011), p. 183). No explicit mention of the city removing monopoly privileges.	U3

$HAMBURG$	Hamburg removed monopoly privileges of local merchant guilds while other cities of the German Hansa, continued with such privileges (Ogilvie (2011), p. 87).	I2
$JEREZ - DE - LA - FRONTERA$	No specific information on the city. But, Andalusian merchant guilds used their privileges to guard their interests in the eighteenth century (Stein & Stein (2014), pp. 95-96).	R3
$KOBENHAVN$	There is a record of establishment of a merchant guild in Copenhagen in 1742 (Andersen (2011), p. 61).	R1
$KOELN$	Until the early nineteenth century, Cologne merchant guilds maintained monopoly privileges, and used the privileges to block non-local merchants from trading in the hinterlands (Ogilvie (2011), p. 69).	R1
$LEIDEN$	After 1581, new Dutch republic refused to grant privileges to merchants, and literature does not mention Haarlem and Leiden (unlike Dordrich and Middleberg) relying on staple rights (Ogilvie (2011), p. 183). No explicit mention of the city removing monopoly privileges.	U3
$LIEGE$	Antwerp emerged as a major trading market for traders of the region Ogilvie (2011). No evidence exists such that individual merchant guilds successfully blocked locals from doing trade through Antwerp. The closeness of Belgian cities hindered them from enforcing exclusive privilege (Ogilvie (2011), pp. 31-32).	U2
$LILLE$	Merchant guild existed until the eighteenth century in Lille; in the eighteenth century, merchant guilds of Lille worked to restrict independence of rural brokers in the textile industry (Ogilvie (2011), p. 37).	R1
$LISBOA$	Portuguese merchants enjoyed monopolies over transatlantic and European trade, unlike English and Dutch merchants (Ogilvie (2011), p. 37).	R1
$LONDON$	The London livery companies (guilds of merchants and craftsmen) found it increasingly difficult to enforce their economic privileges as the sixteenth century progressed (Ogilvie (2011), pp. 32-33).	I2
$LUCCA$	In the sixteenth century, Florence had a high level of political centralism, which led to hollowing out of region's urban network (Epstein (2004), pp. 295-296) (Lucca and Pisa), and merchant guild privileges persisted (Ogilvie (2011), pp. 33-34).	R1

$LUEBECK$	In the sixteenth century, Lübeck, to tackle increasing Dutch and English competition, tightened the monopoly of its local merchant guild (Ogilvie (2011), p. 87).	R1
$LYON$	Weaving, finishing and marketing in silk industry of Lyon was controlled until 1789 (Ogilvie (2011), p. 415). French cities resisted any guild reform until at least the eighteenth century. Rouen, Nancy, Metz, Roussillon and Paris had some partial reform during the eighteenth century (Horn (2015), p. 224).	R1
$MADRID$	Merchant guilds enjoyed privileges in the seventeenth century in Madrid (Herr (2015), p. 168). Epstein (2004) called it a "parasitic city".	R1
$MAGDEBURG$	Magdeburg law was an elaborate system of city privileges that emerged in Magdeburg, and spread across Germany and Eastern Europe. No evidence of Magdeburg laws changing in the sixteenth century.	R3
$MALAGA$	In 1667, Malaga sought permission to trade with the Atlantic (Smith (1940), p. 92). Andalusian merchant guilds used their privileges to guard their interests in the eighteenth century (Stein & Stein (2014), pp. 95-96).	R1
$MANTUA$	In the sixteenth century Mantua increased its centralization and control over countryside (Epstein (2004), p. 311).	R1
$MARSEILLE$	No specific information on the city. French cities resisted any guild reform until at least the eighteenth century. Rouen, Nancy, Metz, Roussillon and Paris had some partial reform during the eighteenth century (Horn (2015), p. 224).	R3
$MECHELEN$	Antwerp emerged as a major trading market for traders of the region Ogilvie (2011). No evidence exists such that individual merchant guilds successfully blocked locals from doing trade through Antwerp. The closeness of Belgian cities hindered them from enforcing exclusive privilege (Ogilvie (2011), pp. 31-32).	U2
$MESSINA$	Primarily an agrarian society, Sicilian cities began to grant guilds late, and their significance only increased in early modern period (Epstein 2003). No specific mention of the city.	R3
$METZ$	Rouen, Nancy, Metz, Roussillon and Paris had some partial guild reform but only in 1762 (Horn (2015), p. 224).	R1

$MILANO$	In 1593 Milanese Lombardy, representatives of the village areas managed to have the privileged allocation of raw silk suspended, even though cities maintained their monopoly over most of the activities. (Epstein (2004), p. 308).	U1
$MURCIA$	Privileged guilds of Madrid set factories in the allied cities of Murcia (Herr (2015), p. 168).	R1
$NAPOLI$	Naples resembled a feudal state with significant privileges to nobles and elites (see Chapter 14 on *Kingdom of Naples*, Epstein (2004)).	R1
$NUERNBERG$	Nuremberg was called a "city without guilds" because the city councils dominated by powerful international dealers and merchant-entrepreneurs, regulated activities of craft guilds (Soly 2008).	R1
$ORLEANS$	Orleans received new privileged guilds in the seventeenth century (Horn (2015), p. 224).	R1
$PADOVA$	In Venetian republic while Padova continued to maintain its position in wool production (Epstein (2004), p. 315), but during the second half of the sixteenth century rural representatives were able to negotiate concessions for tax allocation and collection (Epstein (2004), p. 301).	U1
$PALERMO$	Primarily an agrarian society, Sicilian cities began to grant guilds late, and their significance only increased in early modern period (Epstein 2003). No specific mention of the city.	R3
$PARIS$	Paris expanded its guild system in the seventeenth century, and partial reforms only happened in 1762 (Horn (2015), p. 224).	R1
$PIACENZA$	In the sixteenth century, Farnese duchy increased its centralization and control over countryside which meant large part of the state territory lost to urban jurisdiction (Epstein (2004), p. 312).	R1
$PISA$	In the sixteenth century, Florence had a high level of political centralism, which led to hollowing out of region's urban network (Epstein (2004), pp. 295-296) (Lucca and Pisa), and merchant guild privileges persisted (Ogilvie (2011), pp. 33-34).	R1
$PLASENCIA$	No specific information on the city. Spanish cities began receiving merchant guild permissions only in late medieval period (Smith 1940) and cities without such charters lost out.	R3

$PLOVDIV$	Bribes and guild concessions were common in Bulgaria, (Lampe & Jackson (1982), p. 145) until 1878 emerged as an independent state.	R3
$POZNAN$	Merchant guilds enjoyed monopoly in the seventeenth century Poznan, which they defended aggressively against local Jews and foreigners (Ogilvie (2011), p. 82).	R1
$PRAHA$	James (1893) records a 1856 celebration of merchant guilds of Prague.	R1
$RENNES$	Rennes received new privileged guilds in the seventeenth century (Horn (2015), p. 224).	R1
$ROMA$	Early modern Rome had merchant guilds which charged admission fees that added to the wealth requirement of joining the guilds (Ogilvie (2011), p. 60).	R1
$ROUEN$	Rouen, Nancy, Metz, Roussillon and Paris had some partial guild reform but only in 1762 (Horn (2015), p. 224).	R1
$SALONIKA$	No specific information on nature of institutions in Salonika, but given that ruling Ottoman empire had a system of granting privileges (Ogilvie 2011), it is assumed economic organization in Ottoman influenced Salonika was similar.	R3
$SARAJEVO$	Sarajevo had merchant guilds during the sixteenth-seventeenth century. Merchant colonies of Bosnian merchants blocked the settlement of Ragusan merchant colonies in Sarajevo and other Hungarian and Slavonian towns (Ogilvie (2011), p. 124).	R1
$SEVILLA$	In the sixteenth century, Seville restricted naturalization, and thus permission to trade as a long-distance merchant, to men who were subjects of the Holy Roman empire or one of its allies, or those who could prove twenty years residence in Castile and fulfil a property requirement (Ogilvie (2011), p. 54).	R1
$SKOPJE$	Skopje (Molnár 2007) and other Ottoman Balkan cities have been documented to have merchant colonies in the sixteenth century that held commercial privileges (Ogilvie (2011), p. 10).	R1
$TIRGOVISTE$	Targoviste was an important town of Wallachia which in the sixteenth century enjoyed commercial privileges even when some new settlements in Wallachia did not enjoy the same generous privileges as older ones (Rădvan 2009).	R1

$TOLEDO$	In Toledo, during the sixteenth century, craft guilds became even more powerful and autonomous extending control over previously unregulated manufacture. No information specific to merchant guilds (Epstein (2004), p. 288).	R1
$TOULOUSE$	In Languedoc (Toulouse) marketing and finishing of proto-industrial woollens were monopolized by the merchant guilds of Clermont-de-Lodeve and other towns into the later eighteenth century (Ogilvie (2006), p. 415).	R1
$TOURNAI$	Antwerp emerged as a major trading market for traders of the region Ogilvie (2011). No evidence exists such that individual merchant guilds successfully blocked locals from doing trade through Antwerp. The closeness of Belgian cities hindered them from enforcing exclusive privilege (Ogilvie (2011), p. 31-32).	U2
$TOURS$	Tours received new privileged guilds in the seventeenth century (Horn (2015), p. 224).	R1
$VALENCIA$	In Valencia, merchant guilds were existent in the eighteenth century, as a guild court was established in 1762 (Smith (1940), p. 13).	R1
$VALENCIENNES$	In Valenciennes and Cambresis cities retained control of the linen proto-industry until the late eighteenth century (Ogilvie (2006), p. 414).	R1
$VALLADOLID$	From the late fifteenth century, Burgos received exclusive rights to trade at the expense of cities like Valladolid (Smith (1940), p. 69).	R1
$VENEZIA$	In Venice while non-citizens (a large majority) (Ogilvie (2011), p. 53) were excluded from trade in the sixteenth century; during the second half of the sixteenth century rural representatives were able to negotiate concessions for tax allocation and collection (Epstein (2004), p. 301).	U1
$VERONA$	No specific mention of merchant guilds of Verona, but during the second half of the sixteenth century rural representatives of Venetian republic were able to negotiate concessions for tax allocation and collection (Epstein (2004), p. 301).	U3
$WIEN$	It was in the eighteenth century that *Congress of Vienna* decided to remove city staple rights privileges (ART XXV).	R1
$WROCLAW$	In early modern period, Silesia (Wroclaw) had merchant guilds/privileged merchant companies in proto-industrial sectors (Ogilvie (2011), pp. 33-34).	R1

E.3 Urban Agglomeration

Table E.5: City Clusters based on Hierarchal Clustering based on average distance (UPGMA)

City	Cluster Name	Number of Cities
ALMERIA	East Andalusia	9
AMSTERDAM	Netherlands	12
ANGERS	Western France	6
ANTWERPEN	Belgium	22
AQUILA	Roman	5
ATHINAI		1
AUGSBURG		4
AVIGNON	Southern France	7
BARCELONA		1
BELGRADE		1
BOLOGNA	Florence	10
BORDEAUX		2
BOURGES	Central France	3
BRESCIA	Milan	14
BRUGGE	Belgium	22
BRUXELLES	Belgium	22
BURGOS		1
CAEN	Northern France	5
CORDOBA	West Andalusia	18
CREMONA	Milan	14
ERFURT	Western Germany	8
FERRARA	Florence	10
FIRENZE	Florence	10
GDANSK		3
GENOVA		10
GENT	Belgium	22
GRANADA	East Andalusia	9
HAARLEM	Netherlands	12
HAMBURG	Northern Germany	4
JEREZ-DE-LA-FRONTERA		2
KOBENHAVN		1
KOELN	Eastern Germany	4
LEIDEN	Netherlands	12
LIEGE	Eastern Germany	4
LILLE	Belgium	22

LISBOA		3
LONDON		2
LUCCA	Florence	10
LUEBECK	Northern Germany	4
LYON		2
MADRID	Spanish	10
MAGDEBURG	Western Germany	8
MALAGA	West Andalusia	18
MANTUA	Milan	14
MARSEILLE	Southern France	7
MECHELEN	Belgium	22
MESSINA		15
METZ		4
MILANO	Milan	14
MURCIA		2
NAPOLI		3
NUERNBERG		3
ORLEANS	Central France	3
PADOVA	Venice	6
PALERMO		3
PARIS		5
PIACENZA	Milan	14
PISA	Florence	10
PLASENCIA		3
PLOVDIV		3
POZNAN	Poland	4
PRAHA		4
RENNES		2
ROMA	Roman	5
ROUEN	Northern France	5
SALONIKA		1
SARAJEVO		1
SEVILLA	West Andalusia	18
SKOPJE		3
TIRGOVISTE		2
TOLEDO		5
TOULOUSE		3
TOURNAI	Belgium	22
TOURS	Western France	6
VALENCIA		4

VALENCIENNES	Belgium	22
VALLADOLID	Spanish	10
VENEZIA	Venice	6
VERONA	Venice	6
WIEN		3
WROCLAW	Poland	4

E.4 Appendix Figures

E.5 Robustness Tests: Tables

Appendix Tables E.6 (multinomial regression), E.7 (considering number of printing cities within 50km from a city ($PrintCity50km_a$) as a proxy for printing), E.8 (considering different thresholds for distance from port and level of clustering), E.9 and E.10 (considering a limited sample of 50 largest cities in the fifteenth century), E.11 and E.12 (considering a limited sample of cities without code R3 or U3) report robustness checks from section 3.4.7

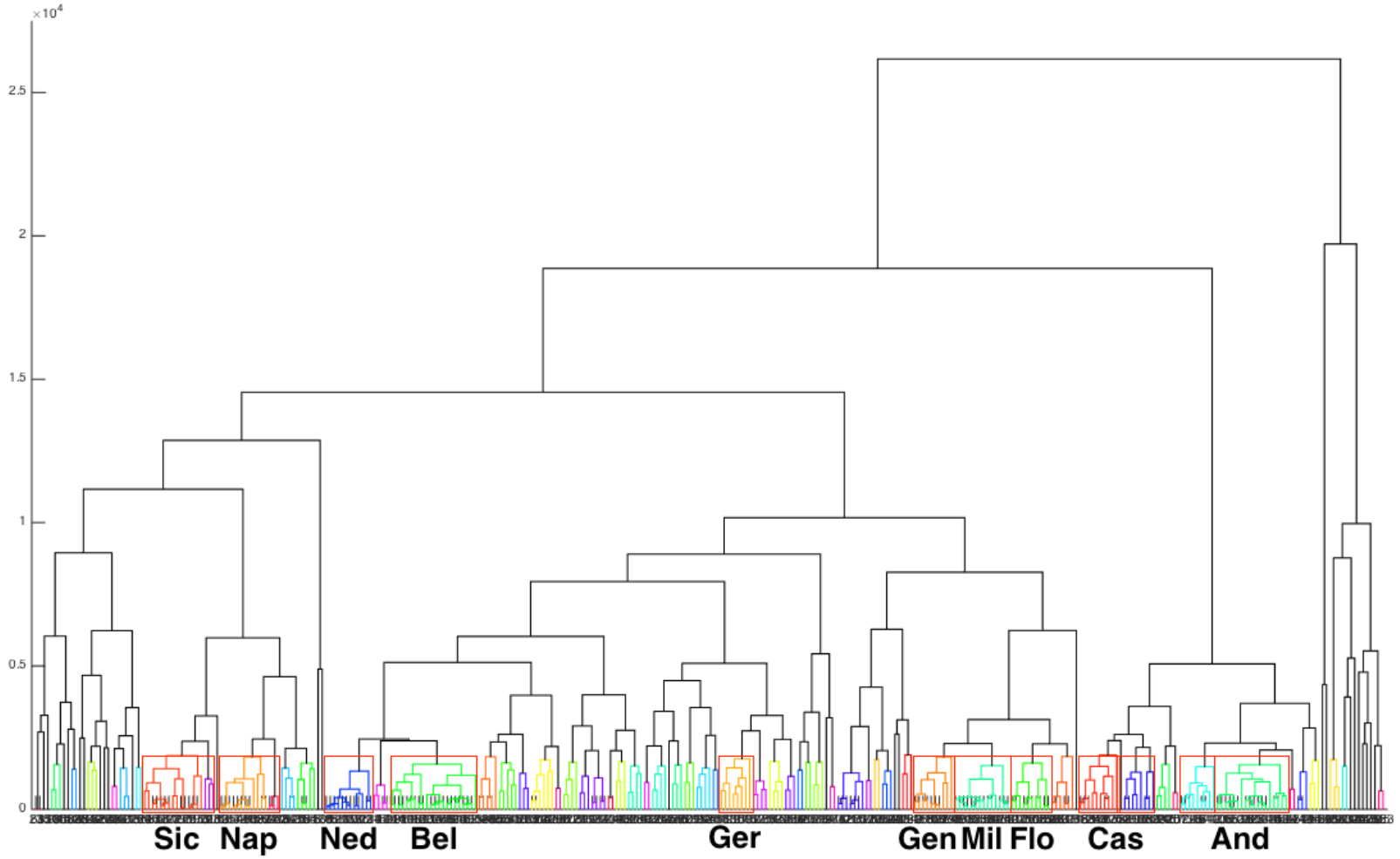

Figure E.1: The figure presents a hierarchal dendrogram of 335 largest cities in fourteenth, fifteenth and sixteenth century Europe.

Table E.6: Columns 2a and 2b report the log odds coefficients for multinomial logit regression between the 16th century economic institutions and the 15th century diffusion of printing with additional variables. Column 1 reports the OLS regression for reference. Column 4a and 4b report the log odds coefficients for multinomial logit regression for cities within 150 km. Column 4 reports the OLS regression for the limited sample for reference. Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2. Printing Penetration is a measure of level of diffusion of printed books in the 15th century in a given city, from the 121 printing cities in the GW database (see section 3.4.4). Closeness Port is the square root of $constant$ - Distance of the city from a sea port (see section 3.4.3). Additional controls are described in section 3.4.6.

	(1)	(2a)	(2b)	(3)	(4a)	(4b)
VARIABLES	OLS	Mlogit	Mlogit	OLS	Mlogit	Mlogit
Dependent Variable:		$Inst_a^{1600}=1$	$Inst_a^{1600}=2$		$Inst_a^{1600}=1$	$Inst_a^{1600}=2$
				$<150km$	$<150km$	$<150km$
Print Penetration	-0.0617	-1.558	-7.180	0.123***	4.657**	-1.411
	(0.0427)	(1.054)	(5.026)	(0.0417)	(1.816)	(2.302)
Closeness Port	-0.0119	-0.792**	-0.620			
	(0.0123)	(0.320)	(0.707)			
Print Penetration X Closeness Port	0.0151***	0.255**	0.426*			
	(0.00469)	(0.105)	(0.253)			
log(# cities in cluster)	0.161*	2.846**	0.554	0.0731	3.172**	0.832
	(0.0825)	(1.210)	(0.868)	(0.0825)	(1.511)	(1.469)
Growth 1500	0.0651	-0.669	-0.481	-0.0873	-2.595	-1.046
	(0.132)	(0.834)	(0.951)	(0.165)	(1.873)	(1.988)
log(Population 1500)	-0.0552	2.569**	-2.903**	0.110	4.883**	-2.161
	(0.0838)	(1.248)	(1.427)	(0.110)	(2.332)	(3.849)
$\sqrt{Elevation}$	0.00140	-0.0614	-0.851*	-0.00267	0.172	0.711
	(0.00898)	(0.118)	(0.444)	(0.00845)	(0.159)	(1.493)
Medieval Fair	-0.0975	-4.403***	4.088**	-0.289**	-12.85***	-4.313
	(0.139)	(1.585)	(1.862)	(0.139)	(4.468)	(3.761)
Near Atlantic				-0.235	0.850	20.71
				(0.318)	(10.62)	(27.36)
Near Atlantic X Print Penetration				0.318**	2.325	6.065
				(0.131)	(2.908)	(4.142)
Constant	0.0127	-10.65	16.39	-0.531	-40.25**	-28.81
	(0.352)	(7.366)	(14.71)	(0.343)	(16.43)	(33.26)
Observations	74	74	74	56	56	56
R-squared	0.360			0.503		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table E.7: The table replicates Table 3.6 with an alternate the 15th century printing variable: Number of the 15th century printing cities within 50 km distance from the observed city. Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS
# Print Cities in 50km	0.223***		0.205***		0.167***		0.152***	
	(0.0386)		(0.0484)		(0.0515)		(0.0463)	
1 Print City in 50km		-0.0649		-0.0914		-0.119		0.0528
		(0.127)		(0.153)		(0.171)		(0.150)
2 Print Cities in 50km		0.364		0.293		0.356		0.163
		(0.348)		(0.361)		(0.299)		(0.217)
3 Print Cities in 50km		0.489*		0.447		0.474*		0.322
		(0.282)		(0.282)		(0.249)		(0.240)
4 Print Cities in 50km		1.114***		1.093***		0.897***		0.862***
		(0.253)		(0.258)		(0.255)		(0.129)
5 Print Cities in 50km		1.114***		1.085***		0.769**		0.796***
		(0.253)		(0.305)		(0.334)		(0.160)
6 Print Cities in 50km		0.864***		0.869***		0.410		0.434
		(0.104)		(0.252)		(0.336)		(0.372)
log(# cities in cluster)			0.0183	0.00226	0.0406	0.0312	0.0366	0.0606
			(0.0969)	(0.0953)	(0.0869)	(0.0880)	(0.0885)	(0.0743)
Growth 1500			0.0512	-0.0200	-0.00243	-0.0252	-0.0187	0.0687
			(0.187)	(0.189)	(0.163)	(0.148)	(0.165)	(0.142)
log(Population 1500)			-0.0386	0.0135	0.0517	0.0681	0.0653	0.0488
			(0.102)	(0.0962)	(0.108)	(0.0945)	(0.114)	(0.0972)
$\sqrt{Elevation}$			-0.00858	-0.00585	-0.000316	0.00343	-0.000737	0.00360
			(0.0117)	(0.0114)	(0.00816)	(0.00847)	(0.00812)	(0.00932)
Near Atlantic					0.540**	0.582**	0.471	0.412
					(0.255)	(0.272)	(0.373)	(0.285)
# Prt. Cities in 50km X Near Atl.							0.0374	
							(0.108)	
0 Prt. City in 50km X Near Atl.								0.124
								(0.566)
1 Prt. Cities in 50km X Near Atl.								-0.492
								(0.316)
2 Prt. Cities in 50km X Near Atl.								1.511***
								(0.354)
3 Prt. Cities in 50km X Near Atl.								1.323***
								(0.398)
4 Prt. Cities in 50km X Near Atl.								0.0484
								(0.572)
5 Prt. Cities in 50km X Near Atl.								0
								(0)
6 Prt. Cities in 50km X Near Atl.								0
								(0)
Medieval Fair			0.0682	0.106	-0.0868	-0.0778	-0.0865	-0.153
			(0.168)	(0.167)	(0.161)	(0.168)	(0.163)	(0.161)
Constant	0.0421	0.136	0.204	0.117	-0.225	-0.234	-0.242	-0.189
	(0.0815)	(0.104)	(0.391)	(0.388)	(0.350)	(0.332)	(0.363)	(0.311)
Observations	59	59	56	56	56	56	56	56
R-squared	0.345	0.401	0.347	0.400	0.449	0.506	0.451	0.655

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table E.8: The table replicates Column 4 of 3.6 considering a limited sample of cities within x km of sea port. Columns 2 and 3 (in contrast to Column 1) have alternate measurement of city clustering (see section 3.4.6). Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2.

VARIABLES	(1) OLS $x = 150$	(2) OLS $x = 150$	(3) OLS $x = 150$	(4) OLS $x = 50$	(5) OLS $x = 75$	(6) OLS $x = 100$	(7) OLS $x = 125$	(8) OLS $x = 175$	(9) OLS $x = 200$
Print Penetration	0.123***	0.121***	0.111**	0.0717	0.140*	0.124**	0.119***	0.122***	0.0898**
	(0.0417)	(0.0407)	(0.0430)	(0.0737)	(0.0780)	(0.0552)	(0.0423)	(0.0419)	(0.0382)
log(# cities in cluster)	0.0731			-0.0613	0.0215	0.0586	0.0654	0.0724	0.0780
	(0.0825)			(0.171)	(0.143)	(0.106)	(0.0828)	(0.0817)	(0.0775)
Growth 1500	-0.0873	-0.0847	-0.103	-0.181	-0.144	-0.164	-0.136	-0.0847	-0.0867
	(0.165)	(0.163)	(0.166)	(0.292)	(0.231)	(0.213)	(0.198)	(0.159)	(0.152)
log(Population 1500)	0.110	0.120	0.121	0.0876	0.0449	0.101	0.110	0.112	0.112
	(0.110)	(0.104)	(0.108)	(0.188)	(0.145)	(0.138)	(0.111)	(0.110)	(0.106)
$\sqrt{Elevation}$	-0.00267	-0.00188	-0.00542	-0.0552	-0.00732	-0.00888	-0.00676	-0.00236	-0.00233
	(0.00845)	(0.00854)	(0.00940)	(0.0834)	(0.0448)	(0.00979)	(0.00916)	(0.00781)	(0.00717)
Near Atlantic	-0.235	-0.166	-0.332	-0.308	-0.249	-0.264	-0.270	-0.238	-0.309
	(0.318)	(0.309)	(0.271)	(0.426)	(0.344)	(0.331)	(0.318)	(0.317)	(0.312)
Near Atlantic X Print Penetration	0.318**	0.292**	0.341***	0.404**	0.334**	0.341**	0.336**	0.320**	0.348***
	(0.131)	(0.134)	(0.119)	(0.164)	(0.152)	(0.143)	(0.133)	(0.131)	(0.128)
Medieval Fair	-0.289**	-0.279**	-0.265*	-0.115	-0.177	-0.306*	-0.276*	-0.290**	-0.239*
	(0.139)	(0.135)	(0.135)	(0.387)	(0.203)	(0.179)	(0.148)	(0.138)	(0.123)
log(# cities in small cluster)		0.127							
		(0.0897)							
log(# cities in large cluster)			0.0960						
			(0.0818)						
Constant	-0.531	-0.566*	-0.637*	-0.0911	-0.323	-0.464	-0.504	-0.531	-0.490
	(0.343)	(0.323)	(0.360)	(0.463)	(0.473)	(0.414)	(0.353)	(0.343)	(0.337)
Observations	56	56	56	29	35	44	53	57	61
R-squared	0.503	0.516	0.507	0.589	0.558	0.547	0.503	0.506	0.497

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

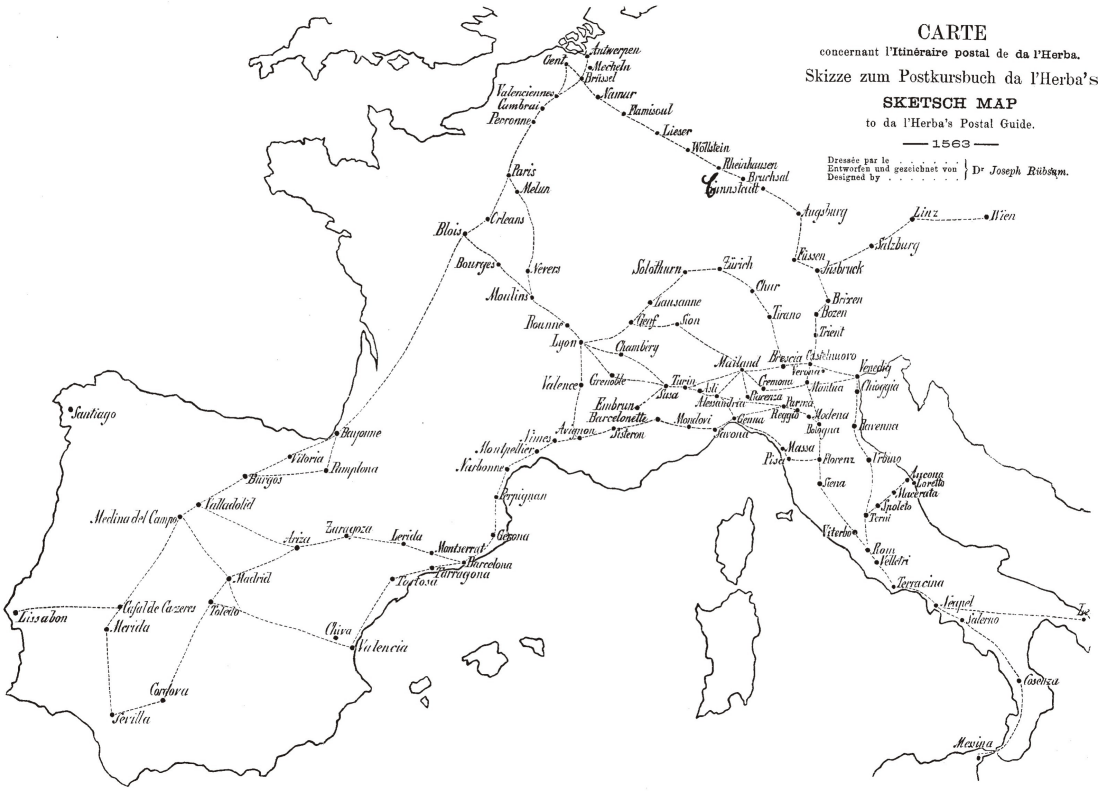

Figure E.2: European postal routes in 1563 according to the travelogue of Giovanni da L'Herba. Sketch in: Joseph Rbsam (1854-1927)'s L'Union postale, revue de l'Union Postale Universelle, Bern 1900, image source: Wikimedia.

Table E.9: The table replicates Table 3.4 considering a limited sample of only the largest cities of fifteenth century Europe (see Table E.2). Dependent variable is 16th century nature of economic institutions in a given city as described in section 3.4.2.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Print Penetration	0.123***	0.137***	-0.0785*	-0.0775*	0.111**
	(0.0443)	(0.0416)	(0.0466)	(0.0446)	(0.0468)
Closeness Port		0.0256**	-0.0289**	-0.0264**	0.0212
		(0.0110)	(0.0109)	(0.0123)	(0.0189)
Print Penetration X Closeness Port			0.0181***	0.0171***	
			(0.00540)	(0.00593)	
log(# cities in cluster)				0.212*	0.226*
				(0.122)	(0.132)
Growth 1500				-0.000366	-0.00215
				(0.191)	(0.210)
log(Population 1500)				-0.00634	-0.0168
				(0.148)	(0.153)
$\sqrt{Elevation}$				0.00567	0.00126
				(0.0141)	(0.0150)
Medieval Fair				-0.0689	0.00444
				(0.159)	(0.167)
Constant	-0.0773	-0.444**	0.224*	-0.147	-0.666
	(0.0891)	(0.179)	(0.112)	(0.613)	(0.605)
Observations	57	57	57	53	53
R-squared	0.101	0.185	0.280	0.352	0.274

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

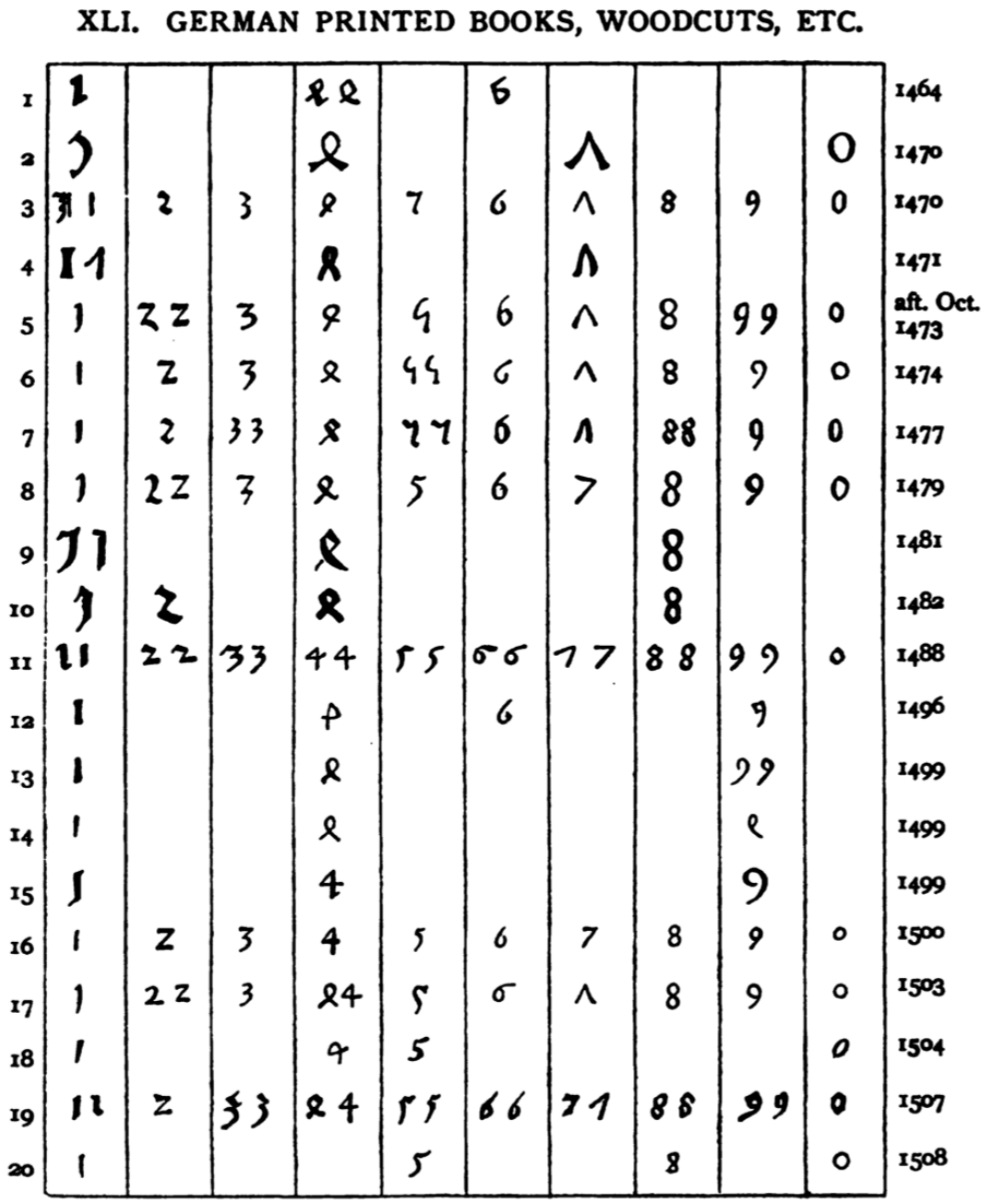

Figure E.3: An illustration of different notations used for Hindu-Arabic numbers in printed books and woodblocks in Germany, from Hill (1915) (p. 92).

Table E.10: The table replicates Table 3.6 considering a limited sample of only the largest cities of the fifteenth century Europe (see Table E.2) within 150km of sea port. Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2.

	(1)	(2)	(3)	(4)
VARIABLES	OLS	OLS	OLS	OLS
Print Penetration	0.198***	0.178**	0.210***	0.149**
	(0.0579)	(0.0724)	(0.0639)	(0.0542)
log(# cities in cluster)		0.197	0.230	0.126
		(0.206)	(0.174)	(0.174)
Growth 1500		0.00338	0.0279	-0.123
		(0.323)	(0.244)	(0.231)
log(Population 1500)		-0.0435	0.0831	0.144
		(0.208)	(0.185)	(0.179)
$\sqrt{Elevation}$		-0.00760	0.00710	0.00467
		(0.0156)	(0.0137)	(0.0156)
Near Atlantic			0.685**	-0.742
			(0.268)	(0.540)
Near Atlantic X Print Penetration				0.471**
				(0.224)
Medieval Fair		-0.0401	-0.334**	-0.338**
		(0.213)	(0.161)	(0.149)
Constant	-0.177	-0.299	-1.076	-0.857
	(0.115)	(0.943)	(0.804)	(0.810)
Observations	41	39	39	39
R-squared	0.188	0.235	0.399	0.493

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

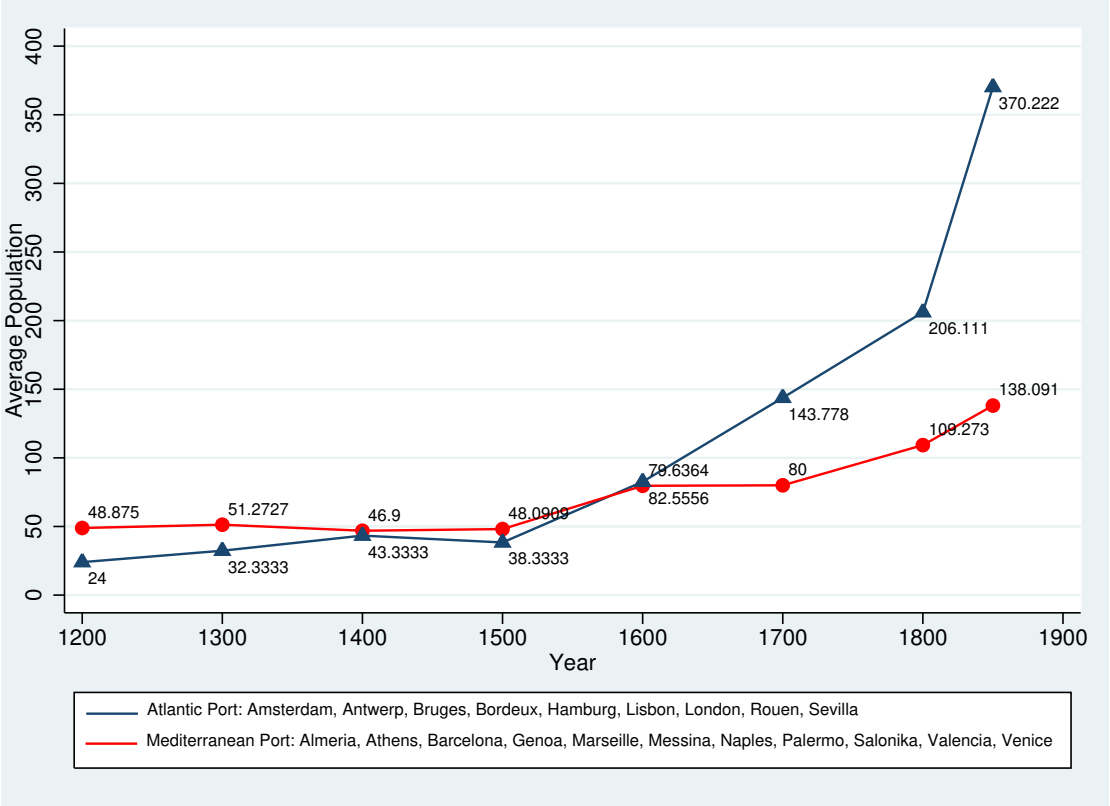

Figure E.4: Average population of the Atlantic and Mediterranean port cities in the sample across time (1200-1850). The average population of the sampled cities do not represent the general population trends of all the Atlantic and Mediterranean port cities.

Table E.11: The table replicates Table 3.4 considering a limited sample excluding those cities that are coded as R3 or U3 (see Table 3.2). Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Print Penetration	0.123**	0.136***	-0.0770	-0.0747	0.122**
	(0.0486)	(0.0431)	(0.0501)	(0.0498)	(0.0483)
Closeness Port		0.0378***	-0.0228	-0.0233	0.0334**
		(0.0124)	(0.0163)	(0.0163)	(0.0149)
Print Penetration X Closeness Port			0.0186***	0.0176***	
			(0.00587)	(0.00523)	
log(# cities in cluster)				0.192*	0.201*
				(0.100)	(0.104)
Growth 1500				0.0864	0.100
				(0.151)	(0.164)
log(Population 1500)				-0.0900	-0.0946
				(0.0994)	(0.112)
$\sqrt{Elevation}$				0.00124	0.000284
				(0.0122)	(0.0130)
Medieval Fair				-0.0293	0.0298
				(0.178)	(0.181)
Constant	-0.0629	-0.580***	0.127	0.136	-0.526
	(0.120)	(0.193)	(0.144)	(0.512)	(0.526)
Observations	59	59	59	55	55
R-squared	0.067	0.219	0.287	0.348	0.290

*** p<0.01, ** p<0.05, * p<0.1

Table E.12: The table replicates Table 3.6 considering a limited sample excluding those cities within 150km of sea port, that are coded as R3 or U3 (see Table 3.2). Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Print Penetration	0.210***	0.238***	0.257***	0.161**
	(0.0690)	(0.0700)	(0.0670)	(0.0708)
log(# cities in cluster)		0.153	0.120	0.0740
		(0.130)	(0.107)	(0.105)
Growth 1500		0.104	0.0537	-0.0852
		(0.245)	(0.176)	(0.185)
log(Population 1500)		-0.202	-0.0670	0.0723
		(0.136)	(0.116)	(0.128)
$\sqrt{Elevation}$		-0.0266	-0.0107	-0.00741
		(0.0183)	(0.0127)	(0.0127)
Near Atlantic			0.862***	-0.557
			(0.277)	(0.533)
Near Atlantic X Print Penetration				0.460**
				(0.201)
Medieval Fair		0.0157	-0.287	-0.326*
		(0.242)	(0.207)	(0.184)
Constant	-0.205	0.257	-0.417	-0.515
	(0.175)	(0.539)	(0.445)	(0.437)
Observations	42	40	40	40
R-squared	0.129	0.236	0.474	0.529

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table E.13: The table replicates Table 3.6 with an alternate the 15th century printing variable: Print^{U9} Penetration. Print^{U9} Penetration is a measure of level of diffusion of printed books in the 15th century in a given city, from the 121 printing cities in the 9 trade friendly categories of the USTC catalog (see section 3.4.7). Dependent variable is the 16th century nature of economic institutions in a given city as described in section 3.4.2.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Print^{U9} Penetration	0.277***	0.254***	0.232***	0.152***
	(0.0476)	(0.0508)	(0.0480)	(0.0424)
log(# cities in cluster)		0.0967	0.0935	0.0629
		(0.0978)	(0.0822)	(0.0747)
Growth 1500		0.0427	-0.0281	-0.123
		(0.194)	(0.159)	(0.145)
log(Population 1500)		-0.0509	0.0603	0.139
		(0.109)	(0.109)	(0.109)
$\sqrt{Elevation}$		-0.0101	7.66e-05	-0.00136
		(0.0115)	(0.00877)	(0.00818)
Near Atlantic			0.610***	-0.139
			(0.201)	(0.257)
Near Atlantic X Print^{U9} Penetration				0.267**
				(0.112)
Medieval Fair		-0.112	-0.266*	-0.259*
		(0.155)	(0.141)	(0.131)
Constant	-0.317***	-0.160	-0.648*	-0.662*
	(0.0883)	(0.369)	(0.363)	(0.339)
Observations	59	56	56	56
R-squared	0.351	0.377	0.513	0.568

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix F

Data on Indian castes and censuses

F.1 Risley's District Committees

Table F.1a: British Provinces where district committees were formed

Census Region	Note	District Committee
Bengal	Bengal was the British province where 1901 census commissioner Herbert Risley, conducted his earlier ethnographic studies and forwarded his social precedence scheme used on Bengal to other provinces as a template. To come up with a social precedence census officer E.A. Gait employed "committees", and arbitrated several points of disputed precedence, although he did not undertake a social precedence classification for the tribal regions of the province. The provincial census was printed in Calcutta.	1
Northwestern Provinces and Oudh	In the adjacent British province of NWP and Oudh, census officer R. Burn writes "District Officers were then asked to appoint representative committees" to determine social precedence. Census officer notes that castes were given an order based on Mr. Risley's scheme and based on the preexisting caste studies of orientalist William Crooke in the province, who wrote the book *The Tribes and Tribes of the Northwestern Provinces and Oudh in 1896* (as a followup book to Mr. Risley's book on Bengal). The provincial census was printed in Allahabad.	1

Central Provinces	In the adjacent British province of Central Provinces, census report had a detailed caste precedence table formed in correspondence with district officials. Census officer R. V. Russell wrote, "the materials, so far as the castes of the Central Provinces are concerned, have been obtained from the replies to a short set of questions, selected from those on which the Ethnographic Survey of Bengal was conducted, and which were circulated to District Officers last year." The provincial census was printed in Nagpur.	1
Berar	In the British province of Berar, adjacent to Central Provinces, a detailed social precedence table was formed and "The Deputy Commissioners were addressed, and with their help committees were formed to advise on the arrangements of various castes found in their districts in the order of their social precedence.", headed by census officer Ardaseer Dinshawji Chinoy. The provincial census was printed in Allahabad.	1
Madras	In the southern British province of Madras, which bordered Bengal, detailed social precedence schemes specific to different lingual groups was developed. Census officer wrote "I drew up the table in accordance with the information available, neither extenuating anything nor setting down aught in malice, and committees of native gentlemen formed for the purpose in every district have criticised and amended the original draft, but I am well aware that the task has been a most delicate one". The provincial census was printed in Madras.	1
Gwalior	Princely state of Gwalior, enclosed within NWP and Central Provinces, developed a detailed social precedence table. Census officer J. W. D. Johnstone wrote, "Subsidiary Table I is an attempt to classify the castes of this State according to social precedence, but though the groups have been formed after consultation with various local authorities, it is not claimed that the arrangement precludes all dispute.". The table was developed with consultation with census officer of Central India Agency. The provincial census was printed in Lucknow.	1

Rajputana Agency	In the princely Rajputana Agency, a detailed social precedence table was developed based on the consultation with different Durbars (princely states) after multiple stages of revision, headed by census officer Captain A. D. Bannerman. The provincial census was printed in Lucknow.	1
Ajmer Merwara	In the British province of Ajmer Merwara, enclosed within Rajputana Agency, a detailed caste precedence table was developed assisted by "a Committee of Native gentlemen", headed by census officer K. C. Bramley. The provincial census was printed in Ajmer.	1
Baroda	In the Princely state of Baroda, enclosed within the Bombay Presidency, an "Order of social precedence was framed from the replies received through the Ethnographic Committees appointed in that behalf", headed by Jamshedji Ardeshir Dalal. The provincial census was printed in Bombay.	1

Table F.1b: British Provinces where district committees were not formed to form detailed caste precedence tables

Census Region	Note	District Committee
Assam	In the British province of Assam, adjacent to Bengal, a detailed social precedence table has been provided. But, there seems mixed evidence on the degree of district-level consultation. The census officer B. C. Allen writes "Within each table I have arranged the castes in the rank which as far as I can ascertain, is assigned to them by popular opinion. This is a point upon which I desire to lay some stress. I have made no attempt to go behind general opinion, or to consider the grounds upon which they are based." On the other hand, the officer did consult local officers, and refers to "Kamrup committee" on whose recommendation it altered the social precedence of a caste "Saloi". Also the census officer received memorials from caste groups, requesting to be levelled upwards in the caste precedence table, pointing towards a satisfactorily detailed consultation. The provincial census was printed in Guwahati.	0 or 1

Central India Agency	In the princely Agency of Central India Agency adjacent to Central Provinces, a detailed caste precedence table was adopted, by following the classification "devised by the Census Superintendent of the North-West Provinces". The census officer, Captain C. Eckford Luard, wrote "it has been almost impossible to reconcile the different positions given to the castes and tribes by different states. All I could do was to arrange them as near the average position as possible noting where there were any great variations in the place assigned." Thus, while a classification scheme existed, it lacked the detailed revision carried out in the Rajputana Agency, or district-based assessment of Northwestern Provinces and Oudh. The provincial census was printed in Lucknow.	0 or 1
Bombay	The census officer of British province of Bombay, R.E. Enthoven showed inhibition in classification of caste and wrote, "In 1881 an attempt to group the castes by social precedence led to widespread discontent and to numerous representations of an embittered character. It is undesirable to provoke for a second time the hostile feeling then aroused in connection with this difficult question of social precedence between caste and caste, nor can the Provincial Superintendent claim the knowledge, even it he assumes the powers, of former native rulers in dealing with such matters." Thus "to avoid friction and discontent" "*a scheme of classification on very broad lines" was adopted* and "instructions were issued to committees to act on this principle". The provincial census was printed in Bombay.	0
Hyderabad	The census officer of the Southern princely state of Hyderabad, Mirza Mehdy Khan, provides a social precedence table, but there is no mention of any consultation with committees to come up with this social precedence, arguably because the census officer is not a British officer. Moreover, castes classification in the province has little resemblance with Risley's detailed classification scheme where castes were put into different classes, and in Hyderabad the many castes were first put in super-caste groups like "Ausala" (artisans) and ranked like a long list of broadly occupational groups. The provincial census was printed in Hyderabad.	0

Kashmir	The census officer of the Northern princely state of Kashmir, Khan Bahadur Munshi Gholam Ahmed Khan, provided a vague social precedence table, but castes mentioned in this table include occupational groups like "Tailors", "Goldsmiths", and "Carpenters", who are classed as "Shudras", preceded by Brahmins, Kshatriyas or Rajputs and "Vaishyas". While the census officer says, "the castes and sub-castes, therefore, recorded in the following pages of this chapter are mainly based upon the materials supplied to me by the reports of the District Officers", but, there seems to be no detail on how the social precedence table was developed, and if proper consultation like in Bengal or United Provinces was conducted. The officer says the reports were "not quite up to mark and lacked in many points, which could not, but, be cleared owing to the shortness of time at my disposal." The provincial census was printed in Lahore.	0
Cochin	The census officer of the Southern princely state of Cochin, M. Sankara Mennon, provides a detailed census precedence table, but there is no record of consultation with district committees, and the census table seems to be developed by the census officer himself, although he mentions the consultation with a priest to ascertain the position of a Brahmin subcaste in the table. It seems, Mr. Mennon not being a British census officer undertook the classification himself. The provincial census was printed in Ernakulam.	0

Table F.1c: British Provinces where no caste precedence tables were developed

Census Region	Note	District Committee
Mysore	The census officer in the Southern princely state of Mysore, T. Ananda Row,, wrote "No attempt is made to group or arrange the Castes according to any scheme of social precedence". The officer points towards the complication of the caste system and the general desire to not generate "the extensive dissatisfaction, which such a classification would even otherwise produce" as some of the reasons to not not apply the classification. The provincial census was printed in Bangalore.	0

Travancore	The census officer in the Southern princely state of Travancore, N. Subramanya Aiyar, wrote "Viewed... from the standpoint of either scientific accuracy or public policy, the formulation of a scale of well-recognized social precedence for the Travancore castes has to be given up for the present." He pointed at the attempts to gain feedback from 111 "competent" persons regarding social precedence, out of which only 6 felt prepared to offer views on precedence. The provincial census was printed in Trivandrum.	0
Punjab	In the Northern British province of Punjab, census officer H.A. Rose wrote "The attempt to classify castes in the order of their social precedence was, as will be seen from the foregoing pages, doomed to failure from the outset. We know far too little of the complex organization of the Punjab castes to be able, at present to classify them in any systematic or instructive way. The complications within the castes have their natural counterpart in the chaotic and uncertain relations between the different castes." The provincial census was printed in Simla.	0
Coorg	In the small British province of Coorg adjacent to Madras, census officer W. Francis provides no social precedence table, and mentions "most of the Coorg castes are referred to in the Madras report". The provincial census was printed in Madras.	0

F.2 Mughal Provinces, and their Zamindari *Qaums*

Table F.2: Mughal Provinces in 1605 and their Zamindari *Qaums*

Sarkar	**Zamindar**	**Lower Caste**	**Enclosed Modern District**	**British district**	**British province**
Agra, Agra	Various (Rajput, Jat, Lodh, Brahmin, Ahir, Sayyid, Shaikhzada, Chishti)	Ahir, Jat, Lodh	Agra	Agra	United Provinces
Ahmadábád, Gujarat	Various (Rajput, Koli)	Koli	Gandhinagar	Mahi Kantha	Bombay
Ahoms Ahoms	Ahom	Ahom	Sivasagar	Sibsagar	Assam
Ajme'r, Ajmer	KachhwÁhah, AfghÁn, ChauhÁn	None	Ajmer	Ajmer Merwara	Ajmer Merwara

Allahábád Allahabad	Various (Brahman, Bhar tribe, Rajput, Kayath, Rahmatullahi)	Bhar	Fatehpur	Fatehpur	United Provinces
Alwar, Agra	Various (Rajput, Meo, Jat, Muslim clans)	Jat, Meo	Alwar	Alwar	Rajputana Agency
Badáon, Delhi	Various (Tyagi, Rajput, Kayastha, Shaikhzada, Kahor)	Kahar	Budaun	Budaun	United Provinces
Baglana, Baglana	Rajput	None	Nashik	Nasik	Bombay
Bahraich, Oudh	Rajput	None	Bahraich	Bahraich	United Provinces
Bahróch (Broach) South, Gujarat	Various (Rajput, Gwalia)	Gwalia	Bharuch	Broach	Bombay
Baroda, South, Gujarat	Rajput	None	Vadodara	Baroda	Baroda
Básim, Berar	Rajput	None	Washim	Basim	Berar
Bariálah (Pitálwári), Berar	Rajput	None	Jalgaon	Khandesh Agency	Bombay
Bayánwán, Agra	Various (Ahir, Gurjar,Mina, Jat, Rajput, Brahman, Marwari)	Ahir, Gujar, Mina, Jat	Shivpuri	Isagarh	Gwalior
Behar, Bihar	Various (Brahman, Kayath, Rajput, Afghan, Shaikhzada)	None	Nalanda	Patna	Bengal
Benáres (East), Allahabad	**Various (Brahman, Rajput)**	**None**	**Varanasi**	**Banares**	**United Provinces**
Bhadrak, Bengal	Various (Khandait, Kayath)	None	Bhadrak	Balasore	Bengal
Bhathkhora (South), Allahabad			Rewa	Baghelkhand	Central India Agency
Bíja'garh, Malwa	Various (Rajput, Bhil, Brahman, Kahar and others)	Bhil, Kahar	Barwani	Bhopawar	Central India Agency
Bikane'r, Ajmer	Rajput (Bhati)	None	Bikaner	Bikaner	Rajputana Agency
Chámpane'r, Gujarat	Various (Rajput)	None	Vadodara	Baroda	Baroda
Champáran, Bihar			Gopalganj	Saran	Bengal

Chanádah (Chana'r) South, Allahabad	Siddique, Faruqui, Ansari	None	Varanasi	Banares	United Provinces
Chanda, Gondwana	Gond	Gond	Sagar	Saugor	Central Provinces
Chande'ri, Malwa	Various (Bagri, Khatri, Ahir, Brahman, Rajput, Kayath, and other tribes)	Ahir	Ashoknagar	Isagarh	Gwalior
Chítór, Ajmer	Rajput	None	Chittaurgarh	Mewar	Rajputana Agency
Dánde's, Dandes	Koli, Ahir, Gond	Ahir, Gond, Koli	Burhanpur	Nimar	Central Provinces
Delhi, Delhi	Various (Rajput, Ahir, Jat, Afghani, Brahman, Gujar, Shaikhzada etc)	Ahir, Gujar, Jat	Delhi West	Delhi	Punjab
Deogarh, Gondwana	Gond	Gond	Chhindwara	Chindwara	Central Provinces
Gágrón, Malwa			Jhalawar	Jhalawar	Rajputana Agency
Garha Madla, Gondwana	Gond	Gond	Mandla	Mandla	Central Provinces
Gáwil, Berar			Amravati	Amraoti and Ellichpur	Berar
Gházípúr (East), Allahabad	Various (Rajput, Brahman, Kayath)	None	Ghazipur	Ghazipur	United Provinces
Ghorághát, Bengal	Various		Dakshin Dinajpur	Dinajpur	Bengal
Godhrá, Gujarat	Various		Panch Mahals	Panch Mahals	Bombay
Gorakhpúr, Oudh	Various (Rajput, Bansi, Afghani)	Bansi	Gorakhpur	Gorakhpur	United Provinces
Gwálior, Agra	Various (Rajput, Gujar, Jat, Bagri, Brahman)	Gurjar, Jat	Morena	Gwalior	Gwalior
Hájipúr, Bihar			Vaishali	Muzaffarpur	Bengal
Hindíah, Malwa	Various		Harda	Hoshangabad	Central Provinces
Hisár Fírózah (Hissár), Delhi	Various (Jat, Jati (Chamar), Rajput, Afghan and others)	Jat, Jatav	Hisar	Hissar	Punjab
I'ríj, Agra	Various (Rajput, Brahman, Kayath, Kunbi, Afghan)	Kunbi	Jalaun	Jalaun	United Provinces

Jale'sar, Bengal	Various (Khandait, Rajputs, Brahman, Bhanj Rajputs (Bhaj))	None	Baleshwar	Balasore	Bengal
Jaunpúr (NORTH), Allahabad	Various (Rajput, Sayyid, Rahmattulah, Kurmi, Brahman, Kayath, Siddique Ansari, Bachgoti)	Kurmi	Jaunpur	Jaunpur	United Provinces
Jodhpúr, Ajmer	Rajput	None	Jodhpur	Marwar	Rajputana Agency
Kálinjar (South), Allahabad	Various (Rajput, Gond, Khandelwal, Sayyid, Bhar, Bagri, Rahmatullah)	Bhar, Gond	Banda	Banda	United Provinces
Kallam (Kalamb), Berar	Gond	Gond	Yavatmal	Wun	Berar
Kálpi, Agra	Various (Rajput, Brahman, Kunbi, Afghan, Turkman)	Kunbi	Jalaun	Jalaun	United Provinces
Kamráj Kashmir	Various (Pandit, Ahir, Akbari etc)	Ahir	Srinagar	Kashmir	Kashmir
Kanauj, Agra	Various (Rajput, Brahman, Kharwar, Shaikhzada, Afghan)	Kharwar	Kannauj	Farukkhabad	United Provinces
Kanauj, Malwa	Gond	Gond	Jabalpur	Jubbalpore	Central Provinces
Kangra, Kangra	Rajput	None	Kangra	Kangra	Punjab
Karrah (West), Allahabad	Various (Rajput, Brahman, Kayath, Afghani)	None	Kaushambi	Allahabad	United Provinces
Katak (Cuttack), Bengal	Various (Brahman, Ahir, Khantait, Rajput, Tenligha)	Ahir	Cuttack	Cuttack	Bengal
Khairábád, Oudh	Various (Brahman, Rajput, Ahir, Ahin, Bcchal)	Ahir	Sitapur	Sitapur	United Provinces
Kherlah, Berar	Gond	Gond	Betul	Betul	Central Provinces
Koch, Koch	Koch	Koch	Koch Bihar	Kuch Bihar	Bengal
Kol (Koil), Agra	Various (Rajput, Brahman, Gurjar, Jat, Afghan, Sayyid, Shaikhzada, Badgujar)	Gurjar, Jat	Aligarh	Aligarh	United Provinces

Korarah (Corah) West, Allahabad	Various(Afghan, Lodhi, Rajput, Brahman)	Lodh	Allahabad	Allahabad	United Provinces
Kótrí Paráyah, Malwa	Various (Rajput, Kayath, Sondha tribe of Rajput)	None	Jhalawar	Jhalawar	Rajputana Agency
Kumáon, Delhi	Various		Almora	Almore	United Provinces
Lakhnauti, Bengal	KÁyaths and Brahmans	None	Maldah	Malda	Bengal
Lucknow, Oudh	Various (Rajput, Brahman, Ansari, Sayyid, Kunbi, Shaikhzada, Afghani, Bacchal)	Kunbi	Lucknow	Lucknow	United Provinces
Madáran, Bengal	Various		Hugli	Hooghly	Bengal
Mahkar, Berar			Buldana	Buldana	Berar
Máhór, Berar	Rajput	None	Yavatmal	Wun	Berar
Mandláer, Agra	Rajput	None	Karauli	Karauli	Rajputana Agency
Mando Malwa	Various		Indore	Indore	Central India Agency
Manikdrug, Berar			Chandrapur	Chanda	Central Provinces
Mánikpúr, Allahabad	Various (Brahman, Rajput, Kayath, Bawariya, Bachgori, Turkoman, Khandelwal)	Baoria	Kaushambi	Allahabad	United Provinces
Marósór, Malwa	Various (Rajput, Ahir, Gond)	Ahir, Gond	Mandsaur	Malwa (Gwalior)	Gwalior
Marráj Kashmir	Various (Pandits, Thakors, Zinah)	Zinah (Ahir)	Srinagar	Kashmir	Kashmir
Monghyr, Bihar			Munger	Monghyr	Bengal
Nádót (Nondod) NORTH, Gujarat			Narmada	Broach	Bombay
Nágor, Ajmer	Various (Kushwaha, Rajput Muslim)	None	Nagaur	Marwar	Rajputana Agency
Narnálah, Berar	Gond	Gond	Akola	Akola	Berar
Nárnol, Agra	Various (Rajput, Jat, Gond, Ahir, Musalman,Muslim Rajput)	Ahir, Gond, Jat	Mahendragarh	Hybrid	Punjab
Narwar, Agra	RajpÚt	None	Shivpuri	Isagarh	Gwalior

Nazarbár, Malwa	Various		Tapi	Navsari	Baroda
Oudh, Oudh	Various (Brahman, Kunbi, Rajput, Ansari, Bachgoti)	Kunbi	Faizabad	Fyzaba	United Provinces
Panár, Berar			Wardha	Wardha	Central Provinces
Páthri, Berar			Parbhani	Parbhani	Hyderabad
Pattan, NORTH, Gujarat	Various (Rajput, Koli)	Koli	Mahesana	Kadi	Baroda
Pinjarah, Bengal	Various		Dakshin Dinajpur	Dinajpur	Bengal
Púrniyah, Bengal			Purnia	Purnea	Bengal
Ráisín, Malwa	Rajput	None	Raisen	Bhopal	Central India Agency
Ránghar (Rámgarh), Berar			Karimnagar	Elgandal	Hyderabad
Ranthamór, Ajmer	Rajput	None	Sawai Madhopur	Jaipur	Rajputana Agency
Re'wári, Delhi	Various (Rajput, Ahir, Jat, Thathar and others)	Ahir, Jat	Rewari	Gurgaon	Punjab
Rohtás, Bihar	Various		Rohtas	Shahabad	Bengal
Sahár, Agra	Various (Meo, Jat, Ahir, Kushwaha Rajput)	Ahir, Jat, Meo	Mathura	Muttra	United Provinces
Saháranpúr, Delhi	Various (Jat, Tyagi, Brahman, Sayyid, Ansari and others)	Jat	Saharanpur	Saharanpur	United Provinces
Sambhal, Delhi	Various (Tyagi, Rajput, Sayyids, Khassiah, Jat, Ahir and others)	Ahir, Jat	Jyotiba Phule Nagar	Moradabad	United Provinces
Sáran, Bihar	Various		Siwan	Saran	Bengal
Sárangpúr, Malwa	Various (Rajput, Kayath, Kacchi)	Kacchi	Shajapur	Malwa (Gwalior)	Gwalior
Sátgáon, Bengal	Various		Kolkata	Calcutta	Bengal
Sharífábád, Bengal	Various		Barddhaman	Burdwan	Bengal
Sirhind, Delhi	Various (Rajput, Brahman, Jat, Shaikh and others)	Jat	Fatehgarh Sahib	Patiala	Punjab
Síróhi, Ajmer	Rajput, Afghan	None	Sirohi	Sirohi	Rajputana Agency
Sórath, Gujarat			Rajkot	Kathiawar	Bombay

Sulaimánábád, Bengal	Various		Hugli	Hooghly	Bengal
Súrat, Gujarat	Various (Rajput)	None	Surat	Surat	Bombay
Tájpúr, Bengal	Various		Uttar Dinajpur	Dinajpur	Bengal
Telingánah, Berar			Nanded	Nander	Hyderabad
The Bet Jálandhar Doáb Lahore	Various (Rajput, Afghan, Lodhi, Bhatti and others)	Lodh	Kapurthala	Kapurthala	Punjab
Tijárah, Agra	Various (Meo, Rajput, Muslim Rajput)	Meo	Mewat	Gurgaon	Punjab
Tirhut, Bihar	Various		Muzaffarpur	Muzaffarpur	Bengal
Udne'r, Bengal			Maldah	Malda	Bengal
Ujjain, Malwa	Various (Rajput, Jadon, Mehtar, Soria other unidentified tribes)	Mehtar	Ujjain	Malwa (Gwalior)	Gwalior